The Semantics of Definite and Indefinite Noun Phrases
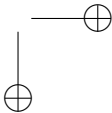
Irene Roswitha Heim

Irene R. Heim
The Semantics of Definite and Indefinite Noun Phrases
M.A., Ludwig-Maximilians-Universität München,
Ph.D., University of Massachusetts
Directed by Professor Barbara Partee

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

DOCTOR OF PHILOSOPHY

September, 1982
Linguistics
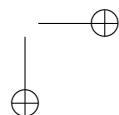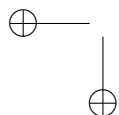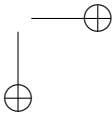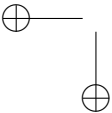
PREFACE

In November 1978, a workshop was held at the University of Massachusetts whose title was "Indefinite Reference" and whose topic Barbara Partee described in a circular that started as follows:

> One standard view among logicians is that indefinite noun phrases like 'a tall man' are not referring expressions, but quantifier phrases, like 'every man', 'no man', and 'most men'. Yet in many respects, indefinite noun phrases seem to function in ordinary language much like definite noun phrases or proper names, particularly with respect to the use of pronouns in discourse. This may be simply a matter of sorting out semantics from pragmatics, but there is not to our knowledge any currently available theory that simultaneously characterizes the logical or truth-functional properties of indefinite noun phrases and accounts for their 'discourse-reference' properties ...

A second installment of the same workshop took place in March 1979, and I was then so preoccupied by the topic that I immediately wrote a paper called "Toward a unified semantics of definite and indefinite noun phrases," in which I treated all definites and indefinites as variables. In the following months I grew disaffected with that paper because it did not tell a convincing story about the reasons why definites and indefinites, both being variables, were so unlike each other. I wished that all the differences could somehow be made to follow from a difference in the presuppositions of definites and indefinites, But it was only in the summer of 1980 that I began to make sense of this idea. Then the writing of this dissertation took almost two more years. It still feels like a rough draft, and I hope the reader will view it as such and excuse the redundancies, inconsistencies, and omissions, not to speak of mistakes.

The workshop at which my dissertation topic was laid out for me was just one of the many stimulating and instructive events that contributed to my education. I was lucky to be first a student and then a regular guest at one of the liveliest linguistics departments in Germany, at the University of Konstanz, and I was even luckier to become moreover a student at one of the liveliest linguistics departments in the world, at the University

ABSTRACT

Logical semanticists have standardly analyzed definite and indefinite descriptions as quantifiers, and definite pronouns as variables. This dissertation explores an alternative analysis, according to which all definites and indefinites are quantifier-free, i.e., consist of an essential free variable and the descriptive predicate (if any). For instance, "A cat arrived" is analyzed as "cat(x) ∧ arrived(x)." The existential force that such sentences carry in their unembedded (and some of their embedded) uses is attributed not to the indefinite article, but to principles that govern the interpretation of variables in general.

The primary motivation for this variable-analysis is that it leads to a straightforward account of indefinites serving as antecedents for pronouns outside their scope, as in the so-called "donkey sentences." These are handled by combining the variable-analysis of indefinites with a treatment of quantifiers as basically unselective, drawing on work by David Lewis.

Since a variable-analysis of both definites and indefinites prima facie obliterates the distinctions between the two, it must be accompanied by a new theory of the definite-indefinite contrast. This theory must account for the different conditions under which definites and indefinites can get bound, and for the exclusive capacity of definites for deixis and anaphora. All these differences can be predicted if the uniform semantic analysis of definites and indefinites is supplemented by suitable assumptions about their contrasting felicity conditions (presuppositions): Felicitous definites must be "familiar" variables, felicitous indefinites must be "novel" variables.

Familiarity may be a matter of having an antecedent in the text, or else of having a contextually salient referent. The underlying concept of familiarity that unites these two cases is best captured in an enriched semantic theory, which includes a further level of analysis, here called the "file" level. A major part of the dissertation is devoted to developing and motivating such a theory of "file change semantics," and to making precise its relation to conventional truth-condional semantics.
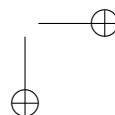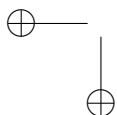
CONTENTS

Chapter I

PROBLEMS CONCERNING INDEFINITES AND ANAPHORA IN LOGICAL SEMANTICS

Logical semantics, i.e., the analysis of natural language in terms of concepts of modern logic, is an enterprise as old as modern logic itself. Logical semanticists have been using such notions as quantification, variable binding, and scope in trying to characterize the semantics of words and constructions in natural languages. In the realm of noun phrase semantics, for instance, they have studied the quantificational force of various determiners and have analyzed certain semantic relations between noun phrases and pronouns as variable binding.

While some of the earlier work in this tradition was directed at purposes of no concern to the linguist (as when philosophers hoped to expose their colleagues' fallacious arguments more convincingly by applying logical analysis to them), there has been increasing interest in logical semantics as a contribution to linguistic theory, more specifically, a contribution to a theory of one subsystem of grammar, the (or one of the) semantic component(s). Seen in this way, logical semantics is not just about algorithms for translating one language into another language, the former natural, the latter artificial. Rather, it implies a commitment to the hypothesis that the grammar of a natural language does in fact include a component in which expressions are analyzed as quantifiers, as variables, as having a certain scope, as standing in binding relations to each other, and the like. In other words, the logical semanticist hypothesizes that the ability to assign to English sentences the logical analyses he or she is proposing is part of knowing English.

By and large, logical semantics has been a very successful enterprise, especially with regard to such topics as the semantics of determiners and the semantics of pronouns, including such issues as are central to this dissertation. Analyses in terms of quantifiers and variables not only have captured the intuitive truth conditions of sentences involving those expressions. They
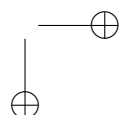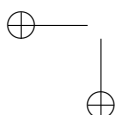
have also made it possible to formulate certain principles that relate the structural properties of sentences to the range of readings of which they admit. The existence of these principles is an entirely contingent fact about natural language, and the finding that they make reference to aspects of the logical analysis of a sentence gives justification (of a sort that could not be anticipated in the early days of logical semantics) to the hypothesis that grammars of natural languages do in fact assign logical analyses to sentences.

Notwithstanding the successes alluded to, there remains a residue of data that are widely believed to have not received an entirely satisfactory treatment in logical semantics. This chapter is about those data and the attempts at accounting for them that have been made. Actually, the scope of the chapter is more limited than that in at least two respects. Firstly, I will focus only on examples that exhibit pronominal anaphora to an indefinite antecedent which is singular and contains the indefinite article "a/an." I will not consider the various types of plural indefinites which show to some extent an analogous behavior. (The limitation to singular NPs will in fact be maintained throughout the dissertation.) I will also avoid talking about certain related phenomena involving a definite description as the antecedent (but see Chapter II, section 6). These omissions will help to keep the discussion perspicuous, but they may also have permitted me to jump to conclusions that will not survive future scrutiny.

Secondly, I will not include in this chapter all existing approaches to the problems under discussion, not even all that are known to me and that fall under the general orientation of logical semantics. Most notably, I will exclude David Lewis' proposal in "Adverbs of Quantification" (1975) from the discussion here, because I see it as a direct precursor of my own proposal and want to present it as such in Chapter II. I will also exclude Hans Kamp's paper, "A Theory of Truth and Semantic Representation" (1981), whose leading ideas are very similar to my own and were developed independently at approximately the same time. There will be a few brief remarks on Kamp's work in comparison to mine in later chapters, but no attempt will be made to present arguments that would bear on a choice between the two. I hope this will be possible at some future time. An approach I will discuss, though without coming to a definitive judgment about it, is the game-theoretical approach (also a variety of "logical semantics," as I would apply that label) that Jaakko Hintikka and Lauri Carlson advocate in their article, "Conditionals, Generic Quantifiers, and other Applications of Subgames" (1979). This approach differs rather fundamentally from the ones to be discussed in this chapter, and one also does not perceive it as related to my own approach. I will append what little I have to say about it to the very end of the chapter.
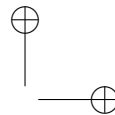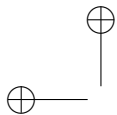
Given all these limitations, it should be clear that the purpose of this chapter cannot possibly be to convince the reader that the problems addressed in it have remained unsolved to this day. No such conclusion could be based on a survey of proposals that neglects precisely the most promising ones. Rather, one of the purposes is to review some facts that will later enter into the motivation of my proposal, and another purpose is to point out weaknesses in some, though not all, of the proposals that are currently being advocated. As for the latter aim, let me add another caveat: The weaknesses that I will find in some of the proposals discussed are by no means dramatic, and they do not justify rejecting them out-of-hand. It remains to be seen what alternatives there are, and at what cost they overcome the weaknesses in question. And, needless to say, this evaluation by comparison must be based on more criteria than mere coverage of data. So this chapter cannot possibly convey all of the reasons I see for choosing the position I will in fact adopt over the available alternatives. More reasons will have to emerge in the course of the following chapters.

1   Do Indefinites Refer?

There is a familiar set of arguments to the effect that indefinites cannot be referring expressions, which I will review very briefly. Then there is one commonly used argument which appears to favor the opposite conclusion. This argument relies on certain assumptions concerning the nature of anaphoric relations between an indefinite antecedent and an anaphoric element. I will devote the bulk of this section to clarifying those assumptions and the argument based on them, and to various attempts at invalidating it that have been made.

1.1   Russell's view

What is the meaning of a noun phrase with the indefinite article, e.g., the noun phrase "a dog" as it occurs in the following sentences?

(1)   A dog came in.
(2)   John is friends with a dog.

According to a most widely held view, originally advocated by Russell[1], such indefinite NPs (for short: "indefinites") have the meanings of existential

---

[1]   In various writings, especially Russell (1919, Ch. 16).
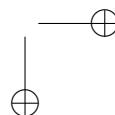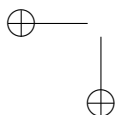
quantifiers. (1) is thus analyzed as meaning that the set of individuals that are both dogs and came in is not empty. And (2) is taken to mean that the set of individuals that are both dogs and friends with John is non-empty. On this view, indefinites do not refer: "a dog" in (1) does not denote any particular dog any more than "every dog" in (3) does:

(3) Every dog came in.

Arguments in favor of Russell's view include the following:[2] Consider a sentence like:

(4) John is friends with a dog, and Mary is friends with a dog.

If "a dog" referred to any particular dog, then (4) would have to be understood as saying about that dog, the referent of "a dog," that John is friends with it and so is Mary. But (4) is clearly consistent with a situation where no dog is a mutual friend of John's. Russell's analysis predicts that: According to it, (4) is true whenever neither the set of dog-friends of John nor the set of dog-friends of Mary is empty.

   This argument, though repeated often, tacitly relies on an extremely dubious premise, which was first exposed by Strawson: It presupposes that "a dog" either refers to one and the same thing on each occasion where it is uttered, or else it never refers to anything at all. A third possibility, viz., that each occurrence of "a dog" refers to something, but not the same thing in each case, is plainly ignored. The argument is thus on a par with the following attempt to show that personal pronouns do not refer: Suppose that "he" referred to some individual x. Then "He likes him" would have to mean that x likes himself, which it does not mean. Ergo, "he" does not refer to anything. What this argument, as well as the analogous one about indefinites, does show is that the *meaning of the expression* "he" (or the *expression* "a dog") is not to be identified with an individual that is supposed to be that expression's referent. But it is a huge step from this insight to Russell's view that indefinites are not referring expressions. Other familiar arguments pro Russell are harder to resist. Consider (5), the negation of (1):

(5) It is not the case that a dog came in.

If "a dog" referred to anything, we would expect (5) to be denyinq that that thing came in, just like (6) denies that the referent of the name "Fido" came in:

[2] These arguments are found in Quine (1960) and Kaplan (1970), among other places.

(6)   It is not the case that Fido came in.

But (5) expresses a much stronger claim than that: It says that no dog whatsoever came in. And this happens to be precisely what Russell's analysis predicts: Given that (1) means that the set of dogs that came in is not empty, (5) , the negation of (1), means that it is not the case that the set of dogs that came in is not empty, or equivalently: (5) means that the set of dogs that came in *is* empty.

Next, consider:

(7)   Every child owns a dog.

If "a dog" referred to anything, then (7) should amount to claiming about that thing that every child owns it, just as (8) claims that every child owns Fido.

(8)   Every child owns Fido.

Again, Russell's analysis of (7) is preferable. It predicts that (7) means that for every child x, the set of dogs owned by x is non-empty. Considerations like these seem to have convinced not only several generations of philosophers, but also a great many linguistic semanticists that are currently working in either the "Montague Granmar" or the REST framework.[3] Even those who are not convinced that *all* uses of indefinites are properly analyzed by Russell's theory have mostly conceded that it captures at least one of the readings that an indefinite can take on.

It has often been pointed out (e.g., in Strawson (1952)) as a difficulty for Russell's analysis of indefinites that indefinites can serve as antecedents for anaphoric pronouns. For instance, (1) can be continued as in (9):

(9)   A dog came in. It lay down under the table.

The pronoun "it" in this text presumably refers to something. But what does it refer to? The first answer that comes to mind is that it refers to the same dog as "a dog" in the preceding sentence. But for that answer to make any sense, "a dog" itself must, contra Russell, refer to some dog. This consideration of course does not prove that "a dog" in (9) refers; it merely shows that, if we assume that "a dog" in (9) refers, we can give a simple and sensible account of what "it" means in a text like this, whereas the role of "it" would seem unaccounted for and mysterious under Russell's assumption that "a dog" does not refer. So we need not yet conclude that the Russellian

---

[3] Montague (1974), Barwise and Cooper (1981), May (1977), to mention only a few.

view is wrong; but we have to acknowledge that its advocates owe us an alternative account of anaphoric pronouns with indefinite antecedents if we are to remain convinced.

There are various ways in which Russellians have risen to this challenge, and I will sketch the most important lines of approach in the next few sections.

1.2 Anaphoric pronouns as bound variables

Example (9) remains puzzling as long as we accept Russell's analysis of the indefinite while presuming that the pronoun must refer. But why should it? It is well recognized that reference is not the only function that natural language pronouns can serve, and that there are cases where they clearly must be understood as bound variables. (9) might be just another case like that. Indeed, we predict intuitively correct truth conditions for this text if we analyze it as saying that the set of dogs that came in and lay down under the table is non-empty, i.e., as equivalent to:

(9′) $\exists x(x$ is a dog $\wedge$ x came in $\wedge$ x lay down under the table)

This analysis was proposed by Geach.[4] It implies as a general moral that the proper unit for the semantic interpretation of natural language is not the individual sentence, but the text. (9′) provides the truth condition for the bisentential text as a whole, but it fails to specify, and apparently even precludes specifying, a truth condition for the individual sentence "It lay down under the table," which appears as part of (9).

Geach's analysis is quite unpopular these days and is sometimes held to have been amply refuted by the following considerations: First of all, how is this analysis supposed to deal with dialogues like the following (example (10) is due to Strawson).[5]

(10)   a. A man fell over the edge.
       b. He didn't fall; he jumped.

(11)   a. A dog came in.
       b. What did it do next?

With (10), where the two speakers contradict each other, an analysis of the whole dialogue as a unit could only amount to assigning it the necessarily false truth condition:

---

[4] Geach (1962, 126ff).

[5] Strawson (1952, 187).

(10′)   ∃x(x is a man ∧ x fell over the edge ∧ ¬x fell over the edge ∧ x jumped
         over the edge)

One can hardly object to the claim that this is indeed the truth condition of the dialogue as a whole. But that just goes to show that there is rather limited interest in an analysis that is confined to assigning truth conditions to whole discourses only. It is intuitively clear about (10) that each speaker has expressed a contingent proposition and that we can imagine circumstances under which we say, e.g., that A said something false and B said something true. There has to be something wrong with a theory which implies that neither utterance has any truth conditions of its own. (11) emphasizes basically the same point. In this case, B's utterance of course does not have truth conditions, since it is a question. Rather, it has a question-meaning, whatever that may be (several plausible proposals are in existence, among them that question meanings are sets of propositions). The relevant point is that Geach's approach to anaphoric pronouns would force us to look for a meaning of the dialogue as a whole, but one cannot even imagine what type of meaning that would be.

Let us be clear about the force of this objection before we move on to the next. Examples like the ones just considered do not show that Geach's analysis makes false predictions; they merely show that it does not make enough predictions. In particular, it falls short of predicting truth conditions for the full range of units of discourse that are capable of truth or falsity in the pretheoretical sense. So the objection is a challenge to, not a refutation of, the Geachian position, and it stands as an objection only as long as no one comes up with a truth criterion for sentence-sized parts of utterances that is consistent with Geach's treatment of intersentential anaphora as variable binding. I emphasize this point, because I will later have to face the same objection with respect to my own approach, and will then try to argue that the challenge can indeed be met (see Chapter II, section 3, and Chapter III, section 3, below).

An additional, independent objection to Geach's analysis is due to Evans[6] who claims that it even predicts inadequate truth conditions in some of the cases where a multisentential discourse basically does amount to the assertion of one complex proposition. His most striking examples are those involving a plural indefinite that serves as an antecedent to a plural pronoun. Plural indefinites presumably mean something like "there are at least two."

(12)   John owns some sheep.

---

[6] Evans (1980, 339).

Apparently (12) is true if and only if the set of sheep that John owns has at least two members. Given this, what truth conditions would Geach's analysis lead us to expect for (13)?

(13)   John owns some sheep. Harry vaccinated them.

By analogy with the singular examples, (13) should amount to the assertion that the set of sheep that John owns and that Harry vaccinated contains at least two members. Consequently, (13) ought to count as true in a situation where John owns six sheep and Harry vaccinates three of them, but does not vaccinate the other three. But intuitively, we are inclined to consider (13) false in this situation. We tend to read (13) as meaning that John owns some sheep and Harry vaccinated *all* the sheep that John owns.

   Evans even goes so far as to claim that a Geachian analysis predicts inadequate readings in cases with singular indefinites. He says that if such an analysis were correct, then we should be able "to report the non-emptiness of the class of Welsh doctors in London by saying:

(14)   There is a doctor in London and he is Welsh."[7]

But (14) is undoubtedly inappropriate for conveying the message. Contrast this with (15), which *would* be appropriate:

(15)   There is a doctor in London who is Welsh.

Geach's problem is that he analyzes (14) in exactly the same way as (15) and thus fails to explain the difference. (I will take a closer look at the contrast between (14) and (15) below and will suggest that Evans himself does not have a satisfactory explanation for it either. But that is an independent point, and it does not change the fact that the observed contrast casts doubt on the correctness of Geach's analysis.)

   There is a third objection that is often brought forward against analyses which, like Geach's, rely on attributing to indefinites a scope that may encompass as much as an entire text. The objection is that an analysis of this type necessitates ad hoc qualifications to what would otherwise be a uniform theory of structural constraints on quantifier scope. (This is of course not an absolute inadequacy of the analysis under consideration, but merely a relative disadvantage in comparison with certain competing analyses, which we have yet to discuss.) The objection draws on the observation that many of the examples where an existential antecedent supposedly binds an anaphoric

---

[7] Evans (1980, 343)

pronoun do not have grammatical analogues with other quantifiers in the antecedent position. The contrast between (9) on the one hand and (16) or (17) on the other is typical:

(9) A dog came in. It lay down under the table.
(16) Every dog came in. It lay down under the table.
(17) No dog came in. It lay down under the table.

Neither (16) nor (17) permits an interpretation where "it" is bound by the quantified NP in the preceding sentence. (In fact, these texts cannot be felicitously used unless a referent for "it" is fixed in some way that is completely independent of the utterance of the preceding sentence.) The generalization behind this fact is that an unembedded sentence is always a "scope-island," i.e. a unit such that no quantifier inside it can take scope beyond it. This generalization (which is just a special case of the structural restrictions on quantifier-scope and pronoun-binding that have been studied in the linguistic literature[8]) is only true as long as the putative cases of pronouns bound by existential quantifiers under Geach's analysis are left out of consideration. Therefore, if Geach's analysis of (9) as an instance of variable binding is adopted, the conditions have to be modified in appropriate ways to predict the seemingly exceptional behavior of existential quantifiers. To be sure, this point can only be considered an objection to the Geachian analysis if it is suspected that such a modification would result in a considerable loss of explanatory power of the overall theory, and that this can be avoided by adopting a different analysis. I postpone the issue until section 2.1.2., where I'll suggest that the objection does not hold up in light of a more fully developed version of Geach's approach.

1.3 Anaphoric pronouns as picking up a speaker's reference

A radically different approach to the relation between an indefinite antecedent and a pronoun anaphoric to it was originally suggested by Grice.[9] The suggestion was taken up by Kripke (1977), on whose paper I will base my discussion. Let us recall once more how the puzzle originally presented itself (e.g., in Strawson's or Geach's exposition of it): On the one hand, anaphoric

---

[8] Various theories of scope-island restrictions and restrictions on bound-variable readings of pronouns are found in Rodman (1976), Reinhart (1976), May (1977), and Higginbotham (1980).

[9] Probably in a talk that Grice gave at some conference in 1971, and perhaps also in the 1967 William James Lectures on "Logic and Conversation."

pronouns seem to pick up the reference of their antecedents; on the other hand, Russell seems to have shown that indefinites do not have a reference. Therefore, they should not be able to serve as antecedents for anaphoric pronouns, but in fact, they are.

Unlike Geach, who attempted, to get out of the dilemma by denying that anaphora need have anything to do with "picking up a reference" previously made, Kripke thinks that there is not really a dilemma at all; there only appears to be one because two distinct notions of reference have been confounded in the statement of the puzzle: "speaker's reference" and "semantic reference," as he calls them. Whether an utterance of an expression has semantic reference, and what its semantic referent is, is purely a matter of the rules that define the language in which the utterance is made. It is in this sense of "reference" that Russell's claim about indefinites is to be understood: No utterance of an indefinite has semantic reference. But this does not mean that no such utterance has speaker's reference. Whether there is speaker's reference and what the speaker's referent is depends on the intentions that the speaker of the utterance happens to have. The speaker's referent is that individual which the speaker "wishes to talk about" (or "has in mind") on the occasion of the utterance. The notion applies to utterances of indefinites as follows:

> When a speaker asserts an existential quantification, $(\exists x)(\phi x \& \psi x)$, it may be clear which thing he has in mind as satisfying 'ϕx' and he may wish to convey to his hearers that that thing satisfies 'ϕx'. In this case, the thing in question (which may or may not actually satisfy 'ϕx') is called the 'speaker's referent' when he makes the existential assertion. (Kripke, 1977, 17)

So it is perfectly conceivable for an utterance that lacks semantic reference to have speaker's reference nevertheless, and this might even be the typical situation with indefinites.

Only one additional assumption is needed now to reconcile Russell's analysis of indefinites with the naive view that anaphoric pronouns pick up the reference of their antecedents: If we assume that they pick up the speaker's reference of their antecedents, i.e., that an anaphoric pronoun's semantic referent is (or at least, may be) the speaker's referent of its antecedent, then examples like (9) are naturally accounted for, with Russell's claims unaffected and the naive identification of anaphora with coreference basically vindicated. In Kripke's words (1977, 26, n32):

> Often one hears it argued against Russell's existential analysis of indefinite descriptions that an indefinite description may be anaphorically referred to by a pronoun that seems to preserve the reference of the indefinite description. I am not sure that these phenomena do conflict

with the existential analysis. (I am not completely sure there are some that don't, either. [sic]) In any event, many cases can be accounted for (given a Russellian theory) by the facts that: (i) existential statements can carry a speaker's reference; (ii) pronouns can refer to the speaker's referent.

As Kripke furthermore notes, it is not at all an unnatural assumption about pronouns that they should be able to get their reference in this way, which depends on pragmatic rather than purely semantic factors. After all, pronouns are also capable of getting their reference by virtue of such purely pragmatic factors as an object's (perceptual or merely associative) salience.

Kripke, if I understand his remark correctly, deliberately refrained from claiming that *all* instances of a pronoun anaphoric to an indefinite can be accounted for by adopting his suggestion. In fact, the scope of his proposal may be much more limited than is often thought. Consider the following sentence:

(18) A dog has been rummaging in the garbage can.

I may utter this sentence in a situation where I have not witnessed the event directly, but am merely inferring what happened on the basis of the mess I am seeing. I have no idea which dog is responsible, so there presumably is no speaker's referent in Kripke's sense. Nevertheless, I can naturally go on to utter (19):

(19) It has torn open all the plastic bags.

The pronoun "it" in this utterance, certainly an anaphoric pronoun in the traditional sense of "anaphoric," cannot be analyzed as picking up a previous speaker's reference, because there was none. (Neither can it be analyzed as a bound variable in the scope of its antecedent, unless we want to burden ourselves with an analysis more or less like Geach's after all.)

Lewis (1979) sketches another essentially Gricean treatment of anaphoric expressions (pronominal and other) with indefinite antecedents that is similar to Kripke's, but might have more potential of extending even to cases like our example involving the utterance of (18) and (19). According to Lewis, a pronoun may refer to whatever object is maximally salient in the situation of its utterance. Anaphoric pronouns are a special instance of this, as one method of raising the salience of an object is by producing a suitable utterance. An utterance may be suited to raising an object's salience by containing an expression which (semantically) refers to that object; but it may also be so suited for reasons other than its involving any reference. Thus an existential statement may raise an object's salience in the following way:

> It is worth mentioning another way to shift comparative salience by conversational means. I may say 'A cat is on the lawn' under circumstances in which it is apparent to all parties to the conversation that there is some one particular cat that is responsible for the truth of what I say, and for my saying it. Perhaps I am looking out of the window, and you rightly presume that I said what I did because I saw a cat; and further (since I spoke in the singular) that I saw only one. What I said was an existential quantification; hence, strictly speaking, it involves no reference to any particular cat. Nevertheless it raises the salience of the cat that made me say it. Hence this newly-most-salient cat may be denoted, by brief indefinite descriptions, or by pronouns, in subsequent dialogue: 'No, it's on the sidewalk.' 'Has Bruce noticed the cat?' … Thus although indefinite descriptions – that is, idioms of existential quantification – are not themselves referring expressions, they may raise the salience of particular individuals in such a way as to pave the way for referring expressions that follow. (Lewis, 1979, 180)

Now the particular example that Lewis discusses in this passage is probably one where Kripke's notion of speaker's reference applies: The speaker who is looking out the window and seeing a cat certainly had that cat "in mind as satisfying 'x is a cat' " – although I find it somewhat less obvious that he "wished to convey to his hearers that that cat was on the lawn," or at least I can think of a way of reading the latter as ascribing him a wish that he might well have lacked. Anyway, let us assume this is a case of Kripkean speaker's reference. Then it would not seem to matter much whether we, account for the anaphoric pronoun that follows in Kripke's or in Lewis' terms. Either one would be satisfactory.

However, as we reconsider the example about (18), (19) in light of Lewis' remark, it might not present a problem at all. It could be argued that my utterance of (18) does raise the salience of a particular dog, namely the dog that "is responsible for the truth of what I say, and for my saying it," in other words, the dog that did rummage in the garbage. Just because I do not know anything else about this dog (what it looks like, what its name is, who it belongs to, …) does not mean it cannot become the most salient dog in the situation. So the anaphoric pronoun in (19) could after all be accounted for in much the same way as in the unproblematic cases of indefinites uttered with a speakers reference.

What if my utterance of (18) is false though, i.e., if the mess I am contemplating is in fact the result of a strong blast of wind? Which object has then been raised in salience? The answer is presumably: none. Thus, if I uttered (18) falsely, then I did not refer to anything when uttering "it" in (19), and my utterance of (19) did not express any proposition at all (although it would have expressed one, had the world in which my utterance took place been as I thought it was). I have no intuitive disagreement with this answer.

What, on the other hand, if my utterance of (18) were true, but true because a couple of dogs were in fact the offenders? Again, no dog can then have been promoted to maximal salience by my utterance, and the subsequent utterance of (19) must be considered a case of failure of reference, with no proposition expressed. Intuitions are quite unclear in cases like this. Sometimes I find it plausible that the utterance of (19) should be judged true just in case one of the dogs that rummaged in the garbage can also tore open all the plastic bags, and it should be judged false if neither one of the dogs that rummaged in the garbage can also tore open all the plastic bags. But that is certainly not the way the utterance was meant by the speaker; if she had only thought of the possibility that more than one dog was involved (which she obviously did not) she would not have committed herself to the completely unwarranted claim that only one of them tore open the bags while the other one stood by. At other times, I tend to think that the whole discourse, including (19), should be judged true, regardless of how many dogs rummaged in the garbage can and tore open the bags, as long as any did. But I certainly would not uphold this judgment if the discourse had run as follows:

(18)   (as above) [A dog has been rummaging in the garbage can.]
(20)   It was a very patient dog, as you can see from the fact that it tore open every single bag.

I think there is little point in scrutinizing intuitions in cases of this sort any further. They are insecure enough for us to accept a theory that predicts automatic truthvaluelessness whenever the existential antecedent is verified by multiple instances, as long as there is not anything else that is wrong with that theory. Against the Grice-Kripke-Lewis account of intersentential anaphora involving indefinite antecedents, I would like to bring to bear a consideration concerning minimal pairs like the following:[10]

(21)   a.   I dropped ten marbles and found all of them, except for one. It is probably under the sofa.
       b.   ? I dropped ten marbles and found only nine of them. It is probably under the sofa.

We might imagine the first sentence of (21b) to be uttered in a situation which is as suited as any to make the tenth, missing marble salient (say, everybody is engaged in the search). Still, the utterance does not seem to be a successful means of raising the salience of the tenth marble to a

---

[10] Examples are due to Barbara Partee.

degree of salience that would suffice for the subsequent pronoun to refer to it. By contrast, an utterance of the first sentence of (21a) does have that capacity. If we adopt Lewis' position on the logical analysis of indefinites and on the pragmatic preconditions for successful pronominal anaphora, then we are compelled to conclude that the salience-shifting potential of an utterance is not predictable from its truth-conditions and the surrounding circumstances alone; it moreover depends on how the utterance is worded. For example, notice that the first sentences of (21a) and (21b) coincide in truth conditions, and that the situations in which they are uttered are by hypothesis identical. (If we took Kripke's, rather than Lewis' view, a slightly different, but essentially analogous, conclusion would be inevitable. I will stick to Lewis' terms, but take it that my point carries over.)

The conclusion that was just drawn raises a question: How exactly does wording influence the capacity of an utterance to raise an individual's salience? Which properties of the sentence uttered are crucial, and what are the relevant generalizations? The phenomenon that (21) illustrates strikes me as a systematic one, and it calls for a systematic account. A first attempt to formulate the generalization of which our example is an instance might be this:

> A necessary condition for an utterance of a sentence S to promote an object x to maximal salience is that S contain either an NP that refers to x or a singular indefinite NP whose predicate is true of x.

This would discriminate between (21a) and (21b) in that the first sentence of (21a) contains the singular indefinite "one," here to be read as "one of the marbles," whose predicate amounts to "is one of the marbles" and is true of the tenth marble, the intended referent of the subsequent pronoun. The first sentence of (21b), on the other hand, contains neither a suitable singular indefinite, nor an expression that refers to the missing marble, so it is ruled out by the above generalization as a means of making that marble maximally salient.

If this generalization, or a more or less refined version of it, holds in fact true, we should wonder why. Does it need to be stipulated, or can it be explained? It seems to me that under the assumptions that Lewis is committed to, there is no way of relating the ability of singular indefinites to contribute to shifts in salience to any other characteristic property they have. Now it is not inconceivable that the salience-shifting potential of an expression is indeed a property that natural languages specify independently of the logical and other properties of that expression. But we would ceteris paribus prefer a theory which predicts it to correlate with other properties.

We have to note in this connection that the need for a separate stipulation disappears if we abandon Lewis' or Kripke's account of cross-sentential anaphora and take, say, Geach's approach. For Geach, the relevant reading of (21a) involves binding of the pronoun "it" by the text-scope existential quantifier "one (of the marbles)." (21b) cannot share that reading because there simply is no existentially quantifying NP to bind the pronoun. The contrast falls out from Geach's assumptions about the logical analysis of indefinites and their relation to anaphoric pronouns, with no additional stipulations needed. Also, if we assume, contra Lewis and Kripke, that some occurrences of indefinites refer and that intersentential anaphora relations involving indefinite antecedents arise when the pronoun picks up the reference of its antecedent (a view that I have attributed to Strawson) , the contrast in (21) can be accounted for much more naturally: (21b) simply does not contain an expression that refers to the missing marble and from which the pronoun could pick up that reference.

Much the same considerations apply to contrasting pairs like the following:

(22) a. John owns a bicycle. He rides it daily.
      b. ? John is a bicycle-owner. He rides it daily.

(23) a. John has a spouse. She is nice.
      b. ? John is married. She is nice.

Again, we find utterances of identical truth conditions exhibiting what Lewis would have to diagnose as different capacities for raising the salience of an object. And again we have a systematic phenomenon: Words (including compound words) are always "anaphoric islands" in the sense of Postal (1969).

To sum up my argument, the Grice-Kripke-Lewis approach to reconciling Russell's quantificational analysis of indefinites with their ability to serve as antecedents to anaphoric pronouns implies that a grammar of English will not only have to assign logical analyses to expressions of English, but in addition, and independently, will have to specify their salience-raising potential. This puts the approach at a prima facie disadvantage in comparison with alternative approaches that manage to get away with assigning logical analyses alone. Certainly this argument does not in any way refute the Gricean type of approach, since the one weakness it points to could turn out to be minor compared to the weaknesses of its competitors.

1.4   Anaphoric pronouns as disguised definite descriptions

There exists a third approach to anaphoric pronouns with definite antecedents that also attempts to maintain Russell's view that indefinites do not refer, but avoids both Geach's conclusion that such pronouns are bound variables, and the Gricean suggestion that they receive reference through pragmatic factors to which the utterance of the antecedent contributes only indirectly. This third approach is best represented by Evans.[11] Evans acknowledges the existence both of bound-variable pronouns and "pragmatic" pronouns (i.e., those referring by virtue of their referent's salience), but maintains that the anaphoric pronoun in our example (9) (repeated here) falls under neither of these two types, but under a separate type, which he calls "E-type".

(9)   A dog came in. It lay down under the table.

E-type pronouns always have quantified NPs as their antecedents, but they are not bound by them. Rather, their meaning is that of a definite description which is determined in a certain systematic way by the minimal sentence in which the antecedent occurs. For instance, the sentence "A dog came in," in which "a dog" occurs, determines the definite description "the dog that came in," and so that is the meaning of the subsequent "it," when interpreted as an E-type pronoun. (9) is thus analyzed as meaning the same as:

(9′)   A dog came in. The dog that came in lay down under the table.

Here are some more examples of E-type pronouns and the corresponding definite descriptions in terms of which Evans proposes to analyze them (some of the examples are found in Evans (1980)):

(24)   Just one man drank champagne. He was ill.
(24′)   Just one man drank champagne. The man who drank champagne was ill.

(25)   Few congressmen admire Kennedy. They are very junior.
(25′)   Few congressmen admire Kennedy. The congressmen that admire Kennedy are very junior.

(26)   John gave a book to a student. She didn't return it.
(26′)   John gave a book to a student. The student who John gave a book to didn't return the book John gave her.

---

[11] Evans (1977; 1980).

It should be clear intuitively from these examples how the appropriate definite description is constructed from the sentence in which the antecedent occurs. (It is possible, but not worthwhile for our present purposes, to specify an exact procedure for this construction.[12])

A paraphrase is not yet an analysis, and we cannot assess Evans' proposal unless we know how he intends to interpret definite descriptions. He assumes that a sentence containing a singular definite description with the predicate "F" implies that there is exactly one object that is F and is true if that unique F satisfies whatever the sentence predicates of "the F." (Whether "imply" is the same as "presuppose," or more like "entail," and whether falsity of the implication results in falsity or in truthvaluelessness of the assertion, is not specified by Evans and seems indeed to make little difference here.) As for plural definite descriptions with the predicate "F," he assumes the implicature that there are at least two Fs, and that the sentence containing "the F" (plur.) is true if *all* Fs satisfy the relevant predicate. These are commonplace assumptions, of course. Consequently, the predictions concerning the interpretations of E-type pronouns come out accordingly: For instance, when "It lay down under the table" is uttered as part of the text (9), it is predicted to imply that exactly one dog came in; it is furthermore predicted to be true just in case the unique dog that came in lay down under the table.

The prediction that E-type pronouns carry uniqueness-implications of this sort is prima facie controversial, and Evans devotes some remarks to defending it. First of all, he certainly does not mean to commit himself to the claim that the object in question should be the only one in the entire universe that satisfies the predicate restricting the description. Any quantification in ordinary conversation requires that we tacitly agree on a domain of relevant instances of evaluation and ignore the countless irrelevant ones that would be included under a strictly literal reading of the quantifier. In the same way, the range of objects within which a definite description, or an E-type pronoun, is presupposed to denote uniquely will always be a pragmatically determined range of relevant individuals. Take an example like (27):

(27)   A wine glass broke last night. It (i.e., the wine glass that broke last night) had been very expensive.

We must not misunderstand Evans as claiming that (27) carries an implication that in the entire world only one wine glass broke on the night before the utterance. Rather, the predicted implication is probably that exactly

---

[12] An exact procedure is spelled out in Parsons (1978).

one wine glass broke on that night in the household where the speaker was staying, or in a similarly confined contextually relevant place. (The same restriction would be needed to arrive at a realistic interpretation of an utterance like, "Every wine glass was refilled at least twice last night.")

Still, Evans I claim that E-type pronouns carry uniqueness-implications is a strong claim, even with the above qualifications. It predicts that the following pairs are by no means paraphrases:

(27)   (see above) [A wine glass broke last night. It (i.e., the wine glass that broke last night) had been very expensive.]

  a.   A glass which had been very expensive broke last night.

(14)   (see section 1.2 above) There is a doctor in London and he is Welsh.

  a.   There is a doctor who is Welsh in London.

(Recall that these *were* pairs of paraphrases under Geach's analysis.) In the case of (27) vs. (27a), I have no strong intuitions as to whether the meanings differ. Evans predicts that (27a) does not imply that just one glass broke on the night in question (within the relevant range of glasses), whereas (27) does imply this. I can well imagine the following scenario: I dropped a tray with three glasses last night, breaking them all. Two of them had been cheap and I did not mind losing them, but the third one had cost a lot and I was upset at its loss. The next morning, a friend asked me why I was in such a bad mood, and I answered by uttering (27). The utterance was, to me, perfectly appropriate. But I am not sure whether it constitutes a counterexample to Evans' analysis. He might simply say that, in the context of my friend's question, the cheap glasses were irrelevant because my mood was not influenced by their loss. This would explain how (27) and (27a), even if interpreted as differently as Evans suggests, end up being interchangeable after all, in the situation described, and probably for similar reasons in most (if not all) other situations as well.

A more perceptible intuitive contrast emerges in (14) vs. (14a), and indeed a contrast that appears to support Evans' analysis: As we noted earlier (when assessing Geach's analysis; cf. section 1.2), (14a) is an acceptable way of claiming that at least one person is both Welsh and a London doctor, whereas (14) is not. Evans offers an explanation for this lack of interchangeability: (14) carries the inappropriate implication that there is just one doctor in London, which is absent from (14a). However, it seems to me that this explanation is both unconvincing and replaceable. Imagine the following dialogue, which provides a natural context for (14a):

(28)   A: This boy is sick, but he speaks only Welsh and we can't
        understand his complaints.

B: There is a doctor who is Welsh in London = (14a)

If we replace (14a) by (14) in (28), the result is odd. But does Evans' analysis really predict this? Doesn't the preceding utterance by A establish a context in which only Welsh-speaking doctors qualify as relevant in the first place? If this were assumed, then we would expect (14) in this context to imply merely that there is just one Welsh-speaking doctor in London, and that would not be very awkward at all. So if the awkwardness of (14) as a response to (28A) is indeed, as Evans would have it, due to the fact that (14) implicates that London has just one doctor, then we must conclude that domains of quantification do not respond to relevancy criteria quite as flexibly as that. But then this casts doubt on the explanation given above for why (27) is not as inappropriate as we might have thought in a situation where I dropped three glasses. It is hard to see why the cheap glasses should not have counted as relevant in the dialogue described above, whereas the non-Welsh London doctors should count as relevant in (28). Explanations of acceptability judgments that have to rely on mysterious differences like this are not very convincing. Furthermore, there is a different reason to which one might attribute the oddity of (14) in most ordinary contexts: The first conjunct of (14), "There is a doctor in London," asserts something that is so well known that one can hardly conceive of a context in which there would be any point in asserting it. So it seems plausible that (14) is ruled out by the conversational maxim that one should not assert what is already presupposed. In order to discriminate between (14) and (14a) in the application of this maxim, we presumably must assume that in an assertive utterance of the form "S_1 and S_2," S_1 and S_2 each count as asserted individually, whereas an utterance of a complex sentence containing a restrictive relative clause is just one assertion, no parts of which must individually conform to the pragmatic appropriateness-criteria for assertions. Note that this line of explanation for the difference in acceptability between (14) and (14a) in no way appeals to a uniqueness implication of the sort that Evans attributes to (14).

I am inclined to deny any systematic validity at all to Evans' predictions concerning uniqueness-implications. Even with an example like (14), a seemingly insignificant variation makes any suggestion of uniqueness disappear:

(29)   There once was a doctor in London. He was Welsh ...

In view of the foregoing considerations, I think that the association of anaphoric interpretations with uniqueness-implications is not well supported by the linguistic facts. This suggests that there is something wrong with Evans' analysis of the kind of anaphoric pronoun he classifies as "E-type," but it presently remains unclear what specifically is wrong with it. Re-

call that I have attributed to Evans two assumptions which are independent of each other: (a) the assumption that certain anaphoric pronouns mean the same thing as certain definite descriptions, and (b) the assumption that definite descriptions are to be analyzed in a certain way, which involves predicting uniqueness-implications for singular definite descriptions. As it turns out upon closer investigation of the facts, it is (b) and not (a) that we should question: To the extent that it seems acceptable to utter (27) ("A wine glass broke last night. It had been very expensive.") without committing oneself to the belief that exactly one wine glass broke last night, it also seems appropriate to utter (27b) without implicating uniqueness:

(27)   b.   A wine glass broke last night. The wine glass that broke last night had been very expensive.

Likewise, (29a) is equally free from suggestions of uniqueness as (29):

(29)   a.   There once was a doctor in London. The doctor in London was Welsh.

(Neither of these sound very natural, but the same point can be made by observing the absence of uniqueness-implications with the definite descriptions in (30):

(30)   A wine glass broke on Monday night, and a beer glass broke on Tuesday night. The wine glass that broke on Monday night was my favorite wine glass, and the beer glass that broke on Tuesday night had been very expensive.)

So if we replaced Evans' assumption (b) by a more adequate analysis of definite descriptions, one which does not necessarily predict uniqueness-implications, could we thereby save Evans' theory of E-type pronouns? Perhaps we could, but the problem is that a more adequate analysis of definite descriptions of the sort needed is not readily available, and the task of developing one faces essentially the same difficulties as the task we have been discussing all along, viz., to account for anaphoric pronouns with indefinite antecedents. Definite descriptions that are anaphoric to indefinites probably work more or less in the same way as pronouns anaphoric to indefinites; and once we have settled on the right theory for the former, it might well carry over to the latter, or vice versa. But while we are in the process of looking for that theory, Evans' proposal to paraphrase anaphoric pronouns away in favor of anaphoric definite descriptions is not a solution – unless we agree with his view that definite descriptions, including those used anaphorically, have been treated satisfactorily in the tradition to which he subscribes.

1.5   Anaphoric pronouns and the ambiguity hypothesis

While I have so far dealt with various attempts to show how indefinites can be existential quantifiers as Russell suggested, and still serve as antecedents for anaphoric expressions, I will now examine the views of those who have concluded that, to the extent that indefinites participate in anaphoric relations, they simply defy Russell's analysis. This view goes back to Strawson (1952), where it is suggested that anaphoric relationships like the one in (9) ("A dog came in. It lay down under the table.") are instances of plain coreference, and that we must therefore acknowledge that indefinites can, at least sometimes, be used to refer.

Among recent defenders of Strawson's position, Chastain (1975) has been particularly outspoken:

> Sentences containing indefinite descriptions are ambiguous. Sometimes 'A mosquito is in here' and its stylistic variant 'There is a mosquito in here' must be taken as asserting merely that the place is not wholly mosquito-less, but sometimes they involve an intended reference to one particular mosquito. (1975, 212)

Note that the claim is not that indefinites always refer. It is still acknowledged that there is a reading for indefinites under which they conform to Russell's analysis. In fact, this seems impossible to deny, given the meanings of sentences that contain indefinites inside the scope of negation or quantifiers. One could not possibly claim, for instance, that "a dog" refers to anything when uttered as part of (31), and Russell's analysis is unrivaled with such examples.

(31)   I don't have a dog.

While examples like (31), where something else takes wider scope than the existentially read indefinite, are the most unmistakable cases of existential readings, Chastain does not think that the existential reading emerges if and only if something else has widest scope. A widest scope existential reading is supposed to exist as well; otherwise a sentence like (32)

(32)   There is a mosquito in here,

which contains no other scoped expression, would not be ambiguous, but admit only for the referential reading. But how are the two readings supposed to differ intuitively with examples of this sort? If I understand Chastain correctly, then the difference between a referentially read and an existentially read utterance of (32) is more or less the same difference as the one that Kripke would describe in terms of the presence or absence of

speaker's reference: When (32) is uttered with a particular mosquito in mind, Chastain considers this a case where "a mosquito" refers to that mosquito; when (32) is uttered with no particular mosquito in mind, he considers this a case where "a mosquito" is read as an existential quantifier. The difference between Chastain and Kripke is just that Chastain does not distinguish between semantic reference and speaker's reference, but uses the notion of reference simpliciter in such a way that it is coextensive with Kripke's notion of speaker's reference.

To the extent that Chastain's view differs from Kripke's other than just terminologically, is there anything to be said for it? We are not going to find an answer to this question in any of the considerations that Chastain himself brings to bear on the issue. As far as I can see, he – like Strawson before him – relies crucially on the plausibility that the ambiguity hypothesis derives from its compatibility with a simple explanation of anaphora: An anaphoric pronoun with an indefinite antecedent (where the latter does not bind it) is coreferential with its antecedent. But it is precisely in this respect that Chastain's proposal is almost identical to and no better than Kripke's: just replace "is coreferential with" by "picks up the speaker's reference made in the utterance of," and you have an equally straightforward explanation for anaphora that combines with Kripke's proposal and covers exactly the same range of examples. (Note that, because of this parallelism, the criticism against Kripke that was based on example-sentences (18)-(19) is equally applicable against Chastain; it can be evaded, though, by recasting Chastain's criteria for reference in such a way that an indefinite counts as referring whenever the utterance it occurs in has the effect of raising a particular object's salience in the sense of Lewis. See above for details.)

So if we stick with Chastain's own reasons – or with the reasons that any other advocate of the ambiguity hypothesis has offered in the philosophical literature I am familiar with — the ambiguity hypothesis buys us nothing, and ceteris paribus, we would of course rank it below a theory that gets around positing an ambiguity. Moreover, there are cases where Chastain's predictions fit poorly with judgments in terms of our pretheoretical notions of truth and falsity. Imagine a situation in which I utter (33):

(33)   There is a mosquito in here. I can hear it buzzing.

The fact is there is no mosquito in here, and the buzzing is just an acoustic illusion. Intuitions concerning truth and falsity may be rather uncertain with respect to the second half of this utterance (i.e., the utterance of the second sentence), but they are firm with respect to the *first* half: My utterance of "There is a mosquito in here" was definitely false, given the facts described. (Recall that this judgment is in line with the predictions of

Kripke's analysis.) Chastain would seem to be committed to saying that this utterance of "There is a mosquito in here" involves an unsuccessful attempt at reference and is consequently neither true nor false. It could be false only if read existentially, but this is against the speaker's intentions, which are clearly signaled by the subsequent use of the anaphoric pronoun. (I suppose that Chastain would not resort to saying that an utterance may have an existential reading even though the speaker means it referentially and is understood to mean it so by the audience.) Short of accepting that the technical distinction between falsity and truthvaluelessness simply fails to match the corresponding intuitive distinction, I see no way for Chastain to respond to this challenge.

But let us not dismiss the ambiguity hypothesis prematurely as a relic from a pre-Gricean past, whose motivation has been undermined as it became conceivable that anaphora might be coreference in a pragmatic, but not a semantic, sense. It might have other advantages, unbeknownst to some of its proponents.

It turns out that a prima facie weakness of Chastain's position, i.e., that he does not attempt to explain the ambiguity he posits in terms of other assumptions that are less ad hoc, serves to render him immune to an objection I have raised against Kripke and Lewis: I noted earlier that anaphoric possibilities are sometimes affected when an indefinite is eliminated in favor of a logically equivalent paraphrase (recall examples (21) through (23) in section 1.3 above). This is problematic for Kripke, who takes the presence of speaker's reference to be a sufficient precondition for a subsequent anaphoric pronoun, and who must therefore concede that it is not only the speaker's intentions and other pragmatic factors which determine whether an utterance with a given semantic interpretation can be accompanied by a speaker's reference, but that the wording of the utterance enters as an additional independent factor. Chastain has no problem here: From his point of view, it is only to be expected that if you paraphrase away a referential expression, such as the referential indefinite "one (of the marbles)" in (21a), you will no longer be able to find a coreferential antecedent for a subsequent pronoun.

Where else could we look for possible arguments in favor of a referential-existential-ambiguity in indefinites? A promising source would seem to be the linguistic literature on the feature ±specific.[13] With indefinites, +specific marks the referential reading, -specific the existential reading. (This is how the feature was interpreted wherever it was interpreted at all, as far as I know.)

---

[13] Especially Baker (1966), Karttunen (1968a), Stockwell *et al.* (1973, Ch. 3).

However, hardly any of the arguments that, at one time or another, were used to motivate the specific/non-specific distinction hold up to closer scrutiny. One reason for this disappointing state of affairs seems to be that linguists were not aware from the outset that many of the contrasts they attributed to specificity vs. non-specificity had perfectly natural explanations in terms of scope-ambiguities. To mention just one example, it was proposed that "only [-SPEC] articles are candidates for undergoing *some-any* suppletion" (Stockwell *et al.*, 1973, 92), but there was no mention of the alternative hypothesis that being in the scope of negation is a sufficient condition for some-any suppletion.[14] If one disregards this and other contrasts that are equally well or better explained as a result of scope constellations, the linguistic literature offers practically no reason to posit a ±specific-ambiguity at all. (There are exceptions one of which I will turn to shortly.) Note that, unlike the philosophers Strawson and Chastain, linguists never really tried to hold specificity responsible for an indefinite's ability to serve as the antecedent to an anaphoric pronoun. Karttunen (1968a; 1976), for instance, devotes much attention to anaphoric relations like the one in (34), where the indefinite antecedent is nonspecific:

(34)   I wish she had a car. She would give me a ride in it.

His conclusion is that specificity is sufficient, but not necessary, for the ability to serve as an anaphoric antecedent.

To my knowledge, there is only one family of linguistic phenomena that are better explained in terms of a referential/nonreferential (or specific/nonspecific) ambiguity than in terms of scope-ambiguities. These phe-

---

[14] It is of course not self-evident that what can be analyzed as a scope ambiguity *should* be so analyzed. Natural language indefinites *could*, in principle, be subject to a two-way ambiguity specific vs. non-specific, rather than be scoped elements. However, there are plenty of arguments to be found in the literature which show a specific-nonspecific-ambiguity to be by itself insufficient to account for the full range of readings that sentences with indefinites exhibit. See Stockwell *et al.* (1973) for examples with an indefinite in the environment of more than one (other) scoped element and the unexpected (from the point of view of a two-way ambiguity hypothesis) multiplicity of readings in such examples. While Stockwell *et al.* (1973) point out that a two-way ambiguity in indefinites cannot account for these examples, they fail to observe that a theory according to which indefinites are unambiguous, but admit of variable scope, could have predicted not only these examples, but all other ambiguities that they cite as manifestations of the specific-nonspecific contrast as well. Only later was this conclusion put forward. See Karttunen (1976), who proposes that the notions of specificity and nonspecificity should, if not discarded altogether, be understood as relative to a given scoped element; in other words, that to be specific should be taken to mean precisely to have wider scope than...

nomena and their relevance for the hypothesis that indefinites are ambiguous were discussed most recently and most sophisticatedly by Fodor and Sag (1982). The point of their argument is that the ambiguity hypothesis is compatible with a simpler and more explanatory theory of scope islands than the alternative hypothesis that all indefinites are existential quantifiers. It is based on data like the following:[15]

(35) John overheard the rumor that each of my students had been called before the dean.

(36) John overheard the rumor that a student of mine had been called before the dean.

As (35) illustrates, a "that"-complement to a noun like "rumor" is a scope island which not even the quantifier "each," which tends to prefer wide scope, can escape: (35) does not have a reading according to which for each student x, John overheard the rumor that x was called before the dean. We would therefore expect that, if "a student of mine" is an existential quantifier, then its scope in (36) should likewise be confined to the "that"-complement. But this is not borne out by the facts: (36) seems quite naturally understood in such a way as to be judged true in a situation where Mary is a student of mine and John overheard the rumor that Mary has been called before the dean.

Note that this situation is one which verifies (36) under a reading where "a student of mine" is given widest scope, but does not verify it under the reading with "a student of mine" taking narrow scope. Yet the latter reading is predicted to be the only one possible if indefinites are existential quantifiers and respect scope islands like other quantifiers do. So the price of defending an analysis of indefinites as unambiguously quantificational is that indefinites must be considered exceptional in their ability to violate the principles which account for scope islands. On the other hand, no such price has to be paid by the proponent of the ambiguity hypothesis: If indefinites have a referential reading in addition to a quantificational reading, then (36) may be interpreted in such a way that "a student of mine" refers to Mary. Under this interpretation, it will be true in the situation described above, i.e., where Mary is a student of mine and John overheard a rumor that Mary has been called before the dean. There is of course no violation of scope islands involved in such an interpretation, since referring expressions have no scope at all.

Other examples (Fodor and Sag give a variety of them) lead to the same conclusion: As far as simplicity and generality of scope-restricting principles

---

[15] Examples from Fodor and Sag (1982).

go, the ambiguity hypothesis compares favorably with the hypothesis that indefinites are unambiguously quantificational.

1.6 Summary

Before I restate the results of the preceding five sections, let me say a few words about a reading of indefinites that was obviously not under consideration in any of the discussion so far: "generic" indefinites, as they are found in the following sentences, under their most natural readings.

(37)   A dog has four legs.

(38)   An up-to-date encyclopedia is expensive.

All of the authors I have cited would consider the generic interpretation as a separate reading of the indefinite article, distinct from the existential-quantifier reading as well as from the referential reading, if the latter exists. None of the reasoning that I am about to summarize was intended to apply to generic indefinites. This said, let us continue to disregard them.

We have considered Russell's analysis of the indefinite as a means of existential quantification, as well as an opposing view, which holds indefinites to be ambiguous between a quantificational and a referential use. The Russellian position was seen to face a prima facie problem in the fact that indefinites can serve as the antecedents of anaphoric pronouns. We looked at three approaches to this problem within the limits of the Russellian position, and found some (possibly minor) defect in each of them:

Geach's proposal to analyze the problematic anaphoric pronouns as bound variables forces him to assign truth-conditions primarily to discourses (i .e., utterances of sequences of sentences), not to monosentential utterances, and thereby falls short of reconstructing our intuitive notion of truth and falsity, which often does apply to each uttered sentence separately. Furthermore, Geach's analysis seems to force him to assume that indefinites as existential quantifiers show an exceptional behavior' with respect to the scope island constraints that otherwise govern the interpretation of quantifiers in natural language. (Besides, Geach's analysis did not seem to carryover to texts with a plural indefinite followed by an anaphoric plural pronoun without practicing inadequately weak truth conditions.)

The Grice-Kripke-Lewis family of proposals has no problems with any of the above challenges, but because of their commitment to a purely pragmatic account of the relation between antecedent and anaphor, they leave unexplained how two structurally different expressions of the same meaning could ever differ in their capacity of making an object available for subsequent anaphoric reference.

Then there was Evans' proposal to analyze the problematic anaphoric pronouns as paraphrases of certain definite descriptions that can be constructed from the antecedent and the sentence it occurs in in a systematic way. This proposal avoids both the objections against Geach and against the Griceans. However, it has little substance if the proper analysis of (anaphoric) definite descriptions remains itself unclear, and it is subject to criticism if coupled with the standard analysis of definite descriptions that Evans presupposes. The problem is that definite descriptions under the standard analysis carry an implication that their referent satisfies the descriptive predicate uniquely, and it is not true, or at the very least doubtful, that this uniqueness is required for an anaphoric use of a pronoun or definite description to be felicitous.

As for the anti-Russellian position, according to which indefinites have a reading as referring expressions, it turned out to be immune to any of the objections that have been raised against the Russellian approaches. On the other hand, it is at a prima facie disadvantage by positing an irreducible ambiguity where its competitors purport to predict all uses of indefinites from a unique lexical meaning by independent principles. Moreover, its predictions contradict pretheoretical semantic judgments on the question of whether certain utterances with indefinites are false or else truthvalueless. Still, the ambiguity hypothesis has the advantage of being compatible with a simpler and more explanatory theory of scope island constraints than any version of the view that indefinites are unambiguously quantificational.

If we were to choose among these four alternatives at this point, we would not be able to settle on any one without some misgivings. And we have not even yet looked at a phenomenon that presents all four of them with difficulties that outweigh most of the weaknesses noted so far. The phenomenon in question is the type of anaphoric relation that indefinite antecedents and pronouns (or definite descriptions) exhibit in the so-called "donkey sentences." This is the subject of section 2.

2 Problems with Donkey Sentences

Donkey sentences are sentences that contain an indefinite NP which is inside an if-clause or relative clause, and a pronoun which is outside that if-clause or relative clause, but is related anaphorically to the indefinite NP. The following are examples:

(1)   If $\left\{ \begin{array}{c} \text{someone} \\ \text{anyone} \end{array} \right\}$ is in Athens, he is not in Rhodes.

(2)   If a man owns a donkey, he beats it.

(3)   Every man who owns a donkey beats it.

Since such sentences (in particular (1)) were first discussed by the Stoic philosopher Chrysippos, some people call them Chrysippos-sentences. The more popular donkey examples, originally from the medieval literature, were revived by Geach (1962).

What are the truth conditions of (1) through (3)? The natural and most commonly held assumption is that they amount to the truth conditions of the following logical formulas:

(1′) $\forall x(x$ is in Athens $\to \neg(x$ is in Rhodes$))$
(2′) $\forall x \forall y((x$ is a man $\land y$ is a donkey $\land x$ owns y$) \to x$ beats y$)$
(3′) $\forall x \forall y((x$ is a man $\land y$ is a donkey $\land x$ owns y$) \to x$ beats y$)$

We will see below that this assumption has been challenged, but meanwhile, we may adopt it without any harm for the validity of the arguments to be presented.

For the rest of this chapter, it will be important to realize that the indefinites in donkey sentences are not instances of "generic" indefinites in the sense that I am using that term. I am not denying, of course, that our understanding of indefinites in the context of donkey sentences like (1)-(3) in effect amounts to a "generic" understanding – if we choose to so describe the truth conditions (1)-(3) that we observe. What I *am* denying is that the observed truth conditions are a result of disambiguating the relevant indefinites in favor of their generic readings, i.e., in favor of those readings that they exhibit in sentences like (4):

(4)   A donkey is grey.

At first sight, this could be a promising line to take in the analysis of donkey sentences. Just as we analyze (4) as the result of quantifying the generic indefinite "a donkey" into the open sentence "... is grey," producing the reading:

(4′) $\forall x(\text{donkey}(x) \to x$ is grey$)$,

we might analyze (2) as a result of first quantifying into the doubly open sentence "If ... owns ... he beats it," the generic indefinite "a donkey" to form "If ... owns a donkey he beats it," and then quantifying into that another generic indefinite, "a man," binding the pronouns in the process. In perfect analogy to (4′), the resulting reading would be:

$$\forall y(\text{man}(y) \to \forall x(\text{donkey}(x) \to (\text{if } y \text{ owns } x, y \text{ beats } x)))$$

which is equivalent to the desired (2′).

There are some reasons against an analysis like that. First, not all indefinites have generic readings. "Someone," for instance, does not (unless it is part of a larger NP, e.g., modified by a relative clause), as the impossibility of a generic reading for (5) shows:

(5)   Someone is grey.

However, "someone" *can* act just like "a donkey" in donkey sentences, as (1) shows. Second, there are certain linguistic environments that, for whatever reason, preclude a generic reading for indefinites, most notably the subject-position of "there"-insertion sentences, and also the object position of the verb 'have." Neither of the following sentences can be read as a generic statement about donkeys:

(6)   John has a donkey.
(7)   There is a donkey in the yard.

However, indefinites in these environments do serve as antecedents for donkey anaphora:

(8)   If John has a donkey he beats it.
(9)   If there is a donkey in the yard John will chase it away

I admit that these considerations do not prove beyond doubt that donkey sentences contain no generic indefinites, especially not in the absence of an analysis of generic indefinites that would explain, e.g., why they do not appear in "there"-insertion contexts. For the time being, i.e., for the remainder of this chapter, I will nevertheless assume – as is generally assumed – that donkey sentences and generic indefinites are distinct phenomena.

From what I just said, it should be clear that the following sentences, despite superficial similarity in both form and meaning, are *not* donkey sentences by my terminology:

(10) $\left\{ \begin{array}{c} \text{someone} \\ \text{anyone} \end{array} \right\}$ who is in Athens is not in Rhodes.

(11)   John beats a donkey if it kicks him.

These contain genuine generic indefinites.

2.1   Donkey anaphora as variable binding

As we compare our examples of donkey sentences to the logical formulas that paraphrase them, we are led to hypothesize the following generalization: An

indefinite that occurs inside an if-clause or relative clause gets interpreted as a universal quantifier whose scope extends beyond this clause. Insofar as this generalization is borne out, it falls out from it that an indefinite thus interpreted as a wide-scope universal is able to bind a pronoun outside its own clause, and the donkey sentence readings are thereby accounted for. There is nothing special to be said about the interpretation of the anaphoric pronoun or about the nature of the anaphoric relation between it and its indefinite antecedent: it is just like other bound-variable pronouns. The special assumptions that are required are assumptions about the *indefinite*, i.e., its quantificational force and scope possibilities.

This is basically the approach to donkey sentences that has been taken by those who started from a Geachian analysls of indefinites and pronouns, as sketched in 1.2 above. The major question that the donkey phenomena, viewed from this angle, pose is this: What are the exact conditions under which an indefinite may receive the wide-scope universal interpretation? And what, if anything, explains why these conditions are as they are? The tentative generalization above is a rather insufficient answer: not only does it fail to subsume the two environments of "if-clause" and "relative clause" under a more revealing common property, but its predictions are not even correct. If they were, the indefinite "a donkey" in sentence (12) should also admit for a wide-scope universal reading, which it does not:

(12)   A friend of mine who owns a donkey beats it.

In fact, this generalization was never seriously entertained by anyone. Geach himself refrained from attempting any generalization at all, and it appears to be widely believed that any such attempt is bound to be a rather hopeless enterprise, leading to an unrevealing list of conditions at the very best.

There are two publications, however, which challenge this pessimistic prejudice: Smaby (1979) and Egli (1979). Both authors construct an artificial language which shares with natural languages the peculiarity that it contains "every"-NPs, indefinite NPs, proper names, and pronouns among its NPs, and they specify an algorithm that converts texts in this language into formulas of a logical language with a known semantics. For the purpose of the present discussion, I am identifying their artificial language with English syntactic structure, and taking it for granted that either one of their algorithms could be turned into one that translates sequences of English syntactic structures into formulas of elementary predicate logic.[16]

---

[16] Egli in fact uses ordinary predicate logic for his logical forms. Smaby, on the other hand, uses his own, very different, formalism. I hope I am correct in assuming that the

2.1.1   Egli's proposal

Egli starts from the observation that the relation between a sentence of the form

(13)   If (... someone ...) then (... he ...)

and its logical formalization

(14) $\forall x((... x ...) \rightarrow (... x ...))$

is reminiscent of an equivalence that is valid in predicate logic, i.e.:

(15)   If ϕ and ψ are formulas, and ψ does not contain the variable x free, then the following formula is logically true: $(\exists x \phi \rightarrow \psi) \leftrightarrow \forall x(\phi \rightarrow \psi)$

Of course it is not possible to see in (15) an explanation for why sentences of the form of (13) receive readings equivalent to (14): In the conversion from (13) to (14), a variable-binding into the "then"-clause is created, whereas (15) explicitly prohibits this by restricting the validity of the equivalence to those cases when x does not occur free in ψ. It is indeed easily shown that

(16) $(\exists x Fx \rightarrow Gx) \leftrightarrow \forall x(Fx \rightarrow Gx)$

is *not* a logically true formula. Nevertheless, the analogy between the translatability of (13) into (14) and the logical law in (15) seems to be more than a coincidence. Egli claims it is part of a wider generalization, which he states informally in more or less the following way:[17]

(I)   Whenever there is a theorem of predicate logic according to which (for certain quantifiers Q_1, Q_2 and a certain truth-functional connective $\oplus$) every instance of the schema:

   $(Q_1 x\ \phi \oplus \psi) \leftrightarrow Q_2 x(\phi \oplus \psi)$, where x is not free in ψ,

is logically true, then a natural language sentence of the form:

   $(... Q_1'...) \oplus' (... \text{pronoun} ...)$

receives a reading:

   $Q_2 x((... x ...) \oplus (... x ...))$

---

particular idea of his that I am interested in here is not essentially tied to his choice of formalism.

[17] Egli (1979, 275, (54))

> (where Q_1' is an expression of the natural language which normally means the same as the quantifier Q_1, and $\oplus'$ is the natural language counterpart of the connective $\oplus$).

This is not an exact formulation of the generalization, but intuitively the claim should be clear. One consequence of (I) is that an indefinite inside an if-clause is read as a wide-scope universal quantifier that may bind a pronoun in the then-clause. This consequence falls out under the assumption that an indefinite's primary logical counterpart is the existential quantifier, and "if-then" corresponds to material implication. Further consequences of (I) include: an indefinite in the first conjunct of an "and"-conjunction can be read as an existential quantifier whose scope extends throughout the second conjunct and can bind a pronoun there. This accounts for the Geachian (cf. section 1.2 above) interpretation of

(17)   Somebody left, and he went home.

as

(18) $\exists x(x \text{ left} \wedge x \text{ went home})$.

The relevant logical law in this case involves the schema:

(19) $(\exists x\, \phi \wedge \psi) \leftrightarrow \exists x(\phi \wedge \psi)$, where x is not free in ψ.

Under the natural additional assumption that juxtaposition of sentences in a natural language text often has the force of conjunction as well (even when there is no overt "and"), (I) could easily be extended to account for the reading of (20) as (18).

(20)   Somebody left. He went home.

How about donkey sentences which contain the critical indefinite inside a relative clause, like example (3) above? These are not directly covered by (I) as it stands, since there is no explicit connective in them, at least not in surface structure. This suggests that (I) must really be viewed in the context of an algorithm, i.e., a system of transformation rules that converts surface structures into logical notation via a number of intermediate steps, only one of which (and not normally the first one) is that step by which a narrow-scope existential quantifier is rewritten as a wide-scope universal. In particular, such an algorithm must first convert a sentence like

(3)   Everyone who owns a donkey beats it.

into an intermediate form with an explicit connective:

(21)   ∀x:((x owns a donkey) → (x beats it))

This intermediate form can now be seen as containing a configuration that is in the relevant sense "reminiscent" of the left half of the logical equivalence in (15), and is therefore predicted by (I) to be further convertible into:

(22)   ∀x:∀y:((x owns y ∧ donkey(y)) → (x beats y))

I am glossing over many details here, but let me emphasize that Egli's paper contains a fully spelled-out system of rules which covers every step we would need here.

    Note that (I), understood in the intended fashion, also accounts for the absence of a wide-scope universal reading in (12) (see above):

(12)   A friend of mine who owns a donkey beats it.

The relevant intermediate representation of this sentence that is comparable to (21) will be (23):

(23)   ∃x:((x is a friend of mine ∧ x owns a donkey) ∧ (x beats it))

This does not contain the connective → and therefore bears no similarity to the left half of the equivalence (15), or of any other equivalence whose right half contains a universal quantifier with widest scope. Instead, (23) recalls (19) and is therefore predicted to amount to (24), which in fact it does.

(24)   ∃x:∃y:(x is a friend of mine ∧ x owns y ∧ y is a donkey ∧ x beats y)

Before I come to considerations bearing on an evaluation of Egli's proposal, let me draw attention to the assumptions it implies with respect to the organization of grammar: Semantic interpretations are not assigned to syntactic representations directly, but to what we may refer to as "logical forms," another level of representation that is related to syntactic representation via a "construal component," i.e., the "rules of construal,"[18] a set of transformational rules. In this respect, Egli works with a model of grammar that corresponds to the Revised Extended Standard Theory (hence

---

[18] I am using the terms "rule of construal" (and "construal component") to refer to what is usually called "interpretive rules" (and the "interpretive component") in the REST literature. Only a subset of the "interpretive rules" are normally called "construal" rules, whereas I extend that term to all of them. My reason is that I want to reserve the term "interpretation" (and accordingly, "interpretive component") for semantic interpretation in the strict sense (e.g., model-theoretic interpretation). So interpretation, in my sense, affects the output of the construal component.

my description of it in terms of REST terminology). However, among his construal rules are not just the quantifier-adjunction and pronoun-reindexing rules that one finds elsewhere; his logical forms are accordingly more abstract, i.e., more removed from the syntactic representations than other peoples' logical forms. Egli's construal component crucially includes rules that spell out connectives (e.g., to get us from (3) to (21) above), and of course the exportation rules that fall under (I).

2.1.2 Evaluation of Egli's proposal

Aside from certain specific difficulties that I will point out below, Egli's proposal is expected to be subject to the objections against Geachian approaches in general (see section 1.2 above). To repeat those briefly:

(a) Truth conditions are only assigned to complete texts, not to smaller units.

(b) When extended to examples involving plural NPs, the analysis predicts insufficiently weak truth conditions in certain cases.

(c) It seems that complicating qualifications must be added to the principles which account for structural constraints on quantifier scope.

I would like to reconsider the third of these objections, i.e., (c), in the light of Egli's proposal, because I think the objection can be answered now.

Recall the point of the objection: As long as we restrict our attention to non-existential quantifiers, e.g., "every"-NPs and "no"-NPs, we find that the position in which the quantifier occurs in syntactic structure limits in certain ways the scope which it can maximally take. (E.g., matrix clauses, relative clauses, if-clauses, and other kinds of subordinate clauses are scope islands.) The simplest theory would be one which predicts that all quantifying NPs observe the same scope islands, including indefinites. Whereas certain alternatives to Geach's analysis of anaphora appear to be compatible with that simplest theory (or nearly so), Geachians are forced to regard indefinites as exceptions. According to Geach's, and likewise Egli's, analysis of (25) through (27), for instance, indefinites can take scope beyond a matrix sentence, relative clause, or if-clause in which they occur.

(25)   John bought a sandwich. He ate it.
(26)   A man who had bought a sandwich ate it.
(27)   If you buy a sandwich I will eat it.

We should reject ceteris paribus a Geachian analysis in favor of one in which indefinites are not exceptions.

Here is how we might defend Egli 's proposal against this objection: Suppose that the scope-island phenomena we observe with non-existential quantifying NPs are a result of constraints on the application of certain construal rules, as proposed by May (1977). For concreteness, let us assume with May that they come about because quantifier adjunction, or rather Egli's counterpart of May's quantifier adjunction (i.e., his rule R_1) obeys the subjacency condition. (R_1 actually does several things at once: it adjoins a quantifier to S, separates off the common noun which restricts the quantifier, and inserts a suitable connective. So it accomplishes more than May's rule, but the difference is irrelevant here.) It turns out that this assumption can be maintained without the slightest qualification, and yet the indefinites in (25) through (27) will end up with the extrawide scopes that Geach attributes to them.

The point considered here is that two entirely distinct construal rules participate in deriving the logical forms of those texts: R_1, alias quantifier adjunction, and one of the exportation rules that fall under (I). For the former to obey subjacency in no way prevents the latter from exporting out of the scope island anyway. The logical form of (27), for example, is derived in two steps: First, R_1 applies, well within the limits of subjacency, to adjoin the existential quantifier in "a sandwich" to the antecedent clause of the conditional; to the result of that, an exportation rule applies which both turns the quantifier from existential to universal and attaches it all the way up. This derivation does not violate May's assumption that quantifier adjunction obeys subjacency.

To conclude, it is not true, as the superficially convincing objection (c) would have it, that a Geachian analysis of anaphora necessitates ad hoc qualifications to the principles that account for scope island constraints. Still, objections (a) and (b) remain in force against Egli's proposal, and to these I want to add the following three crucial considerations.

Before I try to assess the descriptive adequacy of Egli's system, I want to raise a theoretical point. What is the status of the generalization expressed in (I)? (I) specifies a construal rule schema, as well as a constraint on which of the rules instantiating that schema are realized (alternatively, it might be seen as a constraint on rule application). The constraint is of a curious nature: it singles out those rules that bear a certain formal similarity to a valid equivalence in logic. This means that a non-structural property of formulas, i.e., the property of logical validity, must be appealed to in the construal component of the grammar, and that it is impossible, after all, to conceive of the construal component as consisting of strictly structure dependent rules. Now of course there is no a priori reason to prefer a model of grammar which admits only construal rules whose applicability

depends on structural properties alone. But if we have the choice between Egli's theory, which employs construal rules of a relatively great variety of types, as well as of relatively high individual complexity, and a competing theory whose construal rules are simpler and of a more narrowly defined rule type, we would, ceteris paribus, find the latter more interesting. This should be borne in mind when we come to discuss the pronouns-as-definite-descriptions approach (see section 2.3), and also when I present my own approach, as both of these alternative approaches employ a much more constrained construal component.

It is of course possible to specify the rules that fall under (I) individually in purely structural terms and simply list them one by one. This makes it unnecessary to appeal to the generalization (I) in specifying the construal component. (Egli actually proceeds in this way, treating (I) merely as a heuristical generalization, with no status as a part of the grammar at all.) But this only amounts to denying that the generalization is significant at all and leaves us without any answer to the question we started with, viz., why is it precisely that list of unrelated conditions which gives rise to a universal wide-scope reading of the indefinite?

If we take a closer look at the predictions of Egli's proposal, however, we may find enough empirical inadequacies in generalization (I) to convince us that its theoretical implications need not preoccupy us any further. One of the flaws of (I) is that it predicts a symmetry in the behavior of existential and universal quantifiers that is not in fact observed. Consider the following equivalence, which is as much a logical truth as the one in (15):

(28) $(\forall x \, \phi \to \psi) \leftrightarrow \exists x(\phi \to \psi)$, where x is not free in ψ.

If we go by (I), we should therefore expect a sentence like (29) to admit a reading like (30).

(29)   If everyone is in Athens, he is not in Rhodes.
(30) $\exists x(x$ is in Athens $\to \neg(x$ is in Rhodes$))$

No such reading is ever possible. Egli, who notes this overgeneration of readings that his generalization (I) implies, says that it is harmless, since sentences like (29) are "unacceptable syntactically," and the relevant syntactic constraints "can be formulated independently by something very much like surface structure constraints" (1979, 276). He does not spell this out any further, but I take it that he is referring to constraints on coindexing, rather than "syntactic" constraints in the narrowest sense. (After all, (29) is perfectly grammatical if the "he" is read deictically.) Those constraints on coindexing must be sensitive to the particular quantifier involved, so that

they can differentiate between the bad coindexing of "he" with "everyone" in (29), and the-good one of "he" with "someone" (or "anyone") in (1).

There seems to be a problem with using constraints on coindexing to rule out readings like (30) for (29), at least if those constraints are to apply to indexed syntactic structures rather than a more abstract level, which has already undergone some of the construal rules. Consider the following sentence:

(31)   If everyone likes a donkey, it is happy.

The if-clause of (31) is a potentially ambiguous sentence: "everyone likes a donkey" can be interpreted either as "$\forall x(\text{person}(x) \rightarrow \exists y(\text{donkey } (y) \wedge \text{like } (x,y)))$" or as "$\exists y(\text{donkey}(y) \wedge \forall x(\text{person}(x) \rightarrow \text{like}(x,y)))$." But if the pronoun "it" in (31) is to be construed with "a donkey" as its antecedent, only one order of scopes is possible in the if-clause, viz., the second-order, "$\exists y ... \forall ...$" This fact is a problem for Egli for the following reason: His rules of construal assign a reading with "it" anaphoric to "a donkey" with either of the two scope-orders in the if-clause, the two readings being:

(31)   a. $\forall y((\text{donkey}(y) \wedge \forall x(\text{person}(x) \rightarrow \text{like}(x,y))) \rightarrow \text{happy}(y))$
       b. $\exists x \forall y((\text{person}(x) \rightarrow (\text{donkey}(y) \wedge \text{like}(x,y))) \rightarrow \text{happy}(y))$[19]

(31b) does not of course correspond to an actual reading of (31) and must somehow be excluded. Like (30), its derivation crucially involves exportation of a universal quantifier from the antecedent of a conditional by turning it into an existential quantifier. But unlike the case of (29), it is hardly possible here to appeal to a constraint on well-formed indexings of (31): After all, the indexing that gives rise to (31b) is identical to the indexing that underlies the good reading (31a).

Another problem for Egli's proposal is raised by the existence of quantifiers that cannot be analyzed in terms of a truth-functional connective. A well-known example is "most." There can be no quantifier M and truth-functional connective $\oplus$ such that the truth conditions for

$$Mx(Fx \oplus Gx)$$

would capture the intuitive truth conditions for

Most Fs are G,

---

[19] (31b) is derived by Egli's rules R.8a (first line) and R.8b (third line), see p.278. For (31b) to come out true, it takes nothing more than there being somebody who does not like donkeys.

for arbitrary predicates F and G.[20] The logical counterpart of "most" therefore has to be an operator that takes *two* sentential arguments rather than one, i.e., "Most Fs are Gs" must be formalized as something of the form:

$$Mx(Fx, Gx)$$

Suppose we are dealing with an enriched version of elementary logic that contains such "two-place quantifiers" and take that to be the language into which the construal rules translate. Then we will have no problems providing Logical Forms for sentences like (32):

(32)   Most men who own a donkey are wealthy.
(32′) $Mx((man(x) \land \exists y(donkey(y) \land own(x,y))), wealthy(x))$

But what about (33)?

(33)   Most men who own a donkey beat it.

By the available construal rules, we will presumably get to an intermediate representation that looks like this:

(33′) $Mx(\exists y(man(x) \land donkey(y) \land own(x,y)), beat(x, it))$

But then, what further rule would permit us to export the quantifier binding y in such a way that the pronoun "it" comes to be in its scope and can be construed as a bound variable y? Egli's generalization (I) determines no such rule at all.

   But we should hesitate to use this as an objection against Egli, because sentences like (33) have in fact rather nebulous truth conditions. Although one has a strong feeling that they do permit for a reading in which the pronoun is anaphorically linked to the indefinite in the usual donkey-sentence manner, it is not obvious what exactly this reading amounts to. My own intuitions vacillate between the following two paraphrases:

(33)   a.   Most men who own a donkey beat every donkey they own.
       b.   Most men who own a donkey beat one of the donkeys they own.

One might want to account for this vagueness by providing construal rules for either possibility. They would basically look like this:

(R$_1$)   Rewrite $Qx(\exists yS_1, S_2)$
          as: $Qx(\exists yS_1, \exists y(S_1 \land S_2))$

(R$_2$)   Rewrite   Qx($\exists$yS$_1$, S$_2$)
      as: Qx($\exists$yS$_1$, $\forall$y(S$_1$ → S$_2$))

On the other hand, one might take the uncertainty of intuitions with respect to such sentences as a symptom of the fact that grammar simply provides no construal rules to apply to them at all, and that whatever readings they may receive in practice are a result of pragmatic strategies of trying to make some sense or other of any given utterance. The latter position would be compatible with Egli's approach: Generalization (I) could be seen as covering just those cases where an interpretation is grammatically determined.

However, it seems to me that there are examples of a similarly problematic sort whose readings are much less vague. Consider a sentence that contains the adverbial quantifier "in most cases":

(34)   If a table has lasted for 50 years, then in most cases it will last for another 50 years.

(34) means that most tables that have lasted for 50 years will last for another 50 years. This interpretation is beyond the capacities of Egli's theory. Note that adverbial, as opposed to nominal, quantifiers are not in themselves difficult for Egli's system to accommodate, as long as only standard quantifiers, as in (35) and (36), are involved.

(35)   If a table has lasted for 50 years, then it will always (= in all cases) last for another 50 years.

(36)   If a table has lasted for 50 years, then it will sometimes (= in some cases) last for another 50 years.

In order to deal with (35) and (36), one has to avoid translating "if-then" into material implication regardless of context. As Lewis (1975) points out, if-clauses that co-occur with adverbs of quantification serve to restrict the range of the adverbial quantifier and do not carry in themselves the force of any particular connective. For Egli's purposes, roughly the following construal rule would be needed:

(R$_3$)   Rewrite   If S$_1$, Q$_{adv}$ S$_2$
      as: (S$_1$ ⊕ S$_2$)
      where ⊕ = ∧ if Q$_{adv}$ = $\exists$, and ⊕ = → if Q$_{adv}$ = $\forall$.

Here the connective gets introduced depending on the nature of the quantifier. By applying (R$_3$) to (35) and (36), representations of the following form would be created:

(35′) $\exists$x(table(x) ∧ ... (x)) → ...(pronoun)....

(36′) $\exists x(table(x) \wedge ... (x)) \wedge ...(pronoun)....$

These are suitable inputs to the quantifier-exportation rules predicted by (I), and the results are adequate:

(35″) $\forall x((table(x) \wedge ... (x)) \rightarrow ... (x))$
(36″) $\exists x(table(x) \wedge ... (x) \wedge ... (x))$

But with (34), no rule like (R_3) will work because the adverbial quantifier "in most cases," just like its determiner-counterpart "most," fails to determine a connective. One could at best get from a sentence of the form

> If S_1, in-most-cases S_2

to a representation with "in-most-cases" as an operator with two sentential arguments: In the case of (34), S_1 would be headed by an existential quantifier, i.e., it would be of the form $\exists x S_3$. If we want to arrive at the intuitively correct reading, we would need another construal rule which, in effect, converts

> in-most-cases $(\exists x S_3, S_2)$

into

> $Mx(S_3, S_2)$

While such a rule could of course be added ad hoc to Egli's system, it is certainly not an instance of his generalization (I).

To sum up, I have pointed out three more or less unattractive consequences of Egli's proposal. First, Egli must give up the assumption that construal rules make reference only to structural properties of the representations they apply to. Second, he predicts that universal quantifiers should under certain conditions turn into wide-scope existential quantifiers, which they do not. Third, he cannot account for the readings that indefinites exhibit when they occur in clauses restricting adverbial quantifiers, unless he makes ad hoc additions to his system of construal rules.

2.1.3   Smaby's proposal

Smaby (1979) addresses numerous issues that are of no interest to our present concern, which is the treatment of donkey sentences, and by focusing on his contribution to an explanation of the latter, I must abstract away from so much of his paper that the ideas I am attributing to Smaby may not be recognized easily as his own. Peripheral differences aside, Smaby's theory is

much like Egli's, except that instead of the entire group of construal rules that fall under Egli's generalization (I), Smaby has what may be reformulated as a single rule as follows:

(R-Smaby)   Rewrite $\exists x S_1 \wedge (_{S_2} \dots \text{pronoun} \dots)$
            as $\exists x(S_1 \wedge (_{S_2} \dots x \dots))$

This says that an existential quantifier in the left conjunct of a conjunction can be exported to take scope over the entire conjunction. In Egli 's system, this is one special case of quantifier exportation among others. How does Smaby manage to get away with it alone? He ensures that, by setting up the rest of his construal rules accordingly, material implication is not represented as a simple connective, but as a combination of negation and conjunction. In particular, sentences of the form:

> If S_1, S_2
> or: X – every F – Y

are first rewritten as

> $\neg(S_1 \wedge \neg S_2)$
> and: $\forall x \neg(Fx \wedge \neg(X\text{-}x\text{-}Y))$

respectively. This corresponds to adequate truth conditions in those cases where no donkey-pronominalization is involved, because of the logical equivalence between "$p \rightarrow q$" and "$\neg(p \wedge \neg q)$." Furthermore, (R-Smaby) can be applied to such representations and the resulting interpretations happen to be correct for donkey sentences. Sentence (1), for instance, is related to its Logical Form via the following steps:

(1)   If someone is in Athens, he is not in Rhodes.
(1′) $\neg(\exists x \text{ in-Athens}(x) \wedge \neg\neg(\text{in-Rhodes}(he)))$
                 $\Downarrow$ by (R-Smaby)
(1″) $\neg\exists x(\text{in-Athens}(x) \wedge \neg\neg(\text{in-Rhodes}(x)))$

(1″) is logically equivalent to:

> $\forall x(\text{in-Athens}(x) \rightarrow \neg(\text{in-Rhodes}(x)))$

It is easily verified that a sentence like (3), "Every man who owns a donkey beats it," likewise receives an adequate derivation.

In comparison with Egli's proposal, Smaby's has an advantage: His only exportation rule exports only originally existential quantifiers, and no further provisions are needed to avoid assigning a reading like (30) to (29), "If

everyone is in Athens, he is not in Rhodes." The disambiguating effect of anaphora in (31) also falls out. It should be noted however, in Egli's defense, that there are certain types of examples that do seem to require exportation of a universal quantifier in order to be interpreted adequately:

(37) Either John doesn't have a dog, or it is hiding.

In Egli's system, the derivation of (37) proceeds via the following stages:

(37′) ¬∃x(dog-of(x,John)) ∨ hide(it)
(37″) ∀x¬(dog-of(x, John)) ∨ hide(it)
(37‴) ∀x(¬dog-of(x,John) ∨ hide(x))

The step from (37′) to (37″) involves another rule in Egli's system that I have not yet mentioned. It prepares the way for the crucial step, the one from (37″) to (37‴), which is an instance of the schema provided by generalization (I). The corresponding valid equivalence is:

$$(\forall x\ \phi \lor \psi) \leftrightarrow \forall x(\phi \lor \psi), \text{ where x is not free in } \psi.$$

The point of the example is that it is a universal quantifier that is subject to exportation here. For Smaby, the only way of handling this example would be to add a rule that rewrites.

$$\neg S_1 \lor S_2$$
as: $\quad \neg(S_1 \land \neg S_2).$

Unlike his rule that analyzes "if-then" as a negated conjunction, this additional rule does not simply decompose a lexical unit, but manipulates a non-local configuration of negation and disjunction.

Another type of example, which seems to provide even more direct evidence for Egli's implicit claim that universal, as well as existential, quantifiers can be subject to exportation across connectives, is (38):[21]

(38) Everyone must turn in the homeworks, or he flunks the course.

Contrary to the generalization that Smaby, as well as many others, have taken for granted, the scope of "everyone" in (38) is not confined to the left half of the disjunction. Notice, however, that this might equally well be a case of quantifier adjunction applying in exceptional violation of scope-island constraints, rather than a case of exportation of ∀ across a connective licensed by (I). The same is true of examples like (39), where we seem to accept a reading in which "it" is bound by the universal quantifier in the previous sentence:

---

[21] (38), or a similar example, was pointed out to me by Urs Egli in personal communication.

(39)   Every copy must meet the specifications in the handbook. *It* must be on 50% rag bond paper, and *it* has to include the title page and the vita.

So we must conclude that neither Egli, who treats existential and universal NPs as fully alike w.r.t. exportation rules, nor Smaby, who builds an asymmetry between the two into his system, can handle all the facts. If Egli is on the right track, then it remains to be explained why we find no cases of "every"-NPs becoming wide-scope existentials under exportation, and why universal interpretations with scope over a whole text or across a coordinating conjunction are not generally available. From Smaby's point of view, on the other hand, examples like (37) through (39) are the exceptional cases yet to be accounted for. I tend towards the latter view myself, but will leave the exceptions mentioned unaccounted for in this dissertation.

Smaby's proposal derives its attractiveness mainly from the fact that it subsumes all environments in which narrow-scope existential turns into wide-scope universal under the structural description of a single construal rule, and it does so without admitting into the construal component anything but purely structure-dependent rules. However, it is subject to the same objection as Egli's proposal when confronted with the behavior of indefinites in if-clauses that modify non-standard adverbial quantifiers. Consider again (34), "If a table has lasted for 50 years, then in most cases it will last for another 50 years." It is clear from the above discussion of this sentence that it cannot be represented in a form which has the connective $\wedge$ conjoining the existentially quantified to the pronoun-containing clause, and that its intended reading cannot be provided by the rule (R-Smaby).

2.2   *Donkey sentences and pragmatic accounts of anaphora*

I presented in section 1.3 an approach (or a family of approaches) to anaphora which assimilates the anaphoric use of a pronoun to its deictic use, the only difference being that a deictic pronoun refers by virtue of the nonlinguistic circumstances that make its referent salient, whereas an anaphoric pronoun refers to something whose salience is due to the content of a preceding utterance. This approach was found to explain with some plausibility why non-referring expressions such as indefinites can set the stage for subsequent anaphoric reference. Our present concern is whether the same type of explanation is possible for the anaphora relations in donkey sentences. Quite clearly, the answer is "no." What Grice, Kripke, and Lewis were trying to explain was: How can a pronoun receive a well-determined reference from the utterance of a non-referring antecedent? But the pronouns in donkey sentences do not refer at all, and therefore require that an entirely different

question be asked in the first place. Their efforts at trying to understand anaphoric *reference* simply contribute nothing to the understanding of donkey anaphora. If we adopt, for whatever reason, a Gricean account of the anaphora relation in (1), then we must additionally provide for a completely different kind of anaphora in terms of which to analyze (2).

(1) John owns a donkey. He beats it.
(2) If John owns a donkey, he beats it.

2.3 Donkey anaphora and disguised definite descriptions

We are now going to reexamine Evans' notion of an E-type pronoun, introduced above in section 1.4, in light of the donkey-sentence phenomena. Evans actually cites some donkey sentences in illustrating the E-type use of pronouns, e.g. (1980, 342):

(1) If a man enters this room, he will trip the switch.
(2) If there is a man in the garden, John will tell him to leave.

I will therefore attribute to him the contention that the puzzle that is traditionally associated with such sentences is resolved when the crucial pronouns are analyzed as E-type. However, there is a brief remark in a footnote that suggests that this is not his claim:

> ... the interpretation of *a*-expressions is unclear, and we may be forced to recognize that they are sometimes used as equivalent to *any* ... (1980, 343, n.5)

Evans further remarks that he considers "any" a wide-scope-seeking universal quantifier and cites the example (1980, 344):

(3) If any man loves Mozart, he admires Bach.

It thus appears as though he considers some donkey sentences (e.g., (1) and (2)) as instances of E-type pronominalization, and others (e.g., (3) and that variant of (3) which has "a" in place of "any") as instances of bound-variable pronominalization. The latter, to which he pays no further attention, would presumably require him to supplement his account of the former with an account of the distribution of universal "any" and its "a"-variant. It is unclear what the latter would look like, and also, how the line would be drawn between the data that fall under each analysis, and why they could not all be covered by one or the other. So if this is indeed what Evans would want to propose, then his proposal is far too undeveloped for us to discuss here. The proposal I am going to address instead is a different one, i.e., that

donkey anaphora is always to be analyzed as E-type pronominalization. I will pretend that this is in fact Evans' claim; if he did not mean to endorse it, it is still a claim worthy of careful consideration. Along with it, I will consider proposals by Cooper (1979) and Parsons (1978) that embody much the same claim.

An E-type pronoun, as you recall, is interpreted as equivalent to a certain definite description which can be recovered from the E-type pronoun's antecedent and the clause in which that antecedent occurs. For example, if the following donkey sentences are analyzed as containing E-type pronouns, then they are predicted to paraphrase the sentences below:

(1) If a man enters the room, he will trip the switch.
   a. If a man enters the room, the man who enters the room will trip the switch.

(4) If someone is in Athens, he is not in Rhodes.
   a. If someone is in Athens, the one who is in Athens is not in Rhodes.

(5) Every man who owns a donkey beats it.
   a. Every man who owns a donkey beats the donkey he owns.

I will not take issue with the claim that the original sentences and their (a)-variants are equivalent. Rather, I will consider the stronger claim that both they and the (a)-variants are properly analyzed in terms of the standard (Russellian) semantics for definite descriptions. In particular: I will focus on the uniqueness implication (or entailment, or presupposition) that this standard semantics predicts, and try to determine to what extent intuitive judgments confirm this prediction. There is some relevant discussion in Evans' writing, but for ease of exposition, I will first address what Parsons' and especially Cooper's papers have contributed to this issue.

Both Parsons (1978) and Cooper (1979) share Evans' conviction that there is a third, semantically distinct, reading for pronouns aside from the bound-variable reading and the deictic (=referential) reading, and that this third reading is to be treated on a par with definite descriptions. Each of them presents a Montague Grammar fragment which translates not only NPs with the definite article, but also certain personal pronouns, into Russellian descriptions, i.e., into logical expressions of the form:

$$\lambda P.\exists x(\forall y(...(y) \leftrightarrow x = y) \wedge P(x))$$

Both authors consider the treatment of donkey sentences a major application of their proposals, but restrict their attention to those cases where the crucial

indefinite is inside a restrictive relative clause, rather than an if-clause. Beyond these common traits, their proposals differ.

In Parsons' fragment, there is a syntactic transformation that substitutes certain definite descriptions by pronouns, namely those definite descriptions that match a clause somewhere to their left which contains an indefinite. The relevant matching-relation is characterized in purely syntactic terms in a definition whose exact formulation need not concern us here. It amounts to exactly the relation that pertains in the examples I have used to illustrate how Evans' E-type pronouns paraphrase as definites descriptions. For instance, Parsons' transformation will apply to derive (6) from (6a):

(6) John owns a donkey and beats it.
    a. John owns a donkey and beats the donkey such that John owns it.

The transformation is a meaning-preserving one, and therefore (6) and (6a) translate into the same logical formula, viz.:

(6′) $\exists u(\text{own}(j,u) \wedge \text{donkey}(u)) \wedge$
        $\exists u(\forall v((\text{own}(j,v) \wedge \text{donkey}(v)) \leftrightarrow u = v) \wedge \text{beat}(j,u))$

This entails:

    $\exists u \forall v((\text{own}(j,v) \wedge \text{donkey}(v)) \leftrightarrow u = v),$

which is precisely the uniqueness-claim that Evans says (6) implies, i.e., that John owns exactly one donkey.

In the derivation of a donkey sentence, such as (5) above, the substitution transformation applies to material which contains a bound-variable pronoun "he," bound by a previous application of quantifying in. Initially, (i) and (ii) below are generated with translations as specified:

(i)   He$_0$ beats the donkey he$_0$ owns.
    $\exists u(\forall v((\text{donkey}(v) \wedge \text{own}(x_0, v)) \leftrightarrow u = v) \wedge \text{beat}(x_0,u))$

(ii)   every man who owns a donkey
    $\lambda P. \forall w((\text{man}(w) \wedge \exists u(\text{donkey}(u) \wedge \text{own}(w,u))) \rightarrow P(w))$

Then (ii) is quantified into (i), binding *he*$_0$, which results in (iii), with translation (5b):

(iii)   every man who owns a donkey beats the donkey he owns

(5)   b. $\forall w((\text{man}(w) \wedge \exists u(\text{donkey}(u) \wedge \text{own}(w,u))) \rightarrow \exists u(\forall v((\text{donkey}(v) \wedge \text{own}(w,v)) \leftrightarrow u = v) \wedge \text{beat}(w,u)))$

Finally, (iii) undergoes the substitution transformation to form (5b), with the translation remaining the same. We can, once again, single out a uniqueness-entailment from this translation, namely this:

(5) c. $\forall w((man(w) \wedge \exists u(donkey(u) \wedge own(w,u))) \rightarrow \exists u \forall v((donkey(v)$
$\wedge \; own(w,v)) \leftrightarrow u = v))$

Notice that (5c) does not simply say that there is a unique donkey of a certain sort, but merely that, *for each donkey-owner*, there is a unique donkey that he owns, which is much weaker than an absolute uniqueness-claim would have been.

Cooper's fragment generates all pronouns directly and in free distribution, but provides for a richer range of translations for pronouns. Aside from the usual option of translating a pronoun as a variable over individuals, it specifies that a pronoun may translate into any expression of the following form (1979, 78):

$\lambda P.\exists x(\forall y(\Pi(y) \leftrightarrow x = y) \wedge P(x))$
where Π is a property-denoting expression containing only free variables and parentheses.

The choice of translation is free and does not depend on the syntactic environment of the pronoun. This means that a great number of translations are generated by the fragment for any given sentence that contains a pronoun. Cooper further assumes that the actual truth conditions which a sentence receives on a particular occasion of utterance are determined not by its translation alone, but by the translation interpreted with respect to a certain value assignment to its free variables, which is determined by pragmatic factors: Basically, a free variable gets evaluated as denoting the most salient entity in the context that is of the appropriate type. If the context lacks sufficient information to evaluate a certain free variable, then the utterance is either infelicitous, or else it must be disambiguated in favor of a different, more readily interpretable, translation. In this way, pragmatic factors in effect prevent most of the innumerable grammatically generated readings of a sentence from being accepted as actual readings for a given utterance.

Applied to our donkey sentence (5), Cooper's analysis works as follows: One of the grammatically available translations for the pronoun "it" is:

$\lambda P.\exists u(\forall v(R(x_0)(v) \leftrightarrow u = v) \wedge P(u))$

where R and x_0 are variables ranging over 2-place relations and individuals, respectively. (Here, $R(x_0)$ corresponds to Π in the schema above, and it obviously meets the requirement of consisting of free variables and parentheses only). If this translation for "it" is chosen in translating the sentence

he$_0$ beats it,

then this sentence translates as:

$$\exists u(\forall v(R(x_0)(v) \leftrightarrow u = v) \wedge beat(x_0,u))$$

From this, (5) can be derived by quantifying in "every man who owns a donkey" for "he$_0$," and the translation that results is:

(5)   d. $\forall w((man(w) \wedge \exists u(donkey(u) \wedge own(w,u))) \rightarrow$
            $\exists u(\forall v(R(w)(v) \leftrightarrow u = v) \wedge beat(w,u)))$

Under this translation, the sentence requires for its utterance a context which furnishes a value for the free variable R, i.e., in which a 2-place relation is salient. Since one particularly efficient way of making something contextually salient is mentioning it, and since uttering the NP "every man who owns a donkey" amounts to mentioning (among other things) the relation that holds between a donkey and its owner, this relation should be expected to enjoy a high degree of salience whenever (5) has just been uttered and has been processed up to the pronoun "it." Therefore, a most natural candidate for the evaluation of "$R(w)(v)$" in this context is predicted to be: "v is the donkey w owns," and this assignment of value makes (5d) equivalent to Parsons' translation (5b). Cooper thus explains the emergence of donkey-sentence readings by heavily relying on pragmatic factors which he takes to operate within fairly liberal limits of grammatically possible readings. The truth conditions that utterances of donkey sentences ultimately receive according to his explanation seem to coincide with those that Parsons' fragment assigns. In particular, they systematically entail a uniqueness-claim:

(5)   e. $\forall w((man(w) \wedge \exists u(donkey(u) \wedge own(w,u))) \rightarrow$
            $\exists u\forall v(R(w)(v) \leftrightarrow u = v)),$

i.e., that for each donkey-owner, there is exactly one individual that bears the relevant contextually furnished relation to him. On the occasions we are considering, i.e., those where the relevant relation is that of being-a-donkey-owned-by, (5e) amounts to the same as Parsons' (5c).

Before turning to a discussion of the uniqueness implications that are central to the predictions of both Parsons' and Cooper's (as well as Evans') analyses, let me briefly comment on the major difference between them: Whereas Parsons assumes that the descriptive predicate that enters into the definite-description-reading of a pronoun is determined strictly from the preceding linguistic environment, by a rule of grammar, Cooper relies on

grammar only for supplying the logical format of the definite-description-reading, whereas the particular descriptive predicate is filled in by the context, following general pragmatic principles for resolving vagueness. On occasion, these two views lead to different predictions.

Cooper's view leads us to expect that pronouns may exhibit definite description-readings in linguistic environments other than those that fit Parsons' structural description. Cooper's grammar generates definite-description pronouns freely and nothing prevents the free variables from denoting whatever is salient, regardless of whether its salience is due to preceding text of a particular form, or to any preceding text, for that matter. Indeed, Cooper cites various examples of definite description pronouns that fall outside the scope of Parsons' analysis (and also outside the scope of Evans' notion of E-type pronouns). These include, most prominently, the pronouns in Karttunen's paycheck-sentences[22], along with many other so-called "pronouns of laziness." For Parsons, these would have to be cases of yet another (i.e., fourth) kind of pronoun.

On the other hand, Cooper's heavy reliance on pragmatics makes him vulnerable to the same objection I raised earlier against Kripke's and Lewis' attempt at reducing anaphoric reference to a pragmatic phenomenon. Note the contrast between (5) and (7):

(7)   ? Every donkey-owner beats it.

Why does (7) not permit a reading in the sense of "Every donkey-owner beats the donkey he owns?" For Cooper, the only possible reason is that uttering "donkey-owner" is not an effective way of raising the salience of the relation $\lambda x.\lambda y.$(x is a donkey owned by y) — as opposed to uttering "man who owns a donkey," which is. If this is so, however, then salience rankings are not, after all, merely a product of applying Gricean principles on the basis of the meaning that has been extracted from previous discourse, but depend moreover on formal properties of what has been said. We may conclude that Cooper has not succeeded in unburdening grammar by letting independently available pragmatic principles constrain its output. Contrary to superficial appearances, Parsons' view on the division of labor between grammar and pragmatics may well be more realistic.

---

[22] Karttunen (1969); cf. (Cooper, 1979, 73-77).

2.3.1 Pro and contra the uniqueness implication

Both Cooper and Parsons are well aware that the truth conditions they predict for donkey sentences are in open conflict with the judgment that is often taken for granted, namely, that the truth conditions of (5) are:

(5)   f. $\forall w \forall u ((\mathrm{man}(w) \wedge \mathrm{donkey}(u) \wedge \mathrm{own}(w,u)) \rightarrow \mathrm{beat}(w,u))$

(5f) coincides in truth-value with (5b) and (5d) only when the state of affairs that obtains is of one of the following two types:

(i)   Every man owns either no donkey, or else he owns exactly one and beats that one;

or

(ii)   At least one man owns at least one donkey that he doesn't beat.

In the remaining states of affairs, the truth-values diverge:

(iii)   (5f) is true, but (5b) is false, if every man beats all of his donkeys, but at least one man owns at least two.

It is thus incumbent on Parsons and Cooper to argue that, given a state of affairs of type (iii), intuitive judgments disagree with (5f) and agree with (5b). Actually, neither author takes just that position; rather, they both argue, firstly, that we do not judge the utterance true in cases like (iii), and that (5f) is thus inadequate. Secondly, they concede that we do not really judge it false either, which shows that (5b) is not literally correct either. But they go on to suggest, thirdly, that our actual judgments are close enough to (5b) after all. Let me start by examining the first step, i.e., their reasons against (5f).

   Both authors derive an objection to analyses that identify (5) with (5f) from examples like these (which Parsons attributes to Partee):

(8)   Every man who has a daughter thinks she is the most beautiful girl in the world.
(9)   Every man who has a son wills him all his money.

What do such sentences entail about a man who has several daughters (or sons)? According to an analysis by which (5) means (5f), (8) should mean that every man thinks about *every* one of his daughters that she is the most beautiful girl in the world, and (9) should mean that every man wills *every* one of his sons all his money. In other words, (8) should ascribe contradictory beliefs to any father of two or more daughters, and (9) an

impossible will to any father of two or more sons. But in fact, so it is argued, we can make good sense of these sentences: We simply understand them as quantifying only over the range of men with at most one daughter, or at most one son, and within that restricted range, they claim nothing outrageous. Therefore, a theory that analyzes donkey sentences in terms of universal quantification, as in (5f), cannot be correct, at least not generally. This is Cooper's and Parsons' conclusion, at any rate.

Now it emerges upon further reflection about these examples that they do not really decide between a formalization like (5f) and one like (5b), as the latter stands, at all. If (8) and (9) are formalized just like (5b), then they simply come out false in any state of affairs where some father has two or more daughters (or sons), and should thus strike us as blatantly false in the real world. So Parsons and Cooper need additional assumptions to account for the data they have brought into the discussion. More specifically, they need to assume that sentences like (8) and (9) are tacitly understood as ranging over a narrower domain of quantification than that of all men who have a daughter (son). This is, of course, not an objectionable assumption; it is a commonplace that the domains of quantification in ordinary conversation are flexible and respond to all kinds of pragmatic determinants. However, if the data concerning (8) and (9) are supposed to discriminate between the analysis that Cooper and Parsons reject and their own, then the relevant pragmatic principles must be pinned down somewhat more clearly, i.e., clearly enough for it to become plausible that they cannot be appealed to to save the competing analysis as well.

Here we must go a little beyond the actual workings of the fragments that are spelled out and must separate implicatures (presuppositions) from entailments. Both authors, while not implementing the distinction in their fragments, would presumably agree that the uniqueness-claim that they treat as part of the truth conditions of donkey sentences has the status of an implicature (or presupposition — depending on terminological preferences), not only of an entailment. (It may be *only* an implicature and *not* an entailment, or else *both* implicature and entailment; I am assuming the latter, but it makes no difference at this point.) For example, (5c) (or (5e)) ought to be considered an implicature of (5). Furthermore, the authors seem to implicitly assume that implicatures play a role that ordinary entailments do not play in influencing the domains over which we understand quantifiers to range: If a universally quantified utterance carries an *implicature* that only part of the "literal" range of quantification satisfies, then we tacitly restrict our evaluation to that part of the range. On the other hand, if a genuine *entailment* of a universally quantified utterance is true only of part of the literal range, then — ceteris paribus — we judge the utterance false. Given

these assumptions — but not without them, as far as I can see — we can reconstruct the reasoning which leads Cooper and Parsons to interpret (8) and (9) as evidence for their analyses as opposed to an analysis by which the indefinite antecedents in donkey sentences have the force of universal quantifiers.

The reasoning, then, must be this: The range over which the NP "every man who has a daughter" quantifies is normally the set of fathers of at least one daughter. Sentence (8), however, carries the following implicature, according to Parsons' (and similarly Cooper's) analysis:

(8)   a. $\forall u((man(u) \wedge \exists v(daughter\text{-}of(v,u))) \rightarrow \exists v \forall w(daughter\text{-}of(w,u) \leftrightarrow v = w))$

Since this implicature is only true if all men with more than one daughter are disregarded, it does in fact signal to the listener that those men are to be disregarded in understanding and assessing the claim (8) advances. Hence our intuition that (8) discusses only men with at most one daughter. On the other hand, if (8) is assigned the truth condition (8b):

(8)   b. $\forall u \forall v((man(u) \wedge daughter\text{-}of(v,u)) \rightarrow u$ thinks that v is the most beautiful girl in the world)

and no uniqueness implicature is assigned along with it, then (8) is predicted to be unredeemably false (unless there are extraneous factors that delimit a sufficiently narrow domain).

It should be noted that, for both Cooper and Parsons, the argument I have just reconstructed is more or less the only argument for preferring their analyses over one which treats certain indefinites as wide-scope universal quantifiers. (Cooper lists two more arguments (1979: 81f), which turn out to be merely arguments against a particular, rather naive, version of such an analysis.) But for an argument that bears this burden, it is not very convincing. I have two criticisms: First, it is a doubtful assumption that only implicatures, but never genuine entailments, play a role in signaling how the speaker intends the range of quantification to be restricted. It is certainly true that not just any universal statement that is satisfied by only part of its literal domain gets automatically understood as a correspondingly weaker claim: If somebody says "All numbers are even," we are not willing to switch tacitly to a range of numbers in which this will come out true. However, there are cases where we seem to be more flexible. Consider an utterance of (10):

(10)   John thinks that nobody is as smart as him.

Normally, "nobody" ranges over all persons, including John. Yet we readily interpret it as excluding John in this sentence, and we do so just because

otherwise (10) would have a rather weird entailment: that John holds a contradictory belief concerning his own smartness. Note that we need no further (contextual or linguistic) clues to make this adjustment: the mere fact that a silly entailment would result is sufficient. Now my point is that (10) is not unlike (8) or (9). Therefore, if it is plausible to maintain that the restriction of the range of quantification in (10) is not grammatically, but somehow pragmatically, determined, then it is equally plausible that the grammatically determined meaning of (8) is plainly (8b), with the weaker reading we assign in practice an effect of the same pragmatic principles. But this means that (8) and (9) can no longer count as evidence against the analysis that rivals Cooper's and Parsons' anymore than (10) can count against an analysis by which "nobody" means $\lambda P. \neg\exists x(person(x) \land P(x))$.

My second criticism is this: Even if Cooper and Parsons had convinced us that only an analysis that assigns donkey sentences certain uniqueness-implicatures can account for (8) and (9), how could we jump to the conclusion that all donkey sentences carry analogous uniqueness-implicatures? Parsons, in addressing this question, reports that native speakers' judgments about the truth conditions of sentence (5) have been found insecure or non-uniform when they were presented with states of affairs where every donkey-owner beats at least one of his donkeys, but not everyone beats them all (p.18). He cautions that these findings were derived from a small and perhaps unrepresentative sample, but he does not point out that they do not exactly coincide with the predictions of his analysis anyway: People firmly judged (5) true when every donkey-owner beat all of his donkeys, even when some owned several; according to his analysis, the judgment should have been "anomalous" (or even "false") in such cases.

Cooper suggests that his own analysis, which posits a contextually evaluated relation-variable rather than taking the relevant relation to be fixed strictly by the antecedent and its linguistic environment, has a better chance of predicting those judgments. He writes, with reference to sentence (5) and his translation (5d) for it, the following:

> If we were to paraphrase the pronoun in this way [i.e., as 'the donkey he owns"], then the sentence would require that every relevant man own only one donkey and would be false if one of the men owned two donkeys. However, it is not necessary that [R(w)] be determined by the context of use as the property of being the only donkey owned by [w] – it could well be some other property which uniquely picks out one of [w]'s donkeys. This would get us an apparent anaphoric reading without requiring that [w] own just one donkey. Neither will the sentence have to come out false if some of the men beat more than one or even all of their donkeys, as long as the context of use is such that [R(w)] will pick out a particular one of [w]'s donkeys for each [w]. (Cooper, 1979, 84)

Thus Cooper succeeds in showing that his analysis does not predict that it would be downright impossible for (5) to get interpreted in a way which is consistent with a state of affairs in which some of the men in the domain own several donkeys. However, his analysis still predicts that rather special circumstances are required to warrant such an interpretation: the maximally salient relation that gets filled in for the variable R must be more specific than any relation that has been mentioned in the same sentence. According to Cooper, people judge (5) true when every man beats all of his multiple donkeys, just in case the context is rich enough for them to understand that they are to disregard all but a particular one of each man's donkeys. This prediction strikes me as false: If every man beats everyone of his donkeys, then the sentence is felt to be true and felicitous, even in the most impoverished context.

The shortcomings that inhere in both Parsons' and Cooper's analyses and are due to the uniqueness implications that they are committed to (even though Cooper's is a context-dependent and therefore more flexible uniqueness implication) are particularly visible when we consider examples like the following: Imagine that we are talking about a store which sells sage plants not individually, but only by flats of nine. Could I utter one of the following sentences then? I think I could.

(11)   Nobody who buys a sage plant here buys it all by itself.
(12)   Everybody who bought a sage plant here bought eight others along with it.

Or, a similar example:

(13)   Every man who owns a donkey owns a second one to keep it company.

Under Parsons' analysis, these sentences are blatant contradictions on any occasion of utterance. Cooper predicts that they make sense only in sufficiently special contexts: (12), for instance, is predicted to make sense under the presumption that "it" picks out one particular sage plant per buyer in some contextually determined way. But it just does not seem to be true that, in order to make sense of (12), one has to fix one's mind on any particular sage plant per buyer. Imagine somebody uttered (12), and you asked them to make explicit which one of every buyer's sage plants they meant. It would not feel like an appropriate thing to ask.

To summarize this subsection: The uniqueness-implications (or even entailments) that Cooper and Parsons are committed to by analyzing donkey sentences the way they do are on the one hand unnecessary for explaining the facts about sentences that prima facie exhibit them (i.e., (8) and (9)). And on

the other hand, they predict an adequate account of certain examples which force the assumption of non-uniqueness (i.e., (11) through (13)). Since Parsons' analysis appears to agree fully with the one that Evans sketches, this verdict generalizes to the latter, and we may conclude that donkey sentences are not satisfactorily interpreted by treating the crucial pronoun as an E-type pronoun and its indefinite antecedent as an ordinary narrow-scope existential quantifier. At least this conclusion is appropriate with respect to one type of donkey sentence, i.e., the type that involves a relative clause restricting a universally quantified NP, an antecedent with the singular indefinite article, and a singular personal pronoun as the anaphoric element. This is the only type we have dealt with in this section, and also the only type that either Cooper or Parsons address.

Evans talks about a wider range of phenomena, including if-then sentences on the one hand, and cases where various quantified plural NPs act as antecedents to plural pronouns. My next section is devoted to determining whether additional confirmation or disconfirmation for an analysis of the Evans-Cooper-Parsons kind might emerge when conditional sentences are considered. The discussion of plurals, unfortunately, would take me beyond the scope of this dissertation.

2.3.2 Conditionals

Applied to Evans' own example (1) (repeated below), the E-type analysis of the pronoun "he" is based on the following paraphrase:

(1) If a man enters this room, he will trip the switch.

     a. If a man enters this room, the man who enters this room will trip the switch.

I take it that the uniqueness implication that Evans would assign to (1a), and consequently to (1), is this:[23]

---

[23] (1b) would be an entailment of (1a) under a strictly Russellian analysis which does not provide for a notion of presupposition (or conventional implicature) as distinct from entailment. (1b) would be a conventional implicature, but not an entailment, of (1a) in a theory along the lines of Karttunen and Peters (1975), given the following assumptions: (a) A simple affirmative sentence "the F is G" has the truth condition "$\exists x(Fx \wedge Gx)$" and the conventional implicature $\exists x \forall y(Fx \leftrightarrow x = y.)$ (b) "If... then" acts as a so-called filter, i.e.: if A conventionally implicates B, and C conventionally implicates D then "If A then C" conventionally implicates "B and (if A then D)." In a theory along the lines of Gazdar (1979), I do not know whether (1b) would be an entailment, a presupposition, neither, or both of (1a). Gazdar remarks that uniqueness presuppositions fall outside the scope of his

(1)   b.   If a man enters this room, then exactly one man enters this room.

What exactly (1b) amounts to is of course contingent on our interpretation of "if-then." Let me start by identifying "if-then" with material implication. This is a gross simplification, but as I will show below, not one that affects significantly the present discussion. With "if-then" as material implication, (1b) amounts to (14):

(14) $\exists x(\text{man}(x) \wedge \text{enter}(x, \text{this-room})) \to \exists x \forall y((\text{man}(y) \wedge \text{enter}(y, \text{this-room})) \leftrightarrow x = y)$

In other words: "At most one man will enter this room."

Is it a fact that uttering (1) commits one to the truth of (14)? Certainly not. It is not even a fact that, whenever (14) turns out to be false, we feel that (1) does not really claim anything, or has unclear truth conditions. This is not to say that (1) is free from any vagueness and has a definite truth-value regardless of what happens. It is debatable, for instance, whether (1) is true or false if one man enters the room while he carries another man in with him, and only the former trips the switch. But that is not debatable just because more than one man entered the room. If, for instance, two men enter in succession, and each of them trips the switch, we have clear intuitions that (1) is true (assuming that these two are the only men that ever enter).

In a footnote (1980, 343), Evans remarks that apparent counterexamples to his predicted uniqueness-implications frequently can be explained away by making an implicit temporal parameter explicit. I think this remark should be taken to imply that (1) is more accurately paraphrased as (15), and its implication is therefore (16):

(15)   If there is a time at which a man enters this room, then he will trip the switch at that time.
(16)   If there is a time at which a man enters this room, then exactly one man will enter the room at that time.

At first sight, (16) looks like a weaker implication than (1b) (i.e., (14)). In particular, it would be true in the situation just mentioned, where two men enter in succession and each one trips the switch. After all, each man would be the unique man that enters at the time of his entering. However, Evans cannot really strengthen his point by the move from (1b) to (16). For one thing, (16) is itself a donkey sentence with respect to the anaphoric relation

between "a time ..." and "that time," and it remains unexplained how it comes to be read as a universal quantification over times, i.e., as (17):

(17) $\forall t(\exists x(man(x) \wedge x \text{ enters this room at t}) \to \exists!x(man(x) \wedge x \text{ enters this room at t}))$

But even if Evans can somehow predict in a systematic way that a conditional sentence like (1) expresses universal quantification over times and hence ends up with a uniqueness-implication like (17), there remains a problem. Consider a situation in which two men enter simultaneously and simultaneously trip the switch. This situation fails to satisfy the predicted implication, but my judgment would still be that the sentence is neither false nor deviant in this situation, but true. Now perhaps this kind of situation cannot possibly arise in our actual world, in which switches are built the way they are, and therefore we cannot really imagine how we would judge the truth-value of (1) if it did arise. So perhaps this particular example does not provide decisive evidence against the prediction that there is a uniqueness-implication like (17). But there are other examples that clearly do. Consider (4) again:

(4) If someone is in Athens, he is not in Rhodes.

According to Evans' analysis, this is equivalent to (4a) and carries the uniqueness-implication (18):

(4) a. If someone is in Athens, the one who is in Athens is not in Rhodes.

(18) $\forall t(\exists x(x \text{ is in Athens at t}) \to \exists!x(x \text{ is in Athens at t}))$

But it is evident that (4) can be uttered felicitously and truthfully when it is taken for granted that at any time there are many persons in Athens, i.e., when (18) is false and cannot even be rendered true by restricting one's attention to a subset of all times.

In view of examples like these, especially the last-mentioned one, we have to conclude that donkey sentences of the "if-then" variety are not adequately analyzed by interpreting the indefinite antecedent as an ordinary narrow-scope existential quantifier and the pronoun as an E-type pronoun — at least not if "if-then" is simultaneously analyzed as material implication, or as universal quantification over times. I have yet to argue that the same conclusion holds even if we presuppose a more sophisticated treatment of "if-then."

According to currently established views on the semantics of conditionals[24], "if-then" differs from a truth-functional connective in two major respects: (a) all "if-then" sentences are modalized, and (b) "if-then" does not in itself correspond to a logical operator at all, but depends on the operator with which it co-occurs for its interpretation. What this implies is that every conditional sentence is either overtly or covertly modalized. Which modality a given utterance of a conditional sentence expresses is determined by two parameters: a modal relation, and an accessibility relation. Usually, the modal relation is determined grammatically, while the accessibility relation is supplied by the context of utterance within limits imposed grammatically. In the following examples, the italicized expressions determine the modal relations of possibility and necessity, respectively:

(19)   If John enters this room, he *might* trip the switch.
(20)   If John is in Athens, he is *necessarily* in Greece.

A natural use of (19) would be to express "circumstantial" possibility, i.e., possibility in view of certain facts about the situation. This interpretation results when the contextually supplied accessibility relation is as follows (w and w' are variables for possible worlds):

> w' is accessible to w iff the location of the switch and the way it functions are alike in w' and w.

The truth conditions of an "if-then" sentence on a given occasion of utterance are a function of four things: the modal relation, the accessibility relation, the content of the if-clause, and the content of the then-clause. Suppose S_1 is the if-clause, and $M+S_2$ is the then-clause (where M stands for the expression which specifies the modal relation). Then we interpret the utterance as follows:

(i)   If M specifies the modal relation of possibility, then
"If S_1 then $M+S_2$" is true in w iff there is a world w' which is accessible to w and in which S_1 is true and S_2 is true.

(ii)   If M specifies the modal relation of necessity, then
"If S_1 then $M+S_2$" is true in w iff in every world w' which is accessible to w and in which S_1 is true, S_2 is also true.

(19), for instance, with the accessibility relation specified above, receives the following truth-condition according to (i):

---

[24] I am drawing mainly on Kratzer (1978, 1981b); some of the major ideas trace back to work cited there, especially by Stalnaker and Lewis.

(19) is true in w iff there is a world w′ such that:
  - the location and functioning of the switch in w′ is the same as in w;
  - John enters this room in w′; and
  - John trips the switch in w′.

One particularly common way of expressing the modal relation of necessity in conditionals is by using a bare then-clause with no overtly modalizing expressions in it at all. The kinds of accessibility relations that this sort of conditional accepts from the context are circumstantial and epistemic ones. For example, sentence (21) is understood as involving the modal relation of necessity and is thus interpreted according to (ii′) below, which can be viewed as a special case of (ii), where M is not realized lexically.

(21) If John enters this room, he will trip the switch.

(ii′) A "bare" conditional, which has no overt indicator of modality in the then-clause, "If S_1 then S_2" is true in w iff in every world w′ which is accessible to w and in which S_1 is true, S_2 is also true.

If (21) is uttered, for example, in a context that specifies the accessibility relation that I mentioned in connection with (19), then the resulting reading will be:

(21) is true in w iff for every world w′:
if w′ is like w with respect to location and functioning of the switch, and John enters this room in w′, then John trips the switch in w′.

There is a degenerate kind of circumstantial necessity that involves the following accessibility relation:

w′ is accessible to w iff w′ = w.

If we interpret a bare conditional with respect to this degenerate accessibility relation, the predicted reading is equivalent to material implication.

Let us return to Evans' proposal that donkey sentences of the "if-then" variety are to be analyzed by giving an E-type interpretation to the critical anaphoric pronoun. It is easily shown that the uniqueness implications associated with the E-type pronouns still come out inadequately strong, even if Evans' proposal is combined with the improved analysis of "if-then" just outlined. Consider again (4):

(4) If someone is in Athens, he is not in Rhodes.

As a bare conditional, (4) expresses some kind of necessity, most naturally another kind of circumstantial necessity, this time necessity in view of geographical and general physical facts. Let us assume an accessibility relation like the following:

w′ is accessible to w iff w′ is geographically like w.

The truth-condition for (4) that is presumably predicted if we combine (ii′) with Evans' analysis of E-type pronouns is (22):

(22) (4) is true in w iff for every w′ accessible to w:
       if there is an x such that x is in Athens in w′, then there is exactly one
       x such that x is in Athens in w′, and that one x is not in Rhodes in w′.

We can separate out the uniqueness implication from (22) by reading just up to the last comma. The predicted uniqueness application of (4) is thus that in every accessible world, there is at most one person in Athens.

 Now if we assume that the accessibility relation is roughly as described, then the actual world is clearly accessible to itself, and the predicted uniqueness implication is therefore incompatible with what we take to be the case in the actual world, i.e., that there are always many people in Athens. Likewise, if we assume any other accessibility relation which corresponds to an intuitively acceptable use of (4). Thus Evans' prediction remains just as counterintuitive when his analysis is incorporated into a sophisticated treatment of conditionals. In fact, the weaknesses of Evans' approach to donkey-anaphora are seen more clearly when conditional sentences are considered than when only sentences with restrictive relatives on universally quantified NPs are considered.

2.4 Donkey sentences and the ambiguity hypothesis.

Of the four positions that were introduced in the first half of this chapter, I have so far reexamined three in light of the donkey-sentence phenomena. It remains to be asked how the ambiguity hypothesis, i.e., the view that indefinites are ambiguous between a quantificational and a referential reading, fares in confrontation with donkey-anaphora.

 Are the indefinites that serve as antecedents in donkey sentences referential, or are they quantificational? They obviously are not referential; so if one continues to hypothesize this two-way ambiguity, they must be quantificational, i.e., existential quantifiers. But this answer leaves one without an account of the anaphoric relations with pronouns outside their scope into which these quantificational indefinites enter. Short of adopting a hybrid theory that combines the ambiguity hypothesis with some version of

either an Egli/Smaby-type or an Evans-type treatment of donkey sentences, no solution suggests itself. We may conclude that the ambiguity hypothesis offers no advantage over the positions discussed so far in dealing with donkey sentences.

2.5 Summary

Among the approaches to indefinites and anaphora that had shown reasonable success in accommodating the data involving cross-sentential anaphora that we have looked at in section 1, only two were prima facie capable of providing more or less automatically an analysis of donkey sentences as well. One of these approaches is based on the conviction – originally promoted by Geach – that all non-referring pronouns are bound variables, and it leads to Egli's and Smaby's investigations of the conditions under which an intrinsically existential quantifier receives a universal interpretation. The results of these investigations suggest that the relevant conditions are neither surface syntactic nor semantic ones, but apply in representations that arise in the course of deriving highly abstract logical forms from surface structures. This implies a theory of grammar which provides for a wide variety of construal rules, many of them structure-building. The empirical predictions, especially with Smaby's version of the approach, were found adequate up to problems with sentences involving non-standard quantifiers (e.g., "most"), especially non-standard adverbial quantifiers, and up to some instances of universal NPs taking exceptionally wide scope.

The other approach, represented by Evans, Cooper, and Parsons, purports to maintain a maximally simple theory of the quantificational force and scope-possibilities of indefinite NPs: they always express existential quantification, they are subject to the same scope-island restrictions as other quantifiers, and there are no special rules that export them. The donkey-sentence phenomena are not attributed to anything unusual in the behavior of indefinites, but rather to the existence of a third kind of pronoun that is semantically distinct from both referential and bound-variable pronouns and is taken to have the logical characteristics of a Russellian definite description. The theoretical implications of this approach cannot be assessed with complete certainty: Whereas Cooper's version of it is consistent with uncontroversial assumptions concerning the available types of grammatical rules and levels of representation and requires merely an additional lexical meaning for pronouns, it is doubtful whether it is capable of accounting for certain anaphoric-island phenomena. If not, then an anaphora rule with a rather complicated structural description must be posited (presumably in the

form of a construal or translation rule, rather than a syntactic transformation as Parsons proposed).

As for the observational adequacy of the definite-description approach to donkey anaphora, we have seen counterexamples to the uniqueness implications that the pronouns under consideration should carry if they were indeed Russellian descriptions. Particularly clear counterexamples arose among "if-then" sentences. Evidence that had been cited in favor of the controversial uniqueness implications did not withstand closer scrutiny. Let me emphasize again that I made no attempt at evaluating the isolated claim that donkey-pronouns are equivalent to definite descriptions, i.e., at evaluating it independently of the background assumption that definite descriptions receive a basically Russellian analysis. It is my conviction that without this background assumption the claim reduces to a mere observation that certain problematic locutions paraphrase each other, without leading to an analysis of either of them. If I could have resorted to any alternatives to Russell's theory of definite descriptions, I would have considered accordingly modified versions of Evans', Cooper's, and Parsons' proposals. But the only alternative readily available is that definite descriptions are (sometimes) referring expressions, and that would obviously have been of no help as a background to the proposals under discussion.

Despite the overwhelming differences between the two approaches to donkey sentences we have examined, we can find a common denominator. Proponents of either approach would agree to the following conclusion: The donkey sentences are evidence that the following three assumptions – each of them well supported independently – cannot all be true:

(a) Indefinites are existential quantifiers (except for the generic ones, of course), and never change force.

(b) Indefinites obey the same scope-island restrictions as other quantifying NPs (and there is no exportation).

(c) Pronouns are either bound variables, or else refer.

The two approaches diverge when it comes to deciding which assumption among (a), (b) and (c) should be modified.

3   Donkey Anaphora in Game-Theoretical Semantics

Hintikka and Carlson's analysis of donkey sentences, as presented in their (1979) article "Conditionals, generic quantifiers, and other applications of
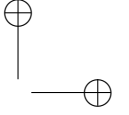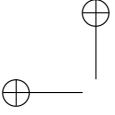
subgames," is cast in the framework of so-called "game-theoretical seman-tics" and irreducibly relies on characteristics of that particular framework. Game-theoretical semantics, a variety of truth-conditional semantics, takes the notion of truth to be derivative on a more basic notion of *verifiability* in an idealized sense of the latter. A sentence need not be verifiable in any concrete, humanly possible sense for it to be verifiable, and hence true, in the theoretical sense intended. For our present purposes, we may in fact assume that, as far as atomic sentences go, verifiability coincides fully with truth, and falsifiability with falsity.

For molecular sentences, i.e., those composed from atomic ones by logical connectives, quantifiers, and operators, verifiability is defined with recourse to a game of *verification*, in which two players, a verifier and a falsifier, play against each other, following rules which define the range of moves that are available to them at any particular point in the game. Each of these rules pertains to a particular logical symbol and applies whenever the sentence under consideration contains that logical symbol as its least embedded logical symbol. Examples of such rules are (G.∧) and (G.¬) below.

(G.∧)   When the game has reached a sentence of the form:

$$\phi \wedge \psi,$$

then the Falsifier may choose either ϕ or ψ, and the game continues with respect to the sentence chosen.

(G.¬)   When the game has reached a sentence of the form:

$$\neg \phi$$

then the players must switch roles and continue the game with respect to the sentence ϕ.

Each game is played until it reaches an atomic sentence, at which point it is decided who wins: if the sentence is verifiable, the Verifier wins; if it is falsifiable, the Falsifier wins. Here is an example of a game played by these rules in the actual world:
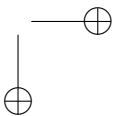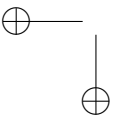
Players	A, B
Original Sentence	(John Paul II was shot) ∧ ¬(Paul VI was shot)
Original Roles	A = Verifier, B = Falsifier
1ˢᵗ step	B chooses "¬(Paul VI was shot)."
2ⁿᵈ step	B becomes Verifier, A becomes Falsifier, and the sentence "Paul VI was shot" comes into play.
3ʳᵈ step	"Paul VI was shot" is falsifiable, hence A, the current Falsifier, wins.
Result	A is winner.

Note that the rules are such that they often permit alternative courses that a game might take, even if the original sentence and the world are held fixed. In the example above, the first step could alternatively have been for B to choose "John Paul II was shot," and the continuation would have been different accordingly. A would have won anyway. But note that alternative courses of a game for a given sentence in a given world do not *necessarily* result in the same player's winning. If, for example, the sentence "(John Paul II was shot) ∧ ¬(John Paul II was shot)" is taken as the starting point of a game in the actual world, either player might end up winning, depending on which conjunct the Falsifier chooses in the first step. In other words, the rules do permit games in which the player that starts out in the role of Verifier emerges as the winner, even though the original sentence is false or even contradictory.

Moreover, the rules permit games in which the original Verifier loses although the original sentence is true. How, then, is it possible to base an adequate definition of truth on such game rules? The key to the answer is the notion of a *winning strategy*: A player is said to have a "winning strategy" in a game if he can win it regardless of the choices that his opponent makes, i.e., if for every sequence of rule-abiding moves that his opponent may choose, there is a sequence of rule-abiding moves that he can choose in order to win. This notion of a winning strategy is used in the following definition of truth:

> A sentence S is true in a world w iff the player who starts out in the role of Verifier has a winning strategy in any game that is played in w and starts with S as the original sentence.

It should be reasonably clear that this definition works and in fact amounts to exactly the same truth conditions for sentences involving ∧ and ¬ that one is familiar with. Further rules can be spelled out for the other truth-functional connectives and for the quantifiers in such a way that the truth definition above continues to coincide with familiar classical semantics. Before I turn

to the proposed application of game-theoretical semantics to pronouns and donkey anaphora, I will spell out three more game rules for later reference:

(G.∀)   When the game has reached a sentence of the form:

$$\forall x \phi$$

the Falsifier may choose any individual *a* and the game continues with respect to the sentence ϕ, all of whose free occurrences of x are henceforth taken as referring to *a*.

(G.∃)   When the game has reached a sentence of the form:

$$\exists x \phi$$

the Verifier may choose any individual *a* and the game continues with respect to the sentence ϕ, all of whose free occurrences of x are hence-forth taken as referring to *a*.

(G.COND)   When the game has reached a sentence of the form:

$$\text{if } \phi, \text{ then } \psi$$

proceed according to the following instructions:

(i) Reverse roles and play a subgame whose original sentence is ϕ. If the winner of this subgame is the player who started out as Verifier in the subgame, move on to (ii).

(ii) Assume the roles you had prior to executing instruction (i) and continue the game with respect to sentence ψ.

The notion of a "subgame" that appears in (G.COND) is self-explanatory. What (G.COND) says is roughly: In order to verify a conditional, first try to falsify the if-clause; if you do not succeed, i.e., if the if-clause gets verified, you must then verify the then-clause. ((G.COND) is not exactly the rule that Hintikka and Carlson give, but merely an approximation, to be revised below.)

    Why does (G.COND) have to be so complicated? If we look at it, we find that a much simpler rule would amount to exactly the same truth conditions:

Instead of the instructions (i) and (ii), one could just as well have specified that the Verifier gets to choose either $\neg\phi$ or ψ, and the game continues with respect to the sentence of his choice. This simpler instruction would have given the Verifier a winning strategy with exactly the same if-then sentences as the complicated rule (G.COND) above. However, there is a good reason for the particular formulation that has been given, which will become apparent as we turn to the game-theoretical treatment of anaphora.

So far, I have sketched game-theoretical interpretations of familiar logical symbols — truth-functional connectives, quantifiers, and the variables they bind — all of which can also be interpreted in "ordinary" semantics in an essentially (for our purposes, at least) equivalent way. Now I will turn to the central claim of Hintikka's and Carlson's paper, i.e., that a game-theoretically interpreted language can be enriched by a notational device that has no sensible interpretation in ordinary semantics: a variable-like element that is outside the scope of, but nevertheless somehow "bound" by, a quantifier. For example, consider a formula like (1):

(1) $\exists x Fx \wedge Gx$

Under familiar assumptions about the interpretation of quantifiers and variables, (1) is fully equivalent with (1′):

(1′) $\exists y Fy \wedge Gx$

In other words, the fact that we have identical variables in (1) where we have different ones in (1′) is a mere notational coincidence with no semantic significance. Within the framework of ordinary semantics, no alternative interpretation of variables that would result in a difference between (1) and (1′) suggests itself naturally. But the game-theoretical approach seems to open up new possibilities. Suppose, for example, we introduce a new game rule like the one below. (This is not a rule that Hintikka and Carlson actually suggest, but my own formulation of what I guess they have in mind.)

(G.PRO)   When the game has reached a sentence ϕ which contains free occurrences of a variable x which have not yet been assigned a referent in accordance with rule (G.∀) or (G.∃), then the free occurrences of x in ϕ are henceforth to be interpreted as referring to the individual *a* which fits the following description:

At an earlier point in the same game, *a* was chosen as the referent for an occurrence of x in some sentence ψ.

Examples will clarify the idea behind this rule. Consider first the following game, which involves applications of the rules (G.cond), (G.∃), (G.¬), and (G.pro):

(2)	**Players**	A, B
	Original Roles	A = Verifier, B = Falsifier
	Original Sentence	If ∃x(x is in Athens), then ¬(x is in Rhodes)
	1st step	A becomes Falsifier, B becomes verifier and "∃x(x is in Athens)" comes into play.
	2nd step	B chooses John, and "x is in Athens" comes into play, where x refers to John.

Since John is in fact in Athens, B has won the subgame w.r.t. "∃(x is in Athens)," and therefore (G.cond) prescribes to move on to the 3rd step:

	3rd step	A becomes Verifier, B becomes Falsifier, and "¬(x is in Rhodes)" comes into play.
	4th step	B becomes Verifier, A becomes Falsifier "x is in Rhodes" comes into play.
	5th step	(G.pro) applies and x in "x is in Rhodes" is assigned the referent John, because of the corresponding assignment earlier in the 2nd step

Since John in fact is not in Rhodes, the current Falsifier wins, i.e.:

| | **Result** | A wins. |

In this example game, I took the actual world to be one where John is in Athens and not in Rhodes. As the reader may confirm by some reflection, any other assumptions about the facts, as long as there is no one person that is both in Athens and Rhodes, would likewise have ensured A's victory.

It is indeed the case that (G.pro), in combination with other available rules, ensures that the Verifier for the sentence "If ∃x(x is in Athens), then ¬(x is in Rhodes)" has a winning strategy if and only if no one who is in Athens is also in Rhodes. So we need only to assume — naturally enough — that one of the logical forms of the Engllish sentence:

(3)   If someone is in Athens then he is not in Rhodes.

looks just like the formula we have been looking at, and a satisfactory account of the donkey-sentence reading of (3) emerges before our eyes.

As I will show shortly, this account is not yet fully satisfactory. But let me first make some general comments about the treatment of donkey sentences that we have just seen an application of. Compared with the other two types of treatment we have discussed, this one has some very appealing features: It shares the main advantage of the Evans/Cooper/Parsons approach in that it maintains a uniform analysis of (non-generic) indefinites as existential quantifiers whose behavior differs in no way from that of other quantified NPs. At the same time, it gets around the problematic uniqueness-implications and instead generates the more adequate truth conditions that Egli and Smaby have assumed: A sentence like (3) above is predicted to have the force of universal quantification over all people that are in Athens. Moreover, the rule for interpreting anaphoric pronouns that makes these predictions fall out strikes one as rather natural; it comes very close to the naive idea that pronouns serve to pick up a previous reference, except that "previous reference" is replaced by "previous instantiation" (as one might call it) and the notorious puzzles concerning non-referring antecedents are thereby avoided. So the game-theoretical approach compares very favorably with its competitors, at least on a superficial examination. I will now take a closer and more critical look.

The system of rules I have so far presented is clearly inadequate as it stands for the following reason. Consider sentence (4):

(4)   If Bill owns every donkey he beats it,

which is not grammatical under a reading where "it" has "every donkey" as its antecedent. This illustrates a systematic contrast in the abilities of existential and universal NPs to serve as antecedents for donkey anaphora, a contrast which Hintikka and Carlson discuss explicitly. The trouble with (G.PRO) is that it would not rule out (4) any more than its counterpart with an existential antecedent (5):

(5)   If Bill owns a donkey he beats it.

A game for (4) – or rather, for its formalized representation (4′):

(4′)   If $\forall x(\text{donkey}(x) \rightarrow \text{own}(\text{Bill},x))$, then beat(Bill,x)

might run as follows: First the original Verifier, A, makes an unsuccessful attempt at falsifying the if-clause "$\forall x(\text{donkey}(x) \rightarrow \text{own}(\text{Bill},x))$," in the course of which he picks an individual a which turns out to be a donkey owned by Bill. In accordance with (G.COND), the game then proceeds to the then-clause, "beat(Bill,x)." Following (G.PRO), x is interpreted as referring a. As a turns out to be, in fact, beaten by Bill, A emerges as the winner and

has verified (4′). Generally speaking, A has a winning strategy for (4′) if and only if either Bill does not own every donkey, or else Bill beats at least one individual, whatever that may be. But this is not an adequate truth condition for any grammatical reading of (4).

Hintikka and Carlson, in their discussion of the contrast between (4) and (5), claim that their theory succeeds in predicting that (4′) cannot be won. They avoid the difficulty I have just pointed out by means of a formulation of (G.cond) that is superficially equivalent to the one I gave above, but in effect serves to prevent certain illicit applications of (G.pro). Basically, they advocate a rule that amounts to (G.cond) with part (ii) reformulated as follows:

(ii′)   Forget all of the moves that were carried out during the subgame about ϕ by that player who started out as *Falsifier* in the subgame. Remember only those moves from the subgame about ϕ that were carried out by the player who was *Verifier* at the start of the subgame. Now assume the roles that you had prior to executing instruction (i) and continue the game with respect to sentence ψ.

This instruction about forgetting certain moves that have taken place within a subgame, but retaining the memory of others beyond the completion of the subgame, would seem mysterious and pointless if we were dealing only with conditionals whose if- and then-clause are closed sentences. However, it becomes meaningful in connection with rule (G.pro): rule (G.pro) assigns to a free variable a value depending on a choice that was carried out earlier in this game, and it therefore cannot apply unless that previous choice has been retained in memory. Let us consider (4′)) again to appreciate the effect of amending (ii) to (ii′):

Suppose A starts as Verifier in a game about (4′). As the if-clause comes into play, A (now in the Falsifier role) picks the individual *a* which happens to be a donkey owned by Bill. A has thus lost the subgame, and we must proceed to instruction (ii′) of the revised rule (G.cond). (ii′) says that all of A's moves during the subgame must be erased from memory. This means in particular that we are to forget that A picked *a* to instantiate x. (We are also supposed to remember all of B's moves in the subgame, but this is irrelevant as there were none in this case.) Now A becomes Verifier and "beat(Bill,x)" comes into play. We try in vain to apply (G.pro) to the free variable x: There has been no previous value-assignment to an occurrence of x, for all we remember. So the game stops short of completion – presumably a criterion for semantic deviance of the original sentence.

Had the original sentence been (5′) rather than (4′)

(5′)   If ∃x(donkey(x) ∧ own(Bill,x)), then beat(Bill,x),

the game would have run smoothly: During the subgame about the if-clause "$\exists x(\text{donkey}(x) \land \text{own}(\text{Bill},x))$," it would have been *B's* turn to pick a value for x. That move, being a move of the player who starts the subgame as *Verifier*, would have been remembered, according to (ii′), throughout the entire game, and (G.PRO) would have applied straightforwardly. So it looks like (ii′) draws just the right line between the cases where we can have donkey anaphora, and the cases where we cannot.[25]

The amendment of (G.COND) that we have just seen to be necessary if the game-theoretical treatment of anaphora is to work properly is only one of many stipulations that will have to be added to the game rules for all the various connectives in order to keep track of what is to be forgotten and what is to be carried along in memory from subgame to subgame. For example, instructions much like those in (ii′) will have to be written into a game rule for sentences conjoined by "and," so that the contrast in anaphoric possibilities between (6) and (7) can be predicted:

(6) Bill owns a donkey and he beats it.
(7) Bill owns every donkey and he beats it.

A question which comes to mind is whether those memory-instructions have to be stipulated individually, i.e., rule by rule for each connective; or can they be made to follow from more general principles? Hintikka and Carlson's paper creates the impression that they are part of the individual rules, but the question certainly remains open.

Even with the amended version of (G.COND), the analysis seems to make wrong predictions about the following sentence:

(8) If Bill doesn't own every donkey he beats it.

No anaphoric relation between "every donkey" and "it" is possible here, regardless of whether "not" or "every" is given widest scope inside the if-clause. If "every" is interpreted with scope over "not," anaphora is ruled

---

[25] There remains a problem with sentences like (4) that Hintikka and Carlson do not seem to be aware of: While (ii′) does block any attempt to verify the then-clause of such a sentence, it still does not rule out *all* conceivable verification strategies: (4′) can still be verified whenever its if-clause happens to be false. Since the existence of a winning strategy for the Verifier is equated with the truth of the sentence, this would seem to result in the counter-intuitive prediction that the sentence (4), under the reading represented by (4′), is true iff Bill does not own every donkey, and deviant otherwise. I believe this problem is not very serious. We might simply adopt a stricter criterion for deviance: A sentence is deviant if the game w.r.t. it stops short of completion in at least one possible world. By this criterion, (4′) would be deviant even though it might be verifiable in the actual world.

out in exactly the same way as it was ruled out with (4). But if "not" takes scope over "every," nothing that has been said so far blocks anaphora. The corresponding logical form would be

(8′)   If ¬(∀x(donkey(x) → own(Bill,x))), then beat(Bill,x)

In the subgame concerning the if-clause of this sentence, the player who gets to pick the value for x is the one who has started the subgame in the role of Verifier (although he has subsequently become Falsifier because of the role-switching prescribed by (G.¬)). Therefore, his choice should not be forgotten when the then-clause comes into play, but remembered through the end — that is what (ii′) says, at any rate. In fact, the current system of rules predicts that there is a winning strategy for the verification of (8′) whenever there is one for the verification of (5′), despite the fact that (8) does not have a grammatical reading equivalent to the donkey-sentence reading of (5).

Various ways to fix up the system might be thought up. My suggestion would be that the game rule for negation should be rewritten to include a stipulation of general amnesia with respect to any moves made "inside the scope of" negation:

(G.¬)   (revised) When the game has reached a sentence of the form:

$$\neg\phi,$$

reverse roles and play a subgame whose original sentence is ϕ. At the end of the subgame, declare the winner of it the winner of the superordinate game, i.e., the game concerning "¬ϕ," and forget all moves that took place within the subgame.

The reader is welcome to verify that this, in conjunction with the other rules, suffices to block any attempts at verifying (8′) (disregarding those attempts which succeed by falsification of the if-clause — cf. footnote 25). It also makes welcome predictions with other examples I can think of. Hintikka and Carlson have a lengthy discussion of their own on the topic of ruling out (8′) and arrive at a different proposal for amending the game rule of negation, which I have been unable to understand, cf. (1979, 26-29).

I have devoted the largest part of this section to a description of Hintikka and Carlson's proposal. I should warn the reader that I am not at all certain that my description is in fact faithful to the authors' intentions. Their paper contains various remarks that are not consistent with the central ideas I attribute to them, so I may well have misunderstood them. At any rate, what I have represented as their theory here is the result of my best efforts to make

sense of their paper, and it is a theory that strikes me as quite promising. Unlike its competitors, which we discussed in the main body of the chapter, this game-theoretical treatment of anaphora preserves an unambiguously existential analysis of indefinites and offers a relatively natural explication of the relation between anaphoric expressions and their antecedents.

The crucial payoff, in my opinion, lies in the various stipulations that must be introduced to regulate which moves are remembered and which are forgotten when a subgame is concluded and the main game resumes. Unfortunately, it is just those stipulations that Hintikka and Carlson are least explicit about in their article. How complicated will a full specification of the required instructions concerning memory turn out to be? Are there general principles involved? Is there any way of predicting, say, from the meaning of negation that none of the moves in the subgame about the negated formula are later accessible, or of predicting from the meaning of "if... then" that just the *Verifier's* moves in the subgame about the antecedent should go on record permanently? As long as such questions are not even addressed, let alone answered, I find it hard to assess the merits of the game-theoretical approach with any confidence. I hope that further research will make the requisite assumptions more explicit and, in particular, provide criteria for a comparison between that approach and the one I will now proceed to develop.
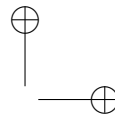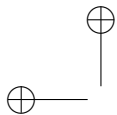
Chapter II

INDEFINITES AS VARIABLES

In this chapter I will start presenting my own solution to the problem that underlay the discussion of the literature in the preceding chapter: How can we reconcile the quantificational nature of the indefinite with its role as an anaphoric antecedent, and in particular, how do we explain the donkey phenomena? Among the approaches we have seen, there were those which concentrated on revising and diversifying the semantics of pronouns, and there were others whose focus was on conceding indefinites a greater range of quantificational forces and scopes. My approach is closer to the latter in that I, too, think it is the semantics of indefinites that needs revising. But the revision will be more drastic than to let indefinites change from existential into universal quantifiers under certain conditions: I am denying that they ever have any quantificational force of their own at all. What appears to be the quantificational force of an indefinite is always contributed by either a different expression in the indefinite's linguistic environment, or by an interpretive principle that is not tied to the lexical meaning of any particular expression at all. What I mean by this will hopefully become understandable by the end of this chapter.

1   The Adaptability of Indefinites, and Lewis on "Adverbs of Quantification"

The reason why we find it so plausible that indefinites have the inherent force of existential quantifiers is that we always look at the simplest and most ordinary examples first: examples like "a policeman came to the office," and "John owns a car." It would not naturally occur to us to start constructing a theory of indefinites by initially ignoring such ordinary examples and instead focusing primarily on logically complex and not very colloquial examples like these:
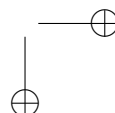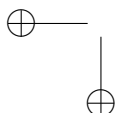
(1)   If a man owns a donkey he always beats it.
(2)   In most cases, if a table has lasted for 50 years, it will last for another 50.
(3)   Sometimes, if a cat falls from the fifth floor, it survives.
(4)   If a person falls from the fifth floor, he or she will very rarely survive.

Yet, I am claiming that these are the examples in which the indefinite exhibits its true semantic nature most openly.

We need to settle some preliminaries before we can look specifically at the indefinite's function in constructions like (1) through (4); in particular, we need to identify the semantic function of the adverb (i.e., "always," "in most cases," "sometimes," and "very rarely," respectively) in each of these sentences. These adverbs are (or can be, under at least one grammatical reading of the sentences under consideration) what Lewis (1975) calls "adverbs of quantification," and I adopt his analysis of them. According to Lewis' analysis, the Q-adverb is the main operator in a sentence like (1)-(4), and it takes two sentential arguments, which are expressed by the if-clause and by the remainder of the matrix clause, respectively. Each of these sentential arguments is, logically, an open sentence, i.e., a formula with one or more free variables. In other words, the "canonical form" (as Lewis calls it; we might also say: the "logical form") of such sentences conforms to schema (5):

(5)   Q-Adv (ϕ, ψ)

The canonical form of (1), for instance, is this:

(1′)   Always ((x is a man $\land$ y is a donkey $\land$ x owns y), x beats y)

(Do not worry, for the moment, about the question of why "a man owns a donkey" amounts to "x is a man $\land$ y is a donkey $\land$ x owns y" in logical form. We will soon enough take up this question in great detail.)

In interpreting these logical forms, Lewis treats Q-adverbs as "unselective quantifiers." In contrast to the familiar, "selective," quantifiers, unselective quantifiers bind not just one particular variable, but an unlimited number of them simultaneously. For instance, "always" is interpreted as follows:

(i) "always (ϕ, ψ)" is true iff every assignment to the free variables in ϕ which makes ϕ true also makes ψ true.

If we apply (i) to the example of (1′), the unselective quantifier "always" will in effect bind two variables, x and y, but with a different example, it might have bound more or less than two, depending on how many there happen to occur in the sentence. (1′) turns out to have truth conditions identical to those of (6), where the unselective quantifier has been paraphrased by selective ones:

(6) ∀x∀y((x is a man ∧ y is a donkey ∧ x owns y) → x beats y)

Note also that ϕ, the left sentential argument of the Q-adverb, which is expressed by the if-clause in the original English sentence, plays the semantic role of restricting the domain of quantification: According to (i), for instance, the truth of "always (ϕ, ψ)" depends not on whether any variable assignment whatsoever satisfies ψ, but merely on whether those variable assignments satisfy ψ which belong to the restricted set of variable assignments satisfying ϕ.

In an analogous way, ϕ functions as a restriction on unselective quantifiers other than the universal one, as (ii), (iii), and (iv) below show. To restrict an adverbial quantifier thus seems to be a typical function of if-clauses.

(ii) "in most cases (ϕ,ψ)" is true iff most assignments to the free variables in ϕ which satisfy ϕ also satisfy ψ.

(iii) "sometimes (ϕ,ψ)" is true iff some assignments to the free variables in ϕ which satisfy ϕ also satisfy ψ.

(iv) "very rarely (ϕ,ψ)" is true iff very few assignments to the free variables in ϕ which satisfy ϕ also satisfy ψ.

With the help of (ii) through (iv), the reader should be able to figure out Lewis' truth conditions for the remaining three of our examples, (1)-(4) above.

Having thus clarified the semantics of Q-adverbs, as well as the contribution that if-clauses make when they co-occur with Q-adverbs, let us return to our central concern: What do the indefinites that occur inside the if-clauses in our example sentences (i .e., "a man," "a donkey," "a table," etc.) contribute to the meaning of each sentence? And more specifically: Do they act as quantifiers in these sentences? There are two different ways in which we might approach these questions:

(a) We might start from the observation that (1) through (4), under the intended readings, admit of paraphrases like the following, in which the indefinites have been eliminated in favor of various undisputedly quantified NPs:

(1) a. For every man and every donkey such that the former owns the latter, he beats it.
(2) a. Most tables that have lasted for 50 years last for another 50.
(3) a. Some cats that fall from the fifth floor survive.
(4) a. Very few people that fall from the fifth floor survive.

On the basis of these paraphrases, we might conclude that, despite their superficial uniformity, indefinites may act as all sorts of different quantifiers, depending on their varying environments: "an F" becomes a universal quantifier over Fs in the context of the universal Q-adverb "always"; it becomes an existential quantifier over Fs in the context of the existential Q-adverb "sometimes"; it assumes the quantificational force of "most Fs" in the neighborhood of "in most cases"; and the force of "very few Fs" in the neighborhood of "very rarely." In short, its quantificational forces vary widely as they adapt to whatever Q-adverb happens to be around. Viewed like this, indefinites are quantifiers, but their quantificational force adapts to the environment like the color of a chameleon.

(b) Alternatively, we might interpret the data under consideration as evidence that indefinites (at least some of them) simply have no quantificational force of their own at all, but are rather like variables, which may get bound by whatever quantifier is there to bind them. In sentence (1), for instance, the logical form of the if-clause contains the variables x and y free, and the quantifier "always" binds them in the logical form of the complete sentence.

In view of what has so far been presented of Lewis' analysis, the second answer, (b), clearly sounds more sensible than the first, (a). Why, after all, should we treat the indefinite as a quantifier when we have already identified the adverb as the one responsible for the quantificational force of the sentence? The rest of this chapter can in fact be seen as an attempt to spell out (b), i.e., the thesis that indefinites resemble variables more than quantifiers. But before I turn to this task, a clarifying note concerning Lewis' own position is in order.

If I read him correctly, Lewis endorses neither the adaptable quantifier nor the variable analysis of indefinites. In the last section of his paper (p. 13ff.), he sketches out a system of transformational rules by which to convert his canonical forms into various types of English sentences, among them sentences of the form of our examples (1)-(4). One among his transformational operations (or rather, a sequence of some of them) effects what he calls the "displacing of restrictive terms": It deletes a clause of the form "α is τ" (where α is a variable and τ an indefinite NP) and simultaneously replaces another occurrence of α by τ (and any additional occurrences of α that the sentence may contain by pronouns of the appropriate gender and number to match τ). For example, this displacement operation may apply to the canonical form (3′) as follows:

(3′)   Sometimes (($\underbrace{x}_{=\alpha}$ is $\underbrace{a\ cat}_{\tau}$ $\wedge$ x falls from the fifth floor), x survives)
$\Downarrow$ ${\uparrow}\atop{=\alpha}$ ${\uparrow}\atop{=\alpha}$

Sometimes (*a cat* falls from the fifth floor, *it* survives)

If we were to adopt some version of this analysis of indefinites in terms of displacement, we would be led to respond differently to the question of whether or not the indefinites in our examples are quantifiers: In the logical forms, they (or their sources, to be more accurate) are predicate nominals, and as such may be analyzed as existential quantifiers in the object position of identity "be" (see Montague (1974) for such an analysis of predicate nominals). They would thus appear to be perfectly ordinary indefinites, which happen to undergo a not-so-ordinary displacement operation, in the course of which, among other changes, a variable disappears. I will not pursue this analysis any further; it might work, for all I know, but I am not aware of any advantages it would have over my own proposal, which is prima facie less ad hoc.

The idea that I am going to pursue is that the indefinite itself has basically the meaning of a variable. Lewis and I thus agree when it comes to the interpretation of a sentential constituent such as (7), as it occurs after "if" within (1):

(7)   a man owns a donkey

For both of us, (7) contains free variables, i.e., is tantamount to a formula like (7′):

(7′)   man(x) ∧ donkey(y) ∧ own(x,y)

But I make the further claim — which Lewis would reject — that the free variables in (7′) have been contributed *by the lexical meaning of the indefinite*. Moreover, I do not just want to say that a variable-reading of the indefinite is *one out of several* lexical meanings, restricted to emerge only in constructions involving adverbial quantification, as it would appear. My aim is to argue that the indefinite never contributes anything more than this variable-reading to the meaning of the sentence in which it occurs, whatever type of sentence it may be. In other words, whenever the indefinite seems to act as a quantifier in some utterance, it is really something else about the utterance that is responsible for the quantificational force. In the examples we have been looking at, this "something else" is simply another expression in the same utterance, the adverb of quantification. But in most examples, it will not be quite so manifest. In (8), for instance, the contributor of

quantificational force (in this case, universal) is a morphologically unrealized necessity operator, as I will argue, kind of an invisible "always."

(8)   If a man owns a donkey he beats it.

In (9), it will turn out to be the "every" which binds the free variable of "a donkey," despite the fact that the head noun of the "every"-NP is "man," and one might therefore expect that "every" could bind only a variable ranging over men.

(9)   Every man who owns a donkey beats it.

In (10), finally, the existential force we observe cannot be blamed on any part or property of the sentence at all, and I will attribute it — at least for the time being — to a rule of "Existential Closure" that applies obligatorily in certain environments.

(10)   A man came in.

All these claims need to be formulated more precisely before we can discuss their validity. To this end, the next section introduces some formalism.

2   Logical Forms for English Texts with Indefinites and Quantifiers

In this section and the next one, I will outline an extensional interpretation for a certain fragment of English. I will proceed in two steps: In the first step, syntactically analyzed expressions of English[1] are associated with certain disambiguated representations; in the second step, these disambiguated representations are assigned truth conditions. The disambiguated representations, which I also call "logical forms," bear resemblances to the logical forms of the Revised Extended Standard Theory, but also to the analysis trees of Montague Grammar. I will not address any issues bearing on a choice between these two frameworks. An English sentence, even if it has just one syntactic analysis and contains no lexically ambiguous elements, may still be ambiguous for one of these two reasons: (a) There may be more than one potential antecedent for an anaphoric element. (b) The scope of a quantifier may be underdetermined. Disambiguated representations are

---

[1] The relevant level of syntactic analysis is presumably somewhat more abstract than surface structure. Perhaps it is "S-Structure" in the sense of Chomsky (1981). I will use the neutral term "syntactic structure" or "syntactic representation" to refer to whatever level of syntax feeds into the construal component (see below).

designed to resolve both types of ambiguity: anaphoric relations are marked by numerical subscripts, and quantifier scope is marked configurationally, by positioning the quantifier in such a way that its scope is the substructure it c-commands. Syntactic structures therefore stand in a one-to-many relation to logical forms. Put differently, logical forms represent sentences under certain "readings."

Syntactic structures will be related to logical forms by means of a number of "rules of construal," which transform the former into the latter. There are, furthermore, wellformedness conditions on logical forms which rule out some otherwise derivable structures.

The first construal rule is NP-Indexing: *Assign every NP a referential index.* By "referential index" I mean a numerical subscript. This is a technical term and not intended to suggest that NPs with referential indices refer to anything. Incidentally, we might not really need NP-Indexing as a construal rule, because perhaps NPs are already indexed in the syntactic representations that are the inputs to the construal component. Be that as it may, all that matters for our purposes here is that the indices be there by the time we need them.

Another rule is NP-Prefixing:[2] *Adjoin every non-pronominal NP to S,* leaving behind a coindexed empty NP. By performing these two operations on sentences (1), (2), and (3), we arrive at structures (1′), (2′), and (3′), respectively.

(1) He arrived.
(2) A man arrived.
(3) Every man arrived.

(1′) [$_S$ he$_3$ arrived]
(2′) [$_S$ a man$_1$ [$_S$ e$_1$ arrived]]   ("e" = empty NP)
(3′) [$_S$ every man$_2$ [$_S$ e$_2$ arrived]]

Furthermore, there is a rule of Quantifier Construal (distinct from NP-Prefixing): Attach every quantifier as a leftmost immediate constituent of S. This will, for example, apply to the quantifier "every" in (3′) and turn (3′) into (3″).

---

[2] I am here disregarding the indisputable fact that NPs may also take scope over N and other non-S constituents. See, e.g., Montague (1974), Higginbotham (1980).

(3″)

```
                        S
         ┌──────────────┼──────────────┐
      every           NP₂              S
                   ┌────┴────┐       ╱────╲
                  __       man     e₂ arrived
```
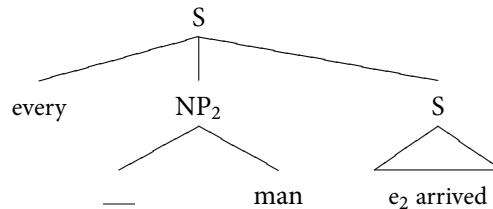
Quantifier Construal will also apply to adverbial quantifiers, e.g., the "always" in sentence (4).

(4)   If a restaurant is good it is always expensive.

In order to derive the disambiguated representation of (4), we must also appeal to a rule which positions the if-clause between the quantifier it restricts and the rest of the sentence. The structure resulting from applying both this rule and Quantifier Construal to (4) should look like this:

(4′)

```
                     S
          ┌──────────┼──────────────┐
       always        S̄              S
                     ╱╲            ╱──╲
                   if ...      it₁ is expensive
```
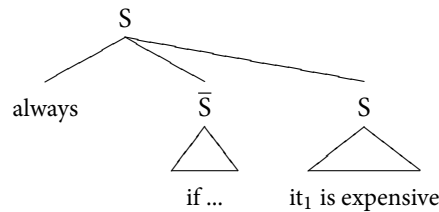
Quantified sentences will generally have tripartite logical forms like (3″) or (4′), where the second part expresses a restriction on the domain of quantification.

Before I move on to more rules of construal and to more detail in the disambiguated representations to be interpreted, a few preliminary remarks on the intended interpretations may be helpful: Pronouns as well as empty NPs are to be treated as individual variables, with identity of referential indices indicating identical variables. I will assume that pronouns and empty NPs are the only NPs that occur in the minimal S-constituents of interpretable disambiguated representations. NPs with full nouns in them are interpretable only in structures where they are sisters to an S-constituent, i.e., where NP-Prefixing has applied. Since I also assume that quantifiers can be interpreted only when they are immediate leftmost constituents of S (i.e., have undergone Quantifier Construal), no NP in an interpretable logical form will ever have a quantifying determiner. For example, (3′) is not possible as a logical form for (3); it is merely an intermediate stage in the derivation of (3″). (2′), on the other hand, is interpretable as it stands, since the indefinite article is not a quantifier.

NPs of either the form [indefinite article $\overline{\text{N}}$]$_i$ or the form [__ $\overline{\text{N}}$]$_i$ (with an empty determiner position from which a quantifier has been removed by Quantifier Construal) will all be interpreted as though they were sentences: An NP of either form will be identified with the open sentence "$\overline{\text{N}}(x_i)$," which is true iff the value of the variable x_i has the property expressed by $\overline{\text{N}}$. (It may be somewhat unconventional to assign propositional denotations to NPs, but I see nothing wrong with it in principle. One could probably devise a semantics substantially equivalent to mine that interprets NPs uniformly as second-order predicates.) The contribution of an NP, interpreted sententially in this way, to the interpretation of the S it is part of depends crucially on whether or not the construction is headed by a quantifier. In the former case, the NP restricts the quantifier that appears as its left sister. For instance, the interpretation of (3″) is roughly: Every variable assignment that satisfies [__ man]$_2$ also satisfies [e$_2$ arrived]. In a quantifier-less structure like (2′), however, a conjunctive interpretation applies, i.e., the whole S is true just in case both its immediate constituents are true. For (2′), we thus arrive at an interpretation identical to that of the open sentence "man(x_1) ∧ arrived(x_1)." Note that this is not an interpretation which reflects our intuitive judgment as to what the truth conditions are of an unembedded occurrence of sentence (2). I will deal with this discrepancy right below.

For if-clauses, as in (4′), the intended interpretation is simply that the "if" is semantically vacuous and the if-clause has exactly the same interpretation as the sentence after "if." I assume that if-clauses must occupy restrictive term positions in logical form as a matter of wellformedness conditions on logical forms. The interpretation of quantifier-initial tripartite constructions is no different when the quantifier is adverbial and the restrictive term an if-clause, than it is when the quantifier is a determiner and the restrictive term an NP. There will be a general rule schema for the interpretation of structures of the form [Quantifier, X, S] which covers (3″) and (4′) simultaneously.

We now have to determine the conditions under which indefinites are read existentially. This happens whenever the indefinite is in an unembedded sentence. It furthermore happens when the indefinite is in the third part of a tripartite structure, i.e., in the scope of a quantifier, but not in the restrictive term. Examples of this are provided by sentences like (5) and (6).

(5) Every man saw *a cat*.
(6) If a man is lonely, he often buys *a cat*.

By the rules of construal that have been introduced so far, (5) and (6) will be represented as follows:

(5′)

```
                          S
        ┌─────────┬───────────────────┐
     every       NP₁                   S
              ┌──────┐          ┌──────────┐
             __      man       NP₂          S
                              ┌────┐      ┌───△───┐
                              a   cat    e₁ saw e₂
```

(6′)

```
                               S
        ┌────────────┬──────────────────────────┐
     often          S̄                            S
              ┌──────────┐              ┌──────────────┐
             if          S             NP₂              S
                    ┌─────────┐      ┌─────┐       ┌────△────┐
                   NP₁         S     a    cat     he₁ buys e₂
                ┌──────┐    ┌──△──┐
                a     man   e₁ is lonely
```
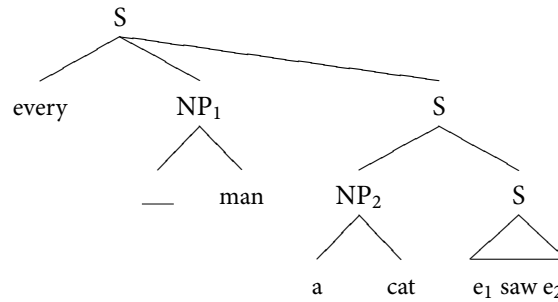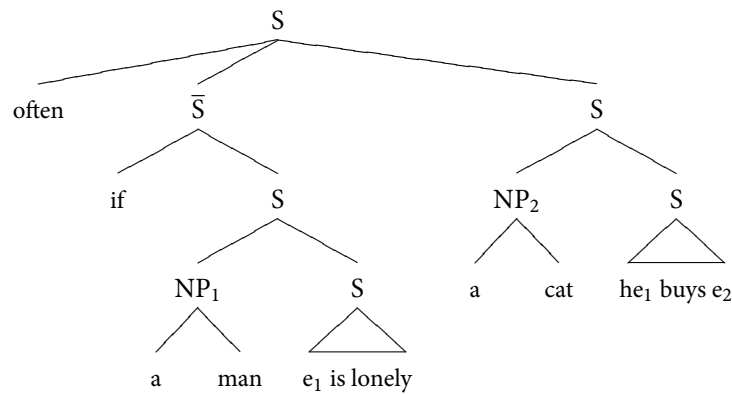
In both cases, the indefinite "a cat" is part of the third term of a tripartite construction headed by a quantifier, i.e., it is part of the "nuclear scope" of the quantifier, as I will henceforth call the third term of any such construction. (The "scope" simpliciter of a quantifier comprises both its restrictive term and its nuclear scope, in my terminology.)

   With regard to the meaning of (5) and (6), we observe that there is existential quantification over cats, and that the existential quantifier is understood to have narrower scope than the quantifiers "every" and "often," respectively.   (To be accurate, this observation applies to the preferred reading of each sentence; see below for the marginal readings.) I therefore introduce another rule of construal, called Existential Closure, which divides into two subrules. One of them is: *Adjoin a quantifier ∃ to the nuclear scope of every quantifier.* This rule is obligatory whenever applicable and dictates that (5′) and (6′) must be converted into (5″) and (6″).

(5″)

```
                          S
          ┌───────────────┼───────────────┐
        every           NP₁              S
                         △        ┌───────┴───────┐
                        ...       ∃             S
                                        ┌────────┴────────┐
                                      NP₂               S
                                       △                △
                                      ...          e₁ saw e₂
```

S-constituents of the form $[\exists, S]$

(6″)

```
                          S
          ┌───────────────┼───────────────┐
        often           S̄               S
                         △        ┌───────┴───────┐
                        ...       ∃             S
                                        ┌────────┴────────┐
                                      NP₂               S
                                       △                △
                                      ...          he₁ buys e₂
```
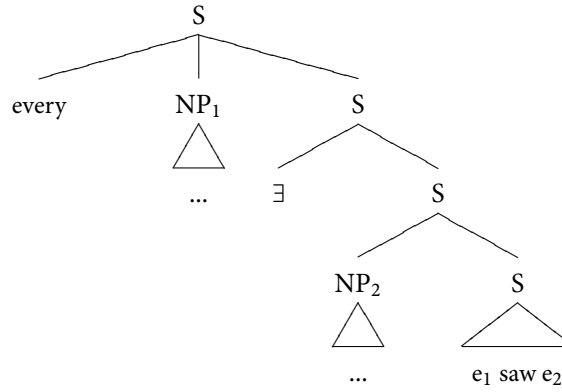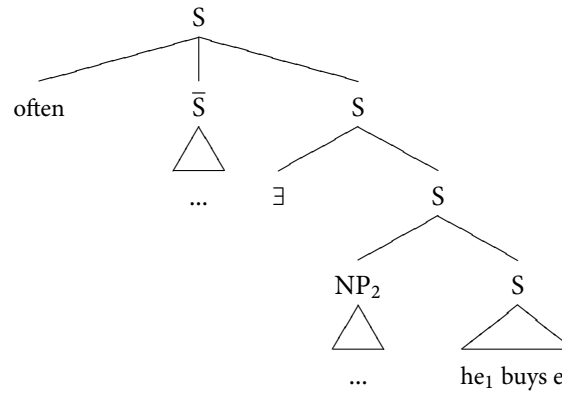
S-constituents of the form [∃,S] will of course be interpreted as existentially quantified sentences. It would be too simple, however, to let ∃ unselectively bind all the free variables in its scope; rather, we want it to bind any number of indefinites, but spare variables like "e_1" in (5″) and "he_1" in (6″), which must remain available for binding from farther outside. I will come back to the problem of selecting the right variables for binding shortly (see the rule of Quantifier Indexing, to be introduced below).

The other subrule of Existential Closure is designed to take care of indefinites in unembedded sentences. In order to formulate it, I have to make reference to expressions larger than sentences, namely *texts*, which are sequences of sentences of unlimited length. The reason for this is that I intend to interpret intersentential anaphoric relations between indefinites and pronouns in the same way as intrasentential ones: The indefinite antecedent and its pronominal anaphor will always be treated as identical variables, bound by the same quantifier. This holds for an example like (4) w.r.t. the anaphorically related pair "a restaurant" — "it" and the quantifier

"always" — and it should hold analogously for example (7), in this case w.r.t. the anaphoric pair "a restaurant" — "it" and the quantifier ∃.

(7)   He went to a restaurant. It was expensive.

Let us assume that complete logical forms correspond to texts, and the S-dominated structures we have so far considered are merely substructures thereof. We may assume that there is a rule of construal called Text Formation, which says: *Attach a sequence of sentences under a T-node.* By this rule and some of our other ones, we can derive the following (intermediate) representation of (7'):

(7')

```
                              T
                  ┌───────────┴───────────┐
                  S                       S
          ┌───────┴───────┐            ┌──┴──┐
        NP₁              S          it₁ was expensive
      ┌──┴──┐          ┌─┴─┐
      a   restaurant  he₂ went to e₁
```
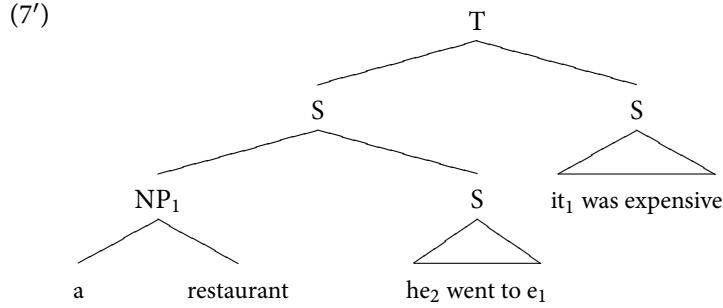
T-constituents of the form $S_1,...,S_n$ will be assigned a conjunctive interpretation, i.e., the T-constituent is true just in case all its daughters are true. But (7') is not yet the logical form of the intended reading of (7), because it still has to undergo the second subrule of Existential Closure, which is obligatory and says: *Adjoin the quantifier ∃ to T.* This will turn (7') into (7''):

(7'')

```
                      T
              ┌───────┴───────────┐
              ∃                   T
                          ┌───────┴───────┐
                          S               S
                  ┌───────┴───────┐    it₁ was expensive
                NP₁              S
              ┌──┴──┐          ┌─┴─┐
              a  restaurant  he₂ went to e₁
```
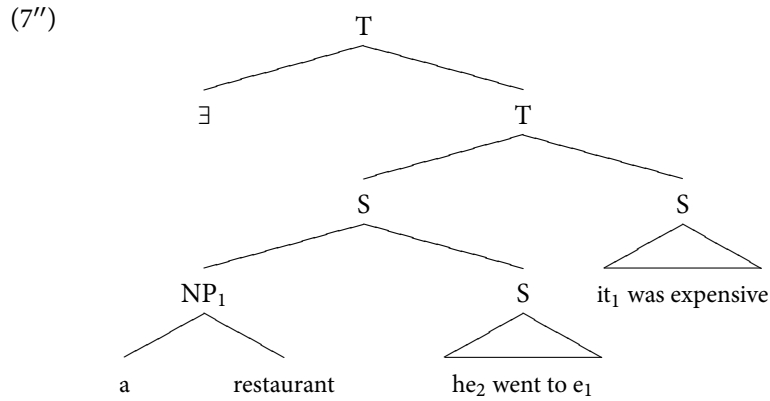
I will assign this an interpretation equivalent to that of:

$$\exists x_1((\text{restaurant}(x_1) \land \text{went-to}(x_2,x_1)) \land \text{expensive}(x_1)).$$

In order to predict this interpretation, it will once again be crucial to assure that ∃ correctly selects the variables it is supposed to bind, but leaves others free, in this case the variable "he$_2$," (which eventually must receive a deictic reading, a phenomenon I will talk about later).

It might not be obvious that the two subrules given are sufficient to generate all existential readings of indefinites that we observe. Is it not possible to read indefinites existentially in at least some cases where they occur in the restrictive term of a quantifier? An apparent case in point would be (8):

(8)   Every man who owns a donkey is rich.

Presumably, this means that every man such that there is a donkey he owns is rich, so we seem to have existential quantification over donkeys with scope inside the relative clause. Existential Closure, as it has been formulated, does not apply to this relative clause, and a third subrule seems to be called for. However, a closer look at the example shows that it does not necessarily involve any existential quantification at all. Consider the logical form we generate by the rules given so far:

(8′)
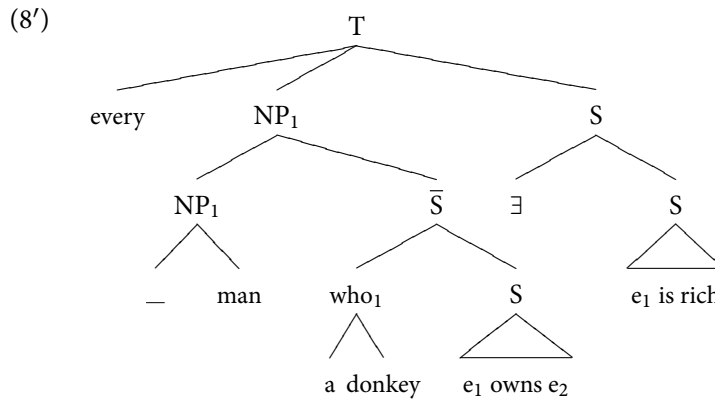
If only we can assume that the quantifier "every" is more or less unselective (see below for how unselective exactly it is), then we should expect it to bind not only the variable over men, but simultaneously the variable over donkeys. We will thus get universal quantification over pairs, as in the following formula:

$$\forall x_1 \forall x_2((man(x_1) \wedge donkey(x_2) \wedge own(x_1,x_2)) \rightarrow rich(x_1))$$

And this happens to be logically equivalent to:

$$\forall x_1((man(x_1) \wedge \exists x_2(donkey(x_2) \wedge own(x_1,x_2))) \rightarrow rich(x_1))$$

In other words: The indefinite does not really have an existential reading in (8); it has a wide-scope universal reading which happens to be equivalent to a narrow-scope existential reading. I claim that a similar state of affairs obtains in any other example that might be thought to show that Existential Closure sometimes applies to the restrictive term of a quantifier.

There is one type of example in which we want Existential Closure to apply and which our current formulation does not quite cover. Take a negated sentence like (9), under the reading where negation has widest scope:

(9)   John didn't buy a cat.

If we are to predict the correct truth conditions for that reading, the logical form must contain an existential quantifier that binds the indefinite and that has narrower scope than the negation operator. For this example, we want a logical form like (9′):

(9′)

$$
\begin{array}{c}
\text{S} \\
\diagdown \\
\text{not} \quad \text{S} \\
\\
\exists \quad \text{S} \\
\\
\text{NP}_1 \quad \text{S} \\
\\
\text{John} \quad \text{NP}_2 \quad \text{S} \\
\\
\text{a cat} \quad e_1 \text{ did buy } e_2
\end{array}
$$
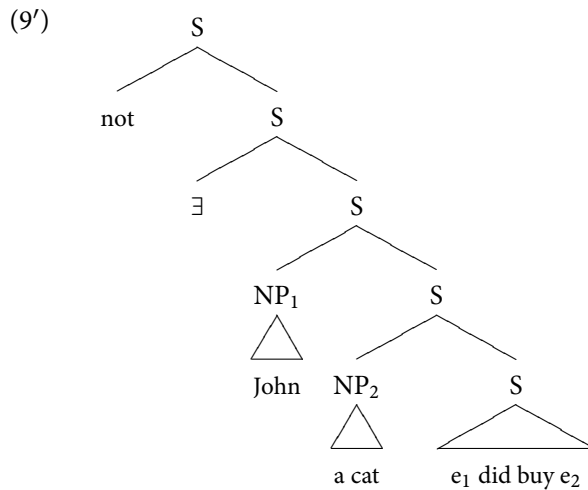
(Disregard once more the problem of getting the existential quantifier in (9′) to bind just variable number 2, not both 1 and 2.) As our construal rules stand, they do not generate (9′), but they will if we make just a minimal change: wherever we have referred to a "quantifier," we write "operator" instead. We take operators to include quantifiers, negation, and temporal and modal operators (which are in some sense quantifiers, i.e., quantifiers over times and possible worlds). If an operator takes two formulas as operators, we call the first its restrictive term and the second its nuclear scope, as specified above. If it takes only one argument-formula, we call that the nuclear scope and assume that there is no restrictive term. Given this, we

not only predict that Operator Construal (formerly: Quantifier Construal) will affect the "not" and adjoin it to its scope, but we also predict that the first subrule of Existential Closure will insert ∃ under "not."

Up to now, I have drawn structures which look as if all quantifiers were strictly unselective. This is usually not true: Not only does ∃ leave some of the variables in its scope unbound (as we saw in examples (5) through (9)), but the other quantifiers, even the adverbial ones, are somewhat selective, too. Consider (10), represented as (10′):

(10)   She knows a man who, if he likes a cat, always buys it.

(10′)

```
                              S
                ┌─────────────┴──────────────┐
              NP₁                             S
           ┌───┴───┐                         △
           a       N̄                    she₂ knows e₁
                ┌──┴───┐
               man     S̄
                    ┌──┴───┐
                  who₁     S
                        ┌───┴────────────┐
           always      S̄                 S
                    ┌───┴───┐         ┌───┴───┐
                   if       S         ∃       S
                        ┌───┴───┐            △
                      NP₃       S       e₁ buys it₃
                   ┌───┴───┐    △
                   a      cat  he₁ likes e₃
```
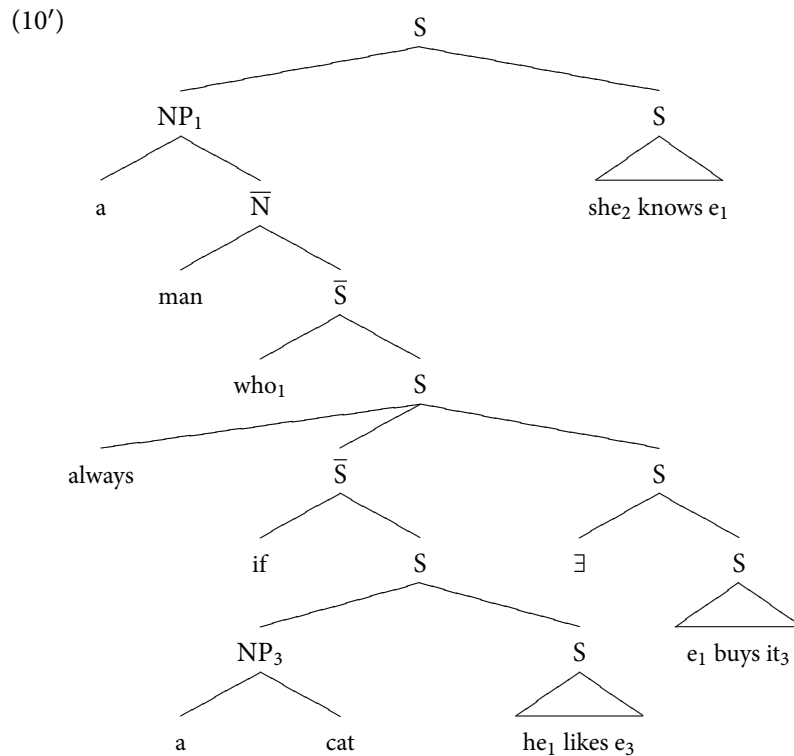
(As for the treatment of relative clauses, just take the interpretation of the S̄ to be identical to that of the S below, as though the "who₁" were not there at all. That will do for now.) The point of the example is that the quantifier "always" must be prevented from binding the variable "he₁" (and the identical variables "who₁" and "e₁" of course). A similar example appears in Lewis (1975), together with a remark that accordingly qualifies the statement that adverbial quantifiers are unselective: "they can bind ...

indefinitely many free variables in the modified sentence, but some variables — the ones used to quantify past the adverbs — remain unbound" (pp. 7-8).

I will represent what selectivity there is in quantifiers by enriching my notation to include "*selection indices*." These are, like the referential indices of NPs, numerical subscripts, except that they appear on quantifiers and that there may be more than one on the same node. The intended interpretation is that a quantifier binds all and only those variables whose referential indices match one of the quantifier's selection indices. Selection indices are assigned in two ways: First, when a quantifier is moved out of an NP as it undergoes Quantifier Construal, it takes the referential index of that NP with it as a selection index. For instance, when Quantifier Construal applies to the "every" in (3) (cf. above), the result is not simply (3″), but, more accurately, (3‴):

(3‴)

$$
\begin{array}{c}
\text{S} \\
\diagup \quad | \quad \diagdown \\
\text{every}_2 \qquad \text{NP}_2 \qquad \text{S} \\
\diagup \diagdown \qquad \diagup \diagdown \\
\underline{\quad} \quad \text{man} \quad \exists \quad \text{S} \\
\diagup \diagdown \\
\text{e}_2 \text{ arrived}
\end{array}
$$
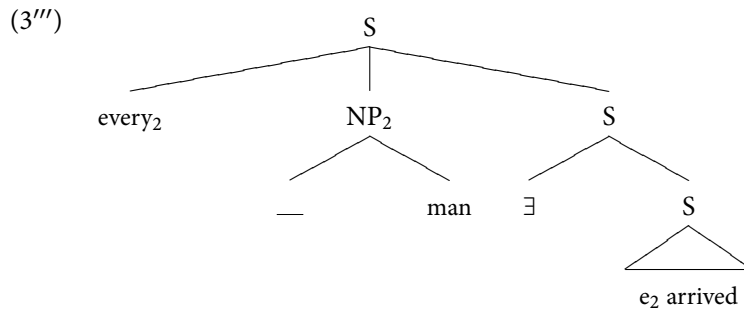
Second, selection indices get assigned by Quantifier Indexing, which is another obligatory rule of construal: *Copy the referential index of every indefinite NP as a selection index onto the lowest c-commanding quantifier.* By "the lowest c-commanding quantifier" I mean that one among all the quantifiers that c-command the indefinite which does not c-command any other one. Quantifier Indexing — or rather, as it should be called: Operator Indexing — applies to quantifying adverbs, quantifying determiners, $\exists$, and whatever other operators there may be, alike. For example, (6″) turns into (6‴).

(6‴)

```
                    S
        ┌───────────┼──────────────────────────┐
      often₁        S̄                           S
              ┌─────┴─────┐              ┌───────┴───────┐
             if           S            ∃₂               S
                    ┌──────┴──────┐            ┌──────────┴──────────┐
                  NP₁             S          NP₂                    S
                 ┌──┴──┐        ╱────╲      ┌──┴──┐             ╱──────╲
                a    man    e₁ is lonely   a    cat        he₁ buys e₂
```
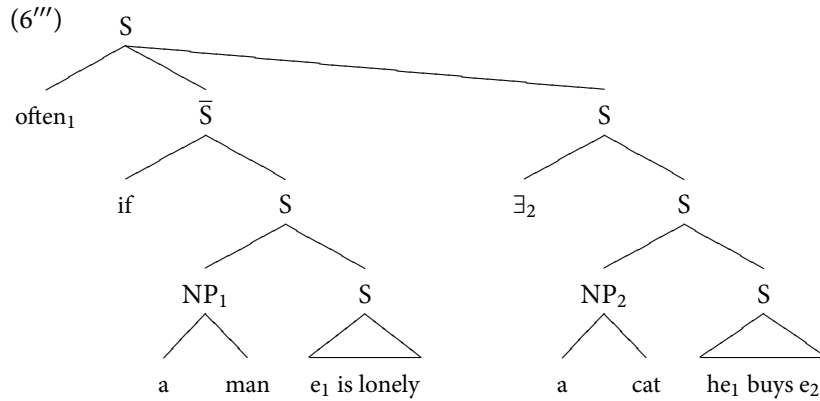
(7″) is affected in such a way that ∃ receives 1, but not 2, as a subscript. In the case of (10′), only 3 gets coupled onto "always" (whereas 1 will end up on some quantifier in the larger structure that (10′) is part of). In all these examples, the predicted selection indices reflect our intuitions about the binding relations.

The two ways in which quantifiers get their selection indices — by bringing them along under movement, and by having them assigned under Quantifier Indexing — might turn out to be non-distinct. Perhaps NPs with empty determiners, such as those left behind when Quantifier Construal has applied, qualify as indefinite. If that is so, then Quantifier Indexing alone can be held responsible for all selectional indices that quantifiers end up with, even the index 2 on "every" in example (3‴). We may leave it open for the time being whether this is the right way to look at things, or whether Quantifier Construal perhaps leaves behind NPs that are neither + nor - definite.

Notice that there is no direct way for a pronoun to impose its referential index onto a quantifier. The only way for a pronoun to get bound is indirectly, by virtue of its being anaphoric to, and thus coindexed with, another NP. For example, the "he₁" in (6‴) is bound by "often" only because it is coindexed with "a man₁" which is indefinite and therefore eligible for Quantifier Indexing. Notice further that some instances of 3 may not get any selection indices at all. This means that they bind nothing, i.e., are semantically vacuous. (3‴) provides an example of this.

Was it correct to write into the rule of Quantifier Indexing that, without exception, it is the *lowest* quantifier that picks up the subscript of an indefinite? Doesn't this rule out any chance for an indefinite to be read existentially with wider scope than the narrowest possible? We will see that it does not. A case in point would be one of the (nonpreferred?) readings of (11), i.e., the

reading which could be paraphrased: "there is a friend of mine such that, if a cat likes him, I always give it to him."

(11)   If a cat likes a friend of mine I always give it to him.

If we liberalized Quantifier Indexing in such a way that the quantifier involved need not be the lowest one, then we could, inter alia, derive the following disambiguated representation for (11):

(11)   a.
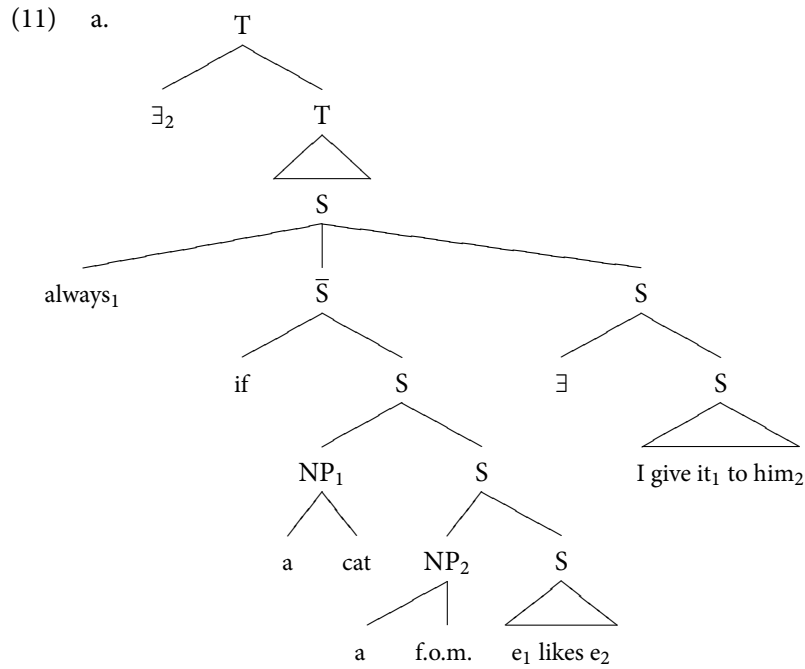
Upon closer inspection, however, (11a) does not have the intended interpretation, and its actual interpretation is not available at all for the English sentence (11). Moreover, it turns out that a representation which *has* the intended interpretation can, after all, be derived without deviating from the strict formulation of Quantifier Indexing that I have given. The problem about (11a) is that it receives the same interpretation as the predicate-calculus formula:

$$\exists x_2 \forall x_1 ((\text{cat}(x_1) \wedge \text{f.o.m.}(x_2) \wedge \text{like}(x_1,x_2)) \rightarrow \text{give}(I,x_1,x_2)),$$

which has very weak truth conditions: It is verified by the mere existence of something that is not a friend of mine. On the other hand, we can derive representation (11b) for (11) by applying NP-Prefixing to "a friend of mine"

in such a way that it is not prefixed to the minimal S that contains it, but rather to the matrix S.

(11) b.

```
                              T
                       ┌──────┴──────┐
                      ∃₂            T
                                   ╱─╲
                                  ╱───╲
                                   S
                          ┌────────┴────────┐
                        NP₂                 S
                        ╱  ╲         ┌───────┼──────────┐
                       a  f.o.m.  always₁  S̄            S
                                      ┌────┴────┐    ┌───┴───┐
                                     if        S    ∃       S
                                          ┌────┴────┐       ╱─╲
                                        NP₁         S   I give it₁ to him₂
                                        ╱─╲        ╱─╲
                                     a cat     e₁ likes e₂
```
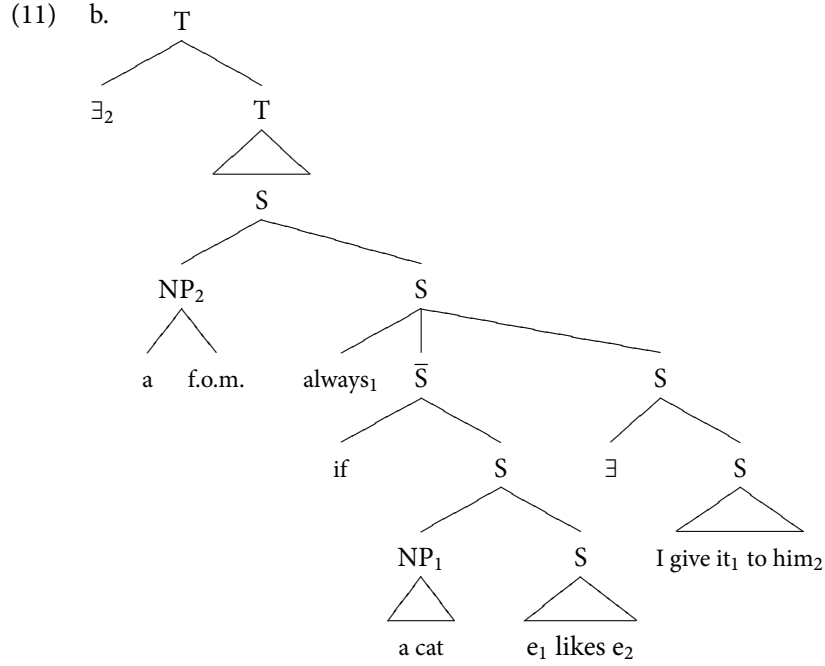
(11b), which conforms to my formulation of Quantifier Indexing, gets an interpretation equivalent to:

$$\exists x_2(\text{f.o.m.}(x_2) \wedge \forall x_1((\text{cat}(x_1) \wedge \text{like}(x_1,x_2)) \rightarrow \text{give}(I,x_1,x_2))).$$

This is the intended reading. We have thus seen that examples with wide scope existential readings of indefinites provide evidence for, not against, the deterministic formulation of Quantifier Indexing that I have chosen. Scope ambiguities generally result from multiple options in applying NP-Prefixing, not in applying Quantifier Indexing.

Having concluded the exposition of the construal rules in my system, I now introduce a condition on wellformedness that filters out certain logical forms which are both derivable by the construal rules from wellformed syntactic structures and interpretable by the interpretation rules to be given below, yet fail to represent grammatical readings for the sentences in question. An example is structure (12′), which is derivable from the syntactic structure of sentence (12).

(12) He likes a cat and she hates a cat.

(12′)

```
                        T
                       / \
                      /   \
                   ∃₁      T
                          /△\
                          S
                     /    |    \
                    /     |     \
                   S     and     S
                  / \          /   \
               NP₁   S      NP₁     S
               /△\  /△\     /△\    /△\
              a cat he₂    a  cat she₃
                   likes e₁      hates e₁
```
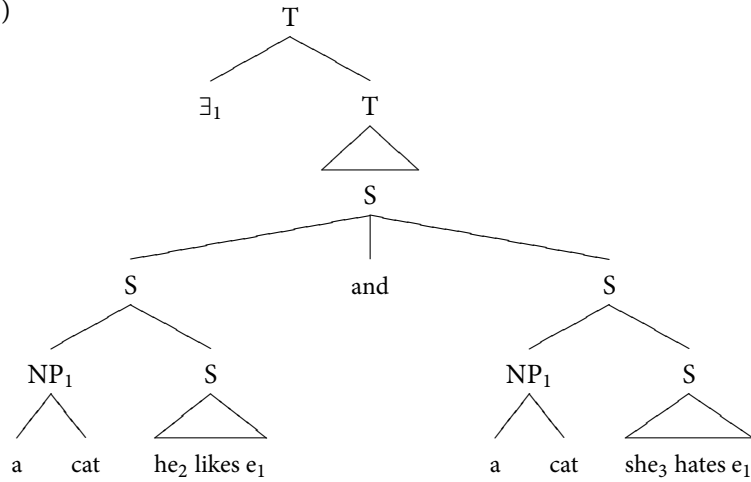
Since the two occurrences of "a cat" are coindexed, (12′) means that he likes a cat and she hates *the same cat*. This is not a possible reading for (12). Note that I am not saying that a state of affairs in which he likes a cat and she hates the same cat is incompatible with the meaning of (12). I am merely denying that (12) has any reading under which it would be incompatible with the two cats being different.

We would therefore want to somehow rule out the option of putting identical referential indices on the two occurrences of "a cat" in (12). It is of interest that this is not ruled out by any of the known prohibitions on coindexing, such as Disjoint Reference and Non-Coreference. (I am tacitly assuming that some version of these two, as well as other conditions on "binding" that have been discovered, are in force in my system.) The example at hand involves a new constraint, which, unlike the ones just mentioned, applies specifically to indefinites. I call it the Novelty Condition:[3] *An indefinite NP must not have the same referential index as any NP to its left.* In other words, an indefinite must always carry a "new" referential index, i.e., one that has not yet been used as the referential index of any other NP earlier in the same text. The condition applies regardless of the distance between the two NPs involved, as long as they are in the same text. In example (12), both of the NPs are indefinite, but this is not necessarily so. We also want to rule out coindexing of an indefinite with an earlier definite, as in (13):

---

[3] Not to be confused with something that Wasow (1972) called the "Novelty Constraint." An analogue of the latter will appear in Chapter III, section 5 below.

(13)   He likes $\left\{ \begin{array}{c} \text{the cat}_i \\ \text{it}_i \end{array} \right\}$, and she hates a cat$_j$.   *i=j

Contrast this with an example where the definite comes after the indefinite:

(14)   He likes a cat$_i$, and she hates $\left\{ \begin{array}{c} \text{the cat}_j \\ \text{it}_j \end{array} \right\}$.

In this case, the Novelty Condition does not apply, and nothing rules out i=j. This is as it should be, since an anaphoric reading of the definite is indeed acceptable.

3   Semantic Interpretation of Logical Form

I will spell out in a somewhat more precise fashion the informal remarks that I have so far made about the intended interpretations for my logical forms. In this connection I will introduce the notions of satisfaction, truth, and felicity.

3.1   *Semantic categorization of elements of logical form*

Logical forms, the representations on which semantic interpretation is defined, are unlike the formulas that logicians employ in that they contain a lot of things that are irrelevant to semantic interpretation. For instance, pronouns look different from empty NPs in logical form, but semantic interpretation is insensitive to the difference and treats all variables alike. Similarly, logical forms contain information about syntactic category, e.g., in showing whether a quantifier is a determiner or an adverb, but this information is not needed to determine the semantic interpretation. Other "irrelevant" characteristics of logical forms include: whether there is a definite or an indefinite article; whether the determiner position is empty or filled; and whether a sentence is or is not preceded by "if." Of course, all these things have a good reason to be there: logical forms derive via construal from syntactic representations and thereby automatically inherit properties which are relevant from the point of view of the syntactic component.

Moreover, some of the features that are irrelevant to the component of semantic interpretation are crucial for the wellformedness of the logical forms in which they occur. A case in point is definiteness, which is referred to by the Novelty Condition. (Note in this connection that I distinguish between wellformedness and interpretability: a logical form that is illformed by the Novelty Condition may still have a welldefined interpretation; on the other hand, failure to apply an optional construal rule, say Quantifier

Construal, can result in an uninterpretable, though not ill formed, logical form.)

Having talked about features that play *no* role in semantic interpretation, I must now specify those that do. I assume that the constituents of logical forms may fall into certain semantic categories, such as "variable," "predicate," "quantifier," and "formula." These are different from, and cut across, the syntactic categories. Some constituents fall under no semantic category at all. I take it that information about semantic category membership can be read off logical forms, along with information about syntactic category membership and various other information. In some cases, it is simply part of the lexicon entry for a terminal symbol that it belongs to such and such semantic category. For instance, "every" is a quantifier, "cat" is a predicate, and "buy" is also a predicate, and these are lexical properties of those words. In other cases, it appears as though semantic categories are assigned to constituents of logical forms by way of rules of one sort or another, such as: "mark all empty NPs variables." I will leave it open just how it is that semantic category information gets to be present in logical form. I simply rely on it being present, and specifically, on the following elements being semantically categorized as indicated:

(a) **Variables**: pronominal NPs with indices; empty NPs with indices; indices on predicates.

(b) **Predicates**: nouns (both common and proper nouns, but not pronouns); verbs.

(c) **Quantifiers and other operators**: "every$_{7,5,3}$," "always," "$\exists_2$," "not," etc. (The selectional indices count as part of the operator.)

(d) **No semantic category**: "if," "the," "a," the empty determiner, indices that are not variables, and other things.

Let me give some illustrations. A logical form for the simple sentence (1)

(1)   She met a cat,

may look as follows:
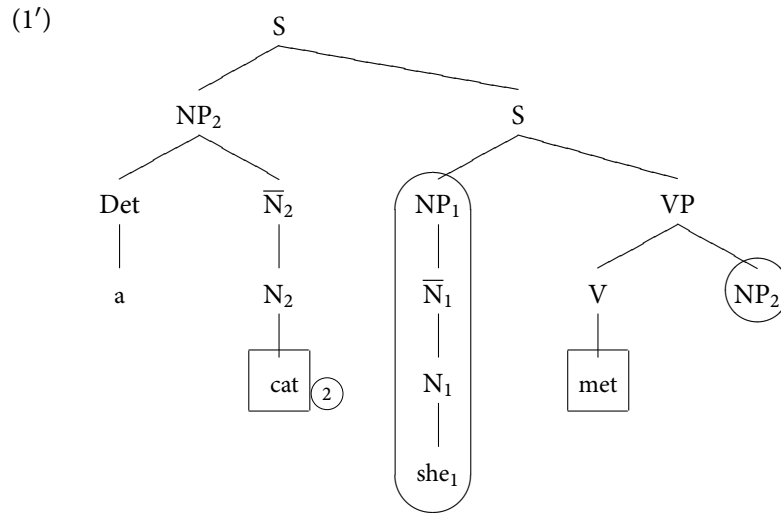
(I have spelled out some more detail here than in the sketches of logical forms I use elsewhere in this dissertation, but (1′) still contains simplifications.) (1′) illustrates an assumption that I have not made explicit so far: The referential index of an NP always "percolates" down to its lexical head, i.e., each node is coindexed with its head. (1′) contains an example of each of the first three kinds of variables that I have listed under (a). I have circled them. In two cases, the variable is an NP; in the third case, it is just an index. (1′) also contains two predicates, which I have boxed.

   Of what semantic category are the other constituents? Some are of none at all, e.g., the Det-constituent and various occurrences of indices. Others are formulas. Formulas come in two varieties: atomic and molecular.

   (e) **Atomic formulas**: minimal constituents that include an n-place predicate and n variables.

   (f) **Molecular formulas**: constituents that have one or more formulas as immediate constituents.

Here is another picture of the logical form (1′), this time with the atomic formulas circled and the molecular formulas boxed:

(1′)

```
                    S
              _____
        NP₂                    S
     ___|___              ____|____
    Det    N̄₂          NP₁        VP
     |      |            |       __|__
     a     N₂           N̄₁      V     NP₂
            |            |       |
          (cat₂)        N₁      met
                         |
                        she₁
```
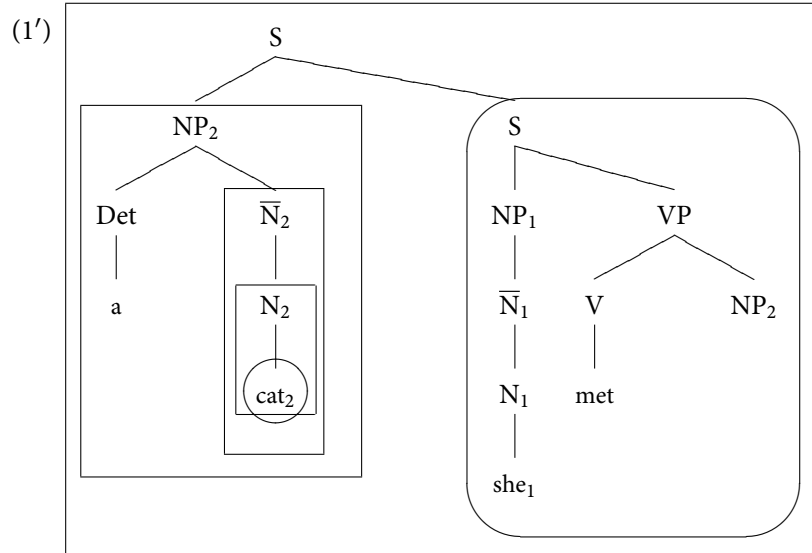
Notice that, to identify the atomic formulas, we must of course know the polyadicity of each predicate, which I presume is specified by the lexicon entry for that predicate. We then can find the smallest constituent which includes the predicate and enough variables to fill all its argument places, and this will be an atomic formula. As (1′) shows, atomic formulas may be either "nominal," i.e., centered around a predicate which is a noun, or "verbal," i.e., centered around a predicate which is a verb.

All the molecular formulas in (1′) are of the type that I will call *cumulative molecular formulas*.[4] There is another type, the *quantified molecular formulas*. They are molecular formulas whose leftmost immediate constituent is a quantifier. Actually, I should use the more general term "operator-headed molecular formulas," because there is really no relevant distinction to be made between cases involving a quantifier and cases involving negation or another kind of operator.

We are now prepared to turn to the rules of semantic interpretation, which make reference to analyses of logical forms in terms of semantic categories along the lines I have outlined here. Just one more piece of terminology: When I speak of "the index of a variable," I mean either the

---

[4] Why do I call them "cumulative?" Because the information they express is in some sense determined by accumulating the information expressed by all their immediate constituents. Chapter III, especially section 1.5, will give a clearer sense of how this process of accumulation could be described.

variable itself, if it consists just of an index, or the index of the NP, if the variable is an NP.

3.2 *An extensional semantics, based on satisfaction*

Following Tarski (1936), I will first give a recursive assignment of satisfaction conditions to formulas, and then define truth in terms of satisfaction. Satisfaction is a relation between infinite sequences of individuals on the one hand, and formulas on the other. (Infinite sequences of individuals are practically the same thing as variable assignments.) Satisfaction is always relative to a model. A model for English is a pair $\langle A, Ext \rangle$, where A is a set (the domain of individuals) and Ext is a function which assigns to every predicate of English an extension. In other words:

$$\text{If } \zeta \text{ is an n-place predicate, then } Ext(\zeta) \subseteq \underbrace{Ax...xA}_{n \text{ times}}.$$

Infinite sequences of individuals of A are functions whose domain is N (the set of natural numbers) and whose values are in A. The set of all such sequences is A^N. I use a_N, b_N, c_N etc., to refer to members of A^N.

Sometimes it is useful to think of a model as determined by two separate factors: (i) a certain state of affairs, and (ii) a certain interpretation of the language. Let W be the set of all possible states of affairs, or "possible worlds," as one calls them. Take an interpretation for English to be a function, Int, which assigns each predicate of English an intension. Suppose intensions for predicates are properties, i.e., functions that assign to each possible world an extension (the extension being a set of n-tuples of individuals, as above). Then each pair $\langle w, Int \rangle$ of a possible world $w \in W$ and an interpretation for English, Int determines a model $\langle A, Ext \rangle$ in the following way: Identify A with the set of all individuals that inhabit w, and let Ext be defined thus: $Ext(\zeta) = Int(\zeta)(w)$, for all predicates ζ. We can pick out a particular model in this way, if we choose as w the actual world and let Int represent the meanings that English predicates actually have in English. The model determined by this choice of w and Int may be called the "actual model."

Given this background, what is the task that our rules of semantic interpretation are to perform? We want them to assign satisfaction conditions to the formulas that make up the logical forms of English texts. For a given model and a given formula ϕ, the rules of interpretation should tell us which sequences satisfy ϕ with respect to that model. Put differently, we want the rules of interpretation to jointly define the relation "x satisfies y with respect to $\langle A, Ext \rangle$" (abbreviated: "x $\underline{SAT}_{A,Ext}$ y"), and the definition ought to apply to formulas of unlimited complexity, up to the complete logical

forms of the longest English texts. The following five rules accomplish this task for a small fragment of English. They jointly amount to a recursive definition of the satisfaction relation, recursive insofar as the rules which assign satisfaction conditions to molecular formulas make reference to the satisfaction conditions of their constituent formulas.

Let a model $\langle A, Ext \rangle$ for English be given:

(i) Let ϕ be an atomic formula, consisting of an n-place predicate ζ and an n-tuple of variables $\langle \alpha^1,..., \alpha^n \rangle$ whose indices are $i_1,...i_n$, respectively. Then, for any sequence $a_N \in A^N$:

$$a_N \ \underline{SAT}_{A,Ext} \ \phi \ \text{iff} \ \langle a_{i_1},...,a_{i_n} \rangle \in Ext(\zeta).$$

(Here and elsewhere, "a_k" means "the k-th individual in the sequence a_N," i.e., $a_k = a_N(k)$.)

(ii) Let ϕ be a cumulative molecular formula with the immediate constituent formulas $\phi^1,...,\phi^n$ (in that order). Then, for any $a_N \in A^N$:

$$a_N \ \underline{SAT}_{A,Ext} \ \phi \ \text{iff for all} \ i \in \{1,...,n\}: a_N \ \underline{SAT}_{A,Ext} \ \phi^i.$$

(iii) Let ϕ be a quantified molecular formula, consisting of a universal quantifier with the selectional indices $i_1,..,i_n$, and of the two formulas ϕ^1 and ϕ^2 (in that order). Then, for any $a_N \in A^N$:

$$a_N \ \underline{SAT}_{A,Ext} \ \phi \ \text{iff for every sequence} \ b_N \ \text{that agrees with} \ a_N \ \text{on all} \ i \notin \{i_1,...,i_n\}: \text{if} \ b_N \ \underline{SAT}_{A,Ext} \ \phi^1, \text{then} \ b_N \ \underline{SAT}_{A,Ext} \ \phi^2.$$

(For b_N to "agree" with a_N on a number i means that $b_i = a_i$.)

(iv) Let ϕ be a quantified molecular formula, consisting of an existential quantifier with the selectional indices $i_1,..,i_n$, and of the formula ψ. Then, for any $a_N \in A^N$:

$$a_N \ \underline{SAT}_{A,Ext} \ \phi \ \text{iff there is a sequence} \ b_N \ \text{that agrees with} \ a_N \ \text{on all} \ i \notin \{i_1,...,i_n\} \ \text{such that} \ b_N \ \underline{SAT}_{A,Ext} \ \psi.$$

(v) Let ϕ be an operator-headed molecular formula, consisting of a negator and the formula ψ. Then, for any $a_N \in A^N$:

$$a_N \ \underline{SAT}_{A,Ext} \ \phi \ \text{iff it is not the case that} \ a_N \ \underline{SAT}_{A,Ext} \ \psi.$$

For example, our logical form (1′) will get interpreted as follows by these rules: The atomic formula "cat_2" (circled in the second picture of (1′), cf. above) consists of the predicate "cat" and the 1-tuple of variables ⟨"2"⟩. The index of the one variable in that 1-tuple is 2. So by rule (i), we have:

$$a_N \ \underline{SAT}_{A,Ext} \ \text{"cat2"} \ \text{iff} \ a_2 \in Ext(\text{"cat"}).$$

Next, consider the molecular formula dominated by the node "N_2". This is a cumulative molecular formula which consists of just one constituent formula, viz., the atomic formula "cat2". Rule (ii) applies and we have:

$$a_N \ \underline{SAT}_{A,Ext} \quad \begin{matrix} N_2 \\ | \\ cat_2 \end{matrix} \quad \text{iff} \ a_2 \in Ext(\text{"cat"}).$$

In other words, this formula has the same satisfaction condition as its constituent formula. Analogously, the formulas dominated by the nodes "$\overline{N}_2$" and "NP_2" will inherit the satisfaction condition of "cat_2" unchanged, because they, too, are cumulative molecular formulas with just one formula among their immediate constituents. (NP_2 also dominates the determiner, but that makes no difference.) So we arrive at:

$$a_N \ \underline{SAT}_{A,Ext} \quad \begin{matrix} NP_2 \\ \triangle \\ a \ cat_2 \end{matrix} \quad \text{iff} \ a_2 \in Ext(\text{"cat"}).$$

Now consider the other atomic formula that (1′) contains, the one dominated by S with the terminal string "she_1 met e_2". This atomic formula consists of the predicate "met" and the pair (2-tuple) of variables ⟨"she_1," "e_2"⟩, whose indices are 1 and 2, respectively.[5] Rule (i) thus yields:

$$a_N \ \underline{SAT}_{A,Ext} \quad \begin{matrix} S \\ \triangle \\ she_1 \ met \ e_2 \end{matrix} \quad \text{iff} \ \langle a_1, a_2 \rangle \in Ext(\text{"met"}).$$

Finally, we apply rule (ii) to the entire tree, which is a cumulative molecular formula with two constituent formulas, "$[_{NP2} \ a \ cat_2]$" and "$[_S \ she_1 \ met \ e_2.]$" The result is:

---

[5] In saying that the formula contains the pair of variables ⟨"she_1," "e_2"⟩, rather than the pair of variables ⟨"e_2," "she_1"⟩, I am presupposing that we can somehow read off logical form which variable belongs into which of the predicate's argument slots. The principles by which this is determined are not under consideration in this dissertation.

a_N $\underline{SAT}_{A,Ext}$ (1′) iff both $a_2 \in Ext(\text{"cat"})$ and $\langle a_1, a_2 \rangle \in Ext(\text{"met"})$.

In other words, (1′) is satisfied by any sequence whose second member is in the extension of "cat" and whose first and second member are a pair which is in the extension of "met." If $\langle A, Ext \rangle$ is the actual model, these are the sequences whose second member is a cat and whose first member met its second member.

3.3 From satisfaction to truth

As long as we consider only closed formulas (i.e., those with no variables free in them), we can choose between several equivalent ways of defining truth in terms of satisfaction. For instance, we may say that ϕ is true if there is at least one sequence that satisfies it (and false otherwise). Or we may follow Tarski (1936) in saying that ϕ is true iff every sequence satisfies it. Or we might pick some particular sequence, say the one that has the sun as its n-th member for all n, and define that ϕ is true iff that particular sequence satisfies it. It all comes to the same thing for closed formulas. Take for instance the logical form of the monosentential text (2), (2′), which is a closed formula.

(2)   A cat arrived.

(2′)
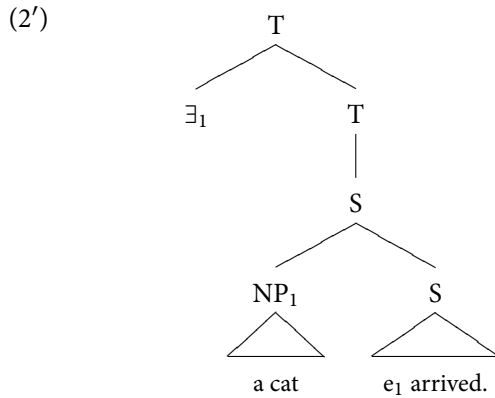
By our rules of interpretation, we determine that for any sequence a_N:

a_N $\underline{SAT}_{A,Ext}$ (2′) iff there is a $c \in A$ such that $c \in Ext(\text{"cat"})$ and $c \in Ext(\text{"arrived"})$.

It is apparent that if any sequence satisfies (2′) at all, then all sequences do. Therefore any of the truth-definitions I have alluded to would result in the same prediction, i.e., that (2′) is true with respect to a model $\langle A, Ext \rangle$ just

in case Ext("cat") and Ext("arrived") have at least one member in common, otherwise it is false w.r.t. $\langle A, Ext \rangle$.

However, matters are not so simple, because we have to take into account those texts whose logical forms contain free variables. Suppose, for example, that (1) is uttered as the only sentence of a text. The text will then have a logical form like (1"):

(1″)

```
                        T
                      /   \
                   ∃₂      T
                           |
                           S    (=(1′))
                         /   \
                      NP₂     S
                      /\      /\
                     /  \    /  \
                  a cat    she₁ met e₂
```
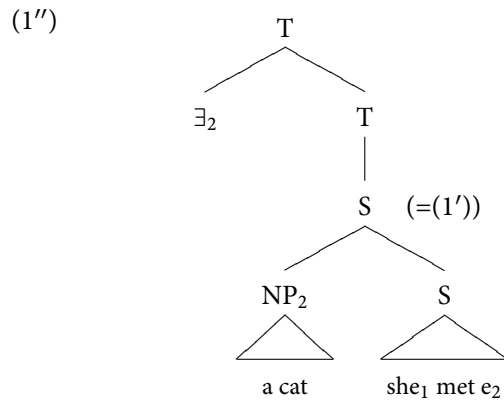
(1″) contains the variable "she₁" free. We do not want to predict that (1') is only true if every sequence satisfies it, because that would require for every individual whatsoever to have met a cat. Nor do we want (1″) to be ruled true as soon as some sequence satisfies it, because then it would suffice for some individual or other to have met a cat. Neither prediction would reflect the way we actually use and understand the text: For (1) to be uttered in isolation, i.e., as a complete text by itself, the speaker must have a definite referent for the pronoun "she" in mind and the context of utterance must somehow be suited to help the audience guess who that referent is. The utterance amounts to a claim that this contextually determined referent of "she" met a cat, and it is judged true or false depending on whether this referent did or did not meet a cat. In other words, a variable that is free in logical form calls for a deictic interpretation, and the truth conditions of formulas with deictically interpreted variables in them depend on the reference of those variables.

The foregoing considerations motivate a context-relative notion of truth, which builds on an ancillary notion of context-relative felicity, in the following manner:

(a) **Felicity**: A formula ϕ is felicitous with respect to a context C and a model $\langle A, Ext \rangle$ only if C furnishes a unique individual $a_{C,i} \in A$ for each number i which is the index of a variable free in ϕ.

(b) **Truth**: If a formula ϕ is felicitous w.r.t. a context C and a model $\langle A, Ext \rangle$, then:

ϕ is true w.r.t. C and $\langle$A, Ext$\rangle$ if there is a sequence $b_N \in A^N$ such that b_N $\underline{SAT}_{A,Ext}$ ϕ, and for each index i of a variable free in ϕ, $b_i = a_{C,i}$;

and ϕ is false w.r.t. C and $\langle$A, Ext$\rangle$ otherwise.

The idea behind (a) is that an utterance will be ruled infelicitous if the context is somehow unsuited to fix referents for all the free variables. I do not want to go into the pragmatics of deictic reference here, and so I will say nothing about the characteristics that a context must have in order to furnish a referent, or about the reasons for which it might fail to furnish one. I wrote (a) merely as a necessary, rather than necessary and sufficient, condition for felicity, since utterances may fall short of felicity for other reasons (say, presupposition failure), even if all their free variables refer properly. As for (b), note that a truth value is defined only for felicitous formulas; infelicitous formulas we take to be neither true nor false.

Like the notion of satisfaction, our defined notions of felicity and truth strictly speaking do not apply to utterances of English texts, but to formulas. In the context of a grammar, however, which links English texts to certain formulas (their logical forms) by way of a construal component, they naturally carry over. Suppose a text T is uttered in a context C under a reading that corresponds to the logical form T′. (Recall that a logical form represents a text under a certain reading.) We may then describe this utterance in terms of the semantic concepts that are defined for the formula T′. For example, if our definitions yield that T′ is felicitous w.r.t. a model $\langle$A, Ext$\rangle$, and that T′ is false w.r.t. C and $\langle$A, Ext$\rangle$, we may derivatively say of the utterance under consideration that it is felicitous w.r.t. $\langle$A, Ext$\rangle$ and false w.r.t. $\langle$A, Ext$\rangle$. This way, we can compare the truth conditions that our theory predicts with the intuitive judgments that speakers of English make about the truth and falsity of utterances in English.

Having introduced a notion of truth, I can now pinpoint a weakness that the approach I am pursuing in this chapter shares with Geach's approach (as discussed in Chapter I, sections 1.2 and 2.1): Our truth definition fails to make adequate predictions about anything but complete texts. Consider, e.g., an utterance of the bisentential text (3) under the reading represented by the logical form (3′).

(3) A cat was at the door. It wanted to be fed.

(3′)

```
                        T
                  ┌─────┴─────┐
                 ∃₁           T
                        ┌──────┴──────┐
                       S¹            S²
                   ┌───┴───┐      ┌───┴───┐
                  NP₁      S     it₁ wanted to be fed
                  /\      /\
                a cat   e₁ was at the door
```
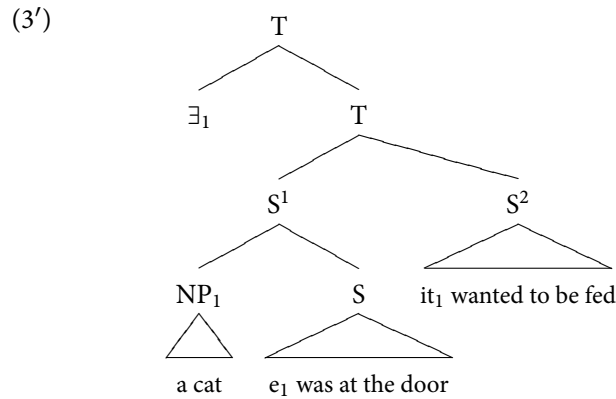
(The superscripts on the S-nodes are just there for easy reference.) As long as we are discussing the truth of this utterance as a whole, the prediction of the theory coincides with intuitive judgment: The utterance as a whole is true (in the actual model) just in case there was a cat at the door that wanted to be fed.

But now suppose we want to discuss separately whether each sentence was uttered truly in the course of the utterance as a whole. In an intuitive sense, it is quite conceivable that the speaker said something true with the first sentence, but something false with the second sentence. We would naturally say this if in fact there was a cat at the door, but no cat that was at the door wanted to be fed. But how does the truth definition apply in such cases? For it to be applicable at all, we must first decide which subformula of (3′) we take to represent the particular subpart of the whole utterance that we are considering, and we must then make sure that subformula is felicitous w.r.t. the context of the utterance. Presumably, if the utterance-part under consideration is the utterance of the sentence "A cat was at the door," then the corresponding subformula of (3′) is the one dominated by S^1. For S^1 to be felicitous in the sense of (a) above, the context must furnish a referent for the variable number 1, which is free in S^1. But texts like (3) are not normally, and certainly not necessarily for them to be felicitous, uttered in situations where the speaker intends to refer to a particular individual that the audience can pick out. So S^1 will be ruled infelicitous in the sense of (a) (even though (3′) as a whole will be ruled felicitous, as it ought to be), and the truth of S^1 is consequently undefined as far as our definition (b) goes. For analogous reasons, (b) will not yield a truth value for S^2, the logical-form-representation of the utterance of "It wanted to be fed."

There is no straightforward way in which (a) and (b) might be amended so as to give predictions which mirror the intuitive judgments we make about

the truth of parts of utterances. I wiII return to the issue in Chapter III, section 3.

4   Invisible Necessity Operators

In section 1, I ended with a strong claim: No indefinite ever has any quantificational force. What appears to be a quantifying reading of an indefinite is always the result of the indefinite's getting bound by a quantifier which is not part of its own meaning. Prima facie, there are more counterexamples to this claim than confirming evidence. One group of apparent counterexamples are those indefinites that are read existentially, without there being any other expression in their environment that could be held responsible for the existential quantificational force. I have eliminated that type of potential counterexamples by stipulating a rule of Existential Closure. Another challenge is posed by simple conditional sentences with an indefinite in the antecedent, e.g., (1):

(1)   If a donkey kicks John, he beats it.

Here the indefinite seems to have universal force, yet the sentence contains no obvious candidate for a universal quantifier, such as an occurrence of "always," or of "every." How, then, can we maintain that the indefinite is just a variable and that some other element in its environment contributes the universal force?

The purpose of this section is to meet this challenge. I will claim that sentences like (1) do, after all, contain sort of a universal quantifier, albeit an invisible one. This sounds like an ad hoc move, designed only to uphold the variable analysis of indefinites at any cost. But it is not. I will refer to research on the semantics of conditionals that has established the existence of such invisible contributors of universal force on grounds completely independent of the semantics of indefinites.

Large parts of this section will be hard, if not impossible, to read for someone who is not familiar to some extent with the semantics of conditionals and modal operators.[6] I could not avoid this, although the point I want to make is a rather simple one and may be summarized in one paragraph as follows:

We observe that indefinites in conditionals like (1) above exhibit universal readings. This is puzzling as long as we think of "if... then" as material

---

[6]   On modal operators, see especially Kratzer (1981b). On conditionals, see Lewis (1973) and numerous other works.

implication. However, the puzzle disappears as we bring together two hypotheses which are both not new, but have been pursued independently of each other. The first hypothesis, of long standing in the study of conditionals, is that "if-then" sentences are not really instances of material implication, but express some sort of conditional necessity: the "then"-clause is read as under the scope of a necessity operator, which in turn is restricted by the "if"-clause. The second hypothesis, implicit in the system of construal and interpretation rules I have set out in the preceding sections, is that operators are basically "unselective," meaning that they attract selectional indices from indefinites in their environment, thereby binding them. The combination of these two hypotheses leads us to expect that an indefinite may in particular get bound by the necessity operator that is part of the analysis of a conditional sentence. Since necessity operators are basically universal in their force (necessity being truth in *every* possible world), an indefinite thus bound by a necessity operator will itself appear to have universal force. The observation that seemed puzzling at first is thus explained.

Unfortunately, one cannot argue quite as straightforwardly as this. It is known that conditional necessity cannot be anything nearly as simple as universal quantification over possible worlds if the analysis of conditionals in terms of conditional necessity is to be maintained with any generality. And it is not at first sight obvious that the universal readings of indefinites in "if"-clauses can still be predicted so naturally when an appropriately sophisticated notion of conditional necessity is assumed. Therefore, I cannot avoid making some reference to recent (and presumably not generally familiar) proposals in the theory of conditionals. I hope I can thereby make it plausible, even to the reader who knows enough to be skeptical, that the observed behavior of indefinites in conditionals can be accounted for without any need for new stipulations.

There are three parts to this section: First I will show how indefinites get bound by overt modal operators, then address, in the second part, the analogous phenomenon in "bare" conditionals, i.e., those conditionals with no overt modalization. As part three, I will attach a sketch of a treatment of generic indefinites, which is loosely related to the primary subject of the section.

4.1 Indefinites bound by overt modal operators

Consider the logical forms that the current system assigns to sentences like (2) through (4).

(2) If a cat has been exposed to 2,4-D, it must be taken to the vet immediately.

(3) If a man is in Athens he cannot be in Rhodes.
(4) If a woman's coat is missing from the coat rack, she may have gone out.

(2′)

```
                          S
              ┌───────────┼──────────────────────┐
          must₁          S̄                        S
                    ┌─────┴─────┐            ┌─────┴─────┐
                   if           S            ∃           S
                          ┌─────┴─────┐            ┌──────┴──────┐
                        NP₁           S          it₁ be taken to the
                         △          ╱──╲           vet immediately
                       a cat    e₁ has been
                              exposed to 2,4-D
```

(3′)

```
               S
          ┌────┴────┐
         not        S
               ┌─────┴─────┐
               ∃           S
                    ┌───────┼──────────────────────┐
                  can₁      S̄                        S
                       ┌────┴────┐              ┌────┴────┐
                      if         S              ∃         S
                            ┌────┴────┐              ╱────╲
                          NP₁         S          he₁ be in Rhodes
                           △        ╱──╲
                         a man   e₁ is in Athens
```

(4′)

```
                    S
        ┌───────────┼─────────────────────┐
      may₁          S̄                      S
              ┌─────┴─────┐          ┌──────┴──────┐
             if           S          ∃             S
                    ┌──────┴──────┐           ┌─────┴─────┐
                   NP₁            S          she₁ have gone out
                    △        ┌────┴────┐
                a woman   e₁'s coat is missing
                          from the coat rack
```
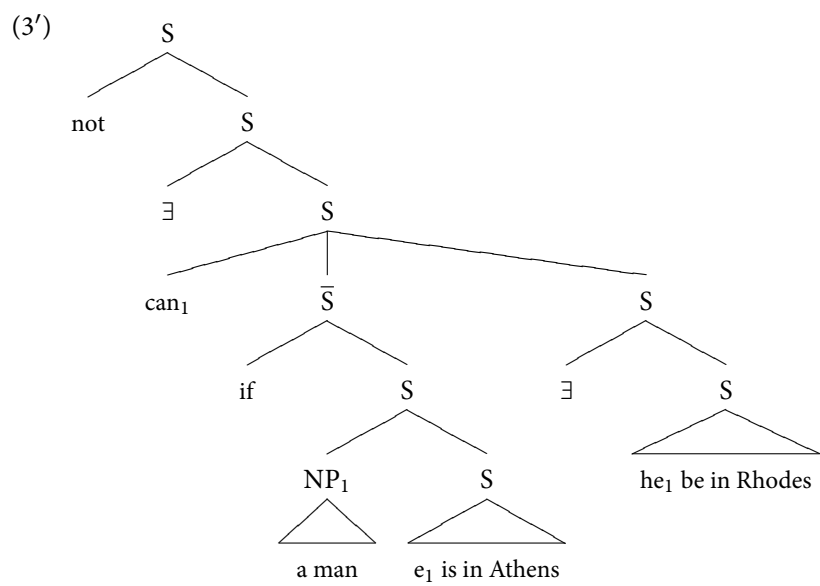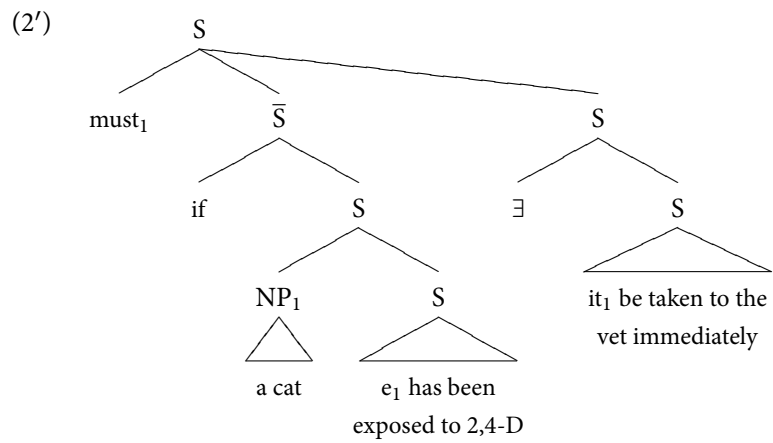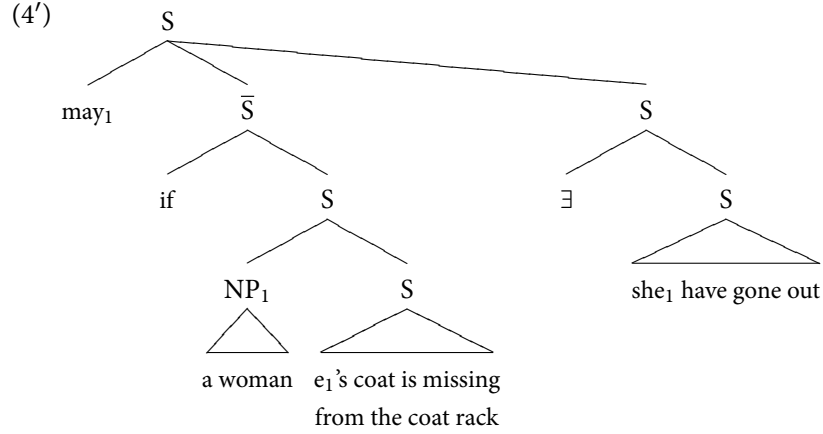
In each of these logical forms, the modal ("must," "can," or "may") has been construed as an operator with a restrictive term and a nuclear scope, the former consisting of the "if"-clause, the latter of the remaining portion of the "then"-clause. In each case, Operator Indexing has copied the index of the indefinite in the restrictive term onto that operator as a selectional index. We accordingly predict a semantic interpretation under which the modal operator somehow binds the indefinite, sort of like an adverb of quantification would.

In order to be slightly more precise about the interpretation of constructions like (2′) through (4′), I need to refer to an intensional version of the semantics given in section 3. Let models be triples $\langle W, A, Int \rangle$, instead of pairs $\langle A, Ext \rangle$, where W is the set of all possible worlds, and Int assigns to each predicate an intension, i.e., a function from possible worlds into predicate extensions. Let satisfaction be a relation between pairs $\langle w, a_N \rangle$ of a world w and a sequence of individuals a_N on the one hand and formulas on the other. The satisfaction condition for an atomic formula ϕ, consisting of an n-place predicate ζ and n variables with indices $i_1,...,i_n$ respectively, amounts now to the following:

(5) $\langle w, a_N \rangle$ $\underline{\text{SAT}}$ ϕ iff $\langle a_{i_1},...,a_{i_n} \rangle \in Int(\zeta)(w)$.

(I should really have written "$\underline{\text{SAT}}_{W,A,Int}$" rather than simply "$\underline{\text{SAT}}$.") The satisfaction conditions for molecular formulas can be taken over from the extensional semantics with routine modifications. Truth is now to be defined as satisfaction by some pair $\langle w_0, a_N \rangle$ where w_0 is the actual world and a_N includes the contextually furnished referents for each of the free variables.

Modal operators are standardly analyzed as quantifiers over possible worlds. Remaining in the spirit of the "unselective" analysis I have adopted

for quantifiers, I will assume that modal operators are, first and foremost, quantifiers over the entire tuple of parameters of evaluation, i.e., quantifiers over world-sequence-pairs $\langle w, a_N \rangle$ given the present system. Quantification over mere worlds emerges as a special case of this when a modal operator happens to occur without any selectional indices. Here is a first approximation to the rule which determines the satisfaction conditions for a formula ϕ which consists of the operator "$\text{must}_{i_1, \ldots, i_n}$" followed by a restrictive term ψ^1 and a nuclear scope ψ^2:

(6) $\langle w, a_N \rangle$ <u>SAT</u> ϕ iff for every pair $\langle w', b_N \rangle$ such that b_N agrees with a_N on all $i \notin \{i_1, \ldots, i_n\}$: if $\langle w', b_N \rangle$ <u>SAT</u> ψ^1, then $\langle w', b_N \rangle$ sat ψ^2.

Applied to example (2′), (6) predicts that (2′) comes out true just in case the following holds: for every world w, and for every individual x that is a cat in w and was exposed to 2,4-D in w, x is immediately taken to the vet in w.

This truth condition is only a very crude approximation of the actual meaning of (2). Ordinarily, we would understand (2) as involving some sort of deontic modality. It is not a claim about all cats in all possible worlds whatsoever; rather, it asserts something only about certain ideal worlds, i.e., worlds that conform to some ideal concerning the proper treatment of cats. About the remaining, non-ideal worlds, (2) does not imply anything. In particular, it does not entail that all cats which get exposed to 2,4-D in the actual world are in fact taken to the vet immediately. A more refined treatment of modal operators will have to take these observations about (2) into account. It remains to be seen whether the interaction of the modal operator with the indefinite will still fallout straightforwardly after we have replaced (6) by an appropriately revised rule.

In outlining the nature of the required refinements, I will once more rely heavily on Kratzer's work (cf. Chapter I, sec. 2.3.2), and especially on the findings reported by Kratzer (1981b). According to the theory proposed there, the observed reading of a given utterance of a modal operator is a product of three different parameters: a "modal relation," a "modal base," and an "ordering source." Modal relations are necessity, possibility, and the like; the modal relation that an operator expresses is, so to speak, its "quantificational force." A modal base defines an accessibility relation on the set of possible worlds. I write R_B for the accessibility relation defined by the modal base B; "w R_B w′" is to be read: "w is accessible from w′." An ordering source defines a relation of relative closeness to an ideal on the set of possible worlds. Given an ordering source 0, I write "w $\leq_0$ w′" (read: "w is at least as close to the ideal as w′"). It does not matter for our purposes what kinds of constructs modal bases and ordering sources *are*: all

we need to refer to is the accessibility and closeness relations they induce.[7] Among these three parameters that jointly determine the interpretation of a modalized utterance, the first one, i.e., the modal relation, tends to be determined as a lexical property of the operator involved, whereas the other two, i.e., the modal base and the ordering source, are generally supplied by the context of utterance, subject to idiosyncratic lexical preferences that particular operators may have for particular types of modal bases and ordering sources. We will get to illustrations shortly.

(You might recall that I mentioned modal relations and accessibility relations in Chapter I, sec. 2.3.2, but omitted the third parameter, the ordering source. I cannot go into the considerations here that establish the indispensability of the third parameter. See Kratzer (1981b) for the relevant arguments.)

Kratzer's analysis for an example like (7) (which is similar to our (2), except that I left out, for the moment, the additional complication of the "if"-clause and the indefinite) would be more or less as follows:

(7)   Felix must be taken to the vet.

"Must" expresses the modal relation of so-called "human necessity." Simplifying Kratzer's definition slightly[8], a proposition is a human necessity under the conditions specified in (8).

(8)   p is a human necessity in w with respect to R_B and $\leq_0$ iff p is true in every world w' which satisfies (i) and (ii):
(i) $w' R_B w$;
(ii) for every $w'' R_B w$: $w' \leq_0 w''$.

Human necessity is thus weaker in two respects than universal quantification over the entire set of possible worlds: it ranges only over accessible worlds — that is what (i) says; and even among the accessible ones, it ranges only over those that are closest to the ideal — that is the point of (ii).

---

[7] Kratzer (1981b) construes both modal bases and ordering sources as functions from possible worlds into sets of propositions. The intuitive rationale behind these concepts and their precise relation to accessibility and ordering relations are elaborated in that article and cannot be reviewed here.

[8] I make the simplifying assumption that there always exist worlds that are maximally close to w. This assumption is often unrealistic, as w might be approximated by an infinite sequence of worlds so that for each world, there is another one that is still closer to w. Cf. Lewis (1973). Kratzer's (op. cit.) definition of human necessity works in the latter case as well, unlike the simplified version I give here.

In the interpretation of (7) (under a natural reading), the modal relation of human necessity combines with certain contextually furnished choices for the modal base and the ordering source. A likely ordering source would be one which ranks worlds in terms of their closeness to an ideal of proper care for pets. "w $\leq_0$ w'" thus amounts to: in w, pets are cared for at least as properly as in w'. The modal base would plausibly be one which excludes from accessibility those worlds in which Felix is already dead, or in which certain other facts obtain that would make taking him to the vet impossible or pointless.

Given these choices, (7) is predicted to be true if Felix is taken to the vet in all worlds that are accessible in the sense alluded to and in which pets are cared for at least as properly as in any other accessible world.

What about examples with "if"-clauses, such as (9)?

(9)   If Felix was exposed to 2,4-D he must be taken to the vet.

In Kratzer's theory, (9) is just like (7), with one difference: The modal base that enters into the interpretation of (9) is not contributed by the context alone, but is a product of contextual factors and the meaning of the "if"-clause. The function of the "if"-clause is to limit accessibility to at most those worlds in which it is true. So if (9) is uttered in a context which is in all relevant respects like the context which we assumed above for the utterance of (7), then R_B will not just exclude worlds where Felix is already dead or the like, but will furthermore exclude any worlds in which Felix was not exposed to 2,4-D. In other words, a necessary condition for w' R_B w to obtain will be that Felix was exposed to 2,4-D in w'. The ordering source will be as before and "must" again expresses human necessity. Consequently, the predicted truth condition for (9) is this: (9) is true iff Felix is taken to the vet in every world w such that (i) Felix was exposed to 2,4-D in w and w also meets the contextually determined requirements for accessibility, and (ii) w is as close as any accessible world to the ideal of proper care for pets.

Up to here, I have only reported Kratzer's theory. Now I will return to examples with indefinites in the "if"-clause and to the idea that modal operators, like quantifiers and other operators, are basically unselective and range, in the general case, over world-sequence pairs, rather than just worlds. A rule of interpretation that combines this idea with Kratzer's concept of human necessity is the following:

(10)   Let ϕ be an operator-headed molecular formula, consisting of "must$_{i_1,\dots,i_n}$" and the two formulas ψ^1 and ψ^2. Then, for any w $\in$ W, $a_N \in A^N$:

$$\langle w, a_N \rangle \; \underline{\text{SAT}} \; \phi \; \text{iff} \; \langle w', b_N \rangle \; \underline{\text{SAT}} \; \psi^2$$

for every pair $\langle w', b_N \rangle$ that satisfies the following conditions:

(i) b_N agrees with a_N on all $i \notin \{i_1,...,i_n\}$;

(ii) $w' \, R_B \, w$;

(iii) $\langle w', b_N \rangle$ <u>SAT</u> ψ^1;

(iv)   for every $w'' \, R_B \, w$ such that $\langle w'', b_N \rangle$ <u>SAT</u> ψ^1, $w' \leq_0 w''$.

This rule is a little more complicated, but not essentially different from (8) above. The additional complexity has two reasons: One is that we have switched from mere worlds to world-sequence-pairs, which necessitates the added clause (i), analogous to a condition to the same effect that had to be built into (6) above. The other reason is that (10), unlike (8), incorporates the contribution that the "if"-clause (i.e., the restrictive term, ψ^1) makes to the interpretation of the whole. That is why (iii) is needed in addition to (ii). (It may ultimately be more insightful to state separate rules for the evaluation of the "if"-clause and the evaluation of the rest[9], but let me stick with (10) here.) The satisfaction relation is presumed in (10) to be relativized to the parameters R_B and $\leq_0$ in addition to the model $\langle W, A, \text{Int} \rangle$. As my earlier remarks suggest, I take it that R_B and $\leq_0$ have to be contextually furnished for an utterance to be felicitous, and that truth is defined as satisfaction w.r.t. the contextually furnished parameters.

(10) applies as follows to example (2) "If a cat is exposed to 2,4-D, it must be taken to the vet immediately," whose logical form we take to be (2'):

$\langle w, a_N \rangle$ <u>SAT</u> (2') iff: b_1 is taken to the vet immediately in w', for every b_1 and every w' which satisfy (i) through (iv):

(i)   — (irrelevant);

(ii) w' is accessible from w in the sense of the context-supplied accessibility relation;

(iii) b_1 is a cat in w' and was exposed to 2,4-D in w';

(iv) w' is as close to the ideal of proper pet care as any world that is accessible from w and in which b_1 is a cat and was exposed to 2,4-D.

(Here I have assumed that the utterance context is again essentially like in the previous examples.) This prediction seems to accord with our intuitive truth conditions for (2). It is weaker in the appropriate way than the truth condition that we initially derived on the basis of rule (6): We no

---

[9]   Kratzer (op. cit.) proceeds in that way.

longer predict (2) to claim anything about the way cats are treated in non-ideal worlds. Still, we retain the prediction that (2) involves in some sense universal quantification over cats, as a consequence of the modal operator binding the indefinite.

Let me complete this subsection by commenting briefly on examples (3) and (4). We may assume that both "can" and "may" express the modal relation of "human possibility," which Kratzer defines as the dual of human necessity. Transposed into the present framework, her definition gives rise to the following rule of interpretation:

(11)   Let ϕ be an operator-headed molecular formula, consisting of "may" or "can" with the selectional indices $i_1,...,i_n$, and the two formulas ψ^1 and ψ^2. Then, for any $w \in W$, $a_N \in A^N$, accessibility relation R_B, and ordering relation $\leq_0$:

$$\langle w, a_N \rangle \text{ \underline{SAT} } \phi \text{ iff } \langle w', b_N \rangle \text{ \underline{SAT} } \psi^2$$

for some pair $\langle w', b_N \rangle$ that satisfies the following conditions:

(i) b_N agrees with a_N on all $i \notin \{i_1,...,i_n\}$;

(ii) $w' R_B w$;

(iii) $\langle w', b_N \rangle$ <u>SAT</u> ψ^1;

(iv)   for every $w'' R_B w$ such that $\langle w'', b_N \rangle$ <u>SAT</u> ψ^1, $w' \leq_0 w''$.

A natural use of sentence (3) (="If a man is in Athens he cannot be in Rhodes.") would seem to involve a so-called "circumstantial" modal base, meaning that the accessible worlds would be just those in which the relevant geographical facts about the relative location of Athens and Rhodes obtain in the same way as in the actual world. As for the ordering source, this example does not really seem to require one at all — which amounts to the same thing as saying that it is evaluated with respect to the "trivial" ordering under which $w \leq_0 w'$ holds for any two worlds whatsoever, i.e., they are all equally ideal. Given these parameters, (11) predicts that (3′) is true if and only if it is not the case that any man in any world that is geographically like ours is both in Athens and in Rhodes. Note that what we have inside the topmost negation operator here is *existential* quantification over men as well as over worlds, which is what the assumption that the indefinite gets bound by the possibility operator leads one to expect.

The evaluation of example (4) (="If a woman's coat is missing from the coat rack, she may have gone out.") should be analogous, except that a natural reading calls for different parameters to be supplied by the context. (A relation of epistemic accessibility, coupled with an ordering that ranks worlds in terms of the extent to which they conform to the "normal course

of events," would seem to fit well.) At any rate, the prediction of (11) is again that (4) amounts to a claim about *some* women in *some* worlds.

4.2   *Indefinites bound by invisible modal operators*

We are now ready for a look at bare conditionals, i.e., those "if-then" sentences which lack an overt adverb or modal verb that the if-clause could be taken to restrict. The most famous examples of donkey sentences usually come in the form of bare conditionals:

(12)   If a man owns a donkey, he beats it.
(13)   If someone is in Athens, he is not in Rhodes.

In fact, Lewis (1975) may be the only discussion of donkey sentences that mentions any *non*-bare conditionals.

   As you recall from Chapter I, sentences like (12) and (13) have usually been judged equivalent to formulas like (12a) and (13a):

(12)   a. $\forall x \forall y ((man(x) \wedge donkey(y) \wedge own(x, y)) \rightarrow beat(x, y))$
(13)   a. $\forall x (in\text{-}Athens(x) \rightarrow \neg(in\text{-}Rhodes(x)))$

From such judgments derives the superficial impression that indefinites in the if-clauses of bare conditionals are read as universal quantifiers. Why should this be so? The question is particularly puzzling if asked in the context of the Fregean tradition according to which "if-then" is material implication, a mere truth-functional connective, and as such unsuspected of sharing any responsibility in the universal force of the sentence.

   Fortunately, an alternative theory is at our disposal which does not treat "if...then" as material implication and which, as we will see, makes the puzzle disappear. This theory of conditionals is founded upon work by Stalnaker and Lewis[10] and was further refined and generalized in subsequent research, especially Kratzer⊠s. It permits us to attribute to "if"-clauses a uniform semantic function wherever they occur, the function of restricting an operator. The operator may be lexically manifest in the form of a quantifying adverb or a modal, as witnessed in the preceding subsection; or else, it may be a syntactically unrealized necessity operator. The latter option gives rise to a bare conditional. Bare conditionals thus express conditional necessity. It happens so that there is no syntactically zero possibility operator, and therefore conditional possibility cannot be expressed by bare conditionals.

---

[10] Stalnaker (1968), Lewis (1973, 1975).

Closer examination of the unrealized necessity operator has shed light on some of its properties:[11] It expresses, like its manifest counterparts (e.g., "must"), the modal relation of human necessity and relies on the speech context to supply a modal base and an ordering source with respect to which it can be evaluated. It imposes its own idiosyncratic limits on the type of modal bases and ordering sources it accepts: apparently both must be "realistic." (A realistic modal base specifies a reflexive accessibility relation: accessible from the actual world are those worlds in which certain actual facts are true. An example would be the circumstantial modal base which was used above in the evaluation of example (3). A realistic ordering source is one which orders worlds in terms of how closely they correspond to certain actual facts.) Because of these idiosyncratic limits, the invisible necessity operator cannot always be replaced by a manifest one (say, "must") without a concomitant change in meaning.

Before I discuss the implications that such an analysis of bare conditionals has for the treatment of donkey sentences, let me stress that everything I have said in the last two paragraphs can be discovered and confirmed by studying "ordinary" bare conditionals: not donkey sentences, but sentences without problematic instances of indefinites and anaphora. The invisible necessity operator and its characteristics were posited for reasons independent of my present attempt to defend the variable analysis of indefinites. This is not the place to recapitulate those independent reasons, but let me briefly allude to them.[12] The main strength of the invisible-operator-analysis of bare conditionals, compared to the rivaling material-implication-analysis, lies in the fact that it permits a uniform account of the semantic function of "if" clauses. Those who maintain that bare conditionals express material implication are forced to concede that there are also some "if ... then" sentences which are not material implication. For instance, when the "if"-clause modifies a possibility operator or one of a variety of nonuniversal adverbs of quantification, a material implication analysis will not work.[13]

Moreover, not even all bare conditionals can be material implication, as the investigation of counterfactuals has shown: Counterfactuals typically

---

[11] The following characterization of the unrealized necessity is likely to be inaccurate in one way or another, but it seems to me at least roughly correct in view of the discussion of bare conditionals that is found in (Kratzer, 1978, part 3, Chapter 3, section 1) and Kratzer (1981b).

[12] The reasons I will allude to are presented explicitly in Kratzer (1981a).

[13] This point is made in Lewis (1975).

disallow certain types of inferences (strengthening the antecedent, for instance) that ought to be valid if material implication were involved. All this superficial diversity in the logical properties of conditionals can be explained without resorting to ambiguities in the semantics of "if," once the material implication analysis is abandoned in favor of the theory that I have been referring to, in the context of which the unrealized necessity operator has been posited.

Aside from this theoretical argument in favor of the invisible-operator-analysis ("theoretical" insofar as it relies on the greater simplicity of a theory that gets away with a uniform interpretation of "if"), there exist arguments that are more directly based on intuitions about the meanings of bare conditionals. It has often been observed, for instance, that mere falsehood of the antecedent does not strike us as a sufficient reason to assent to the truth of a conditional, and that conditionals which lack any kind of "connection" between the meanings of the antecedent and the consequent often sound strange. These observations can also be interpreted as evidence in favor of a theory that treats conditionals as inherently modalized rather than as material implication. However, they are not by themselves very compelling evidence, since they are quite successfully accounted for by appealing to Gricean conversational maxims, an account that makes them consistent with a material implication analysis.[14] So one must ultimately rely on the theoretical argument that I mentioned first if one wants to motivate the invisible operator analysis of bare conditionals compellingly.

So much for the independent motivation we will take for granted as we proceed now to analyze donkey sentences as containing an invisible necessity operator. We will see that such an analysis predicts straightforwardly the apparent universal readings of indefinites in such sentences, given the variable analysis of indefinites. This may be viewed as another argument for the invisible operator analysis of conditionals, or alternatively, as adding plausibility to the variable analysis of indefinites, more specifically, to the hypothesis that indefinites which seem to have quantificational force of their own are really bound by some other quantifying element.

Let us represent the invisible necessity operator by the symbol "□" and assume, for concreteness, that it gets inserted in the course of deriving logical form from syntactic representation. So the logical form of (12) will look like this:

---

[14] See Jackson (1979).

12′

```
                    S
          ┌─────────┼──────────────────────┐
       □₁,₂        S̄                        S
              ┌────┴────┐              ┌────┴────┐
             if         S            ∃ he₁ beat it₂
                   ┌─────┴─────┐
              a man₁ owns a donkey₂
```
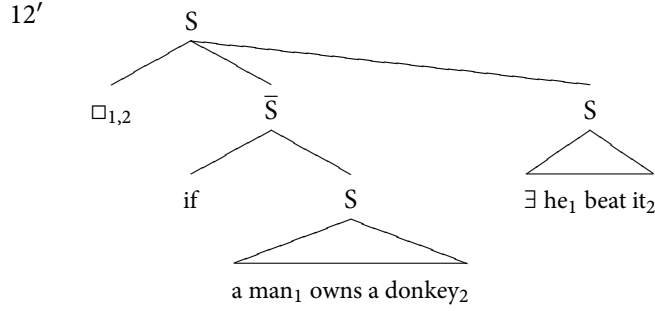
Thanks to Operator Indexing, □ is coindexed with the two indefinites in its restrictive term. (12′) will be evaluated according to rule (10), with □ functioning just like "must." This amounts to universal quantification over men and donkeys in the same way that we saw (2′) amounting to universal quantification over cats. Insofar as we predict the apparent universality of the indefinites, we are thus in agreement with the familiar judgment that (12a) is the correct formalization for (12). However, the agreement is only a rather rough one. What we actually predict is that (12′), because of its dependence upon contextually supplied parameters, corresponds to a whole range of readings, one of which, but by no means the only one, is equivalent to (12a). Let us consider some concrete choices of parameters and the resulting truth conditions for (12′) that our current assumptions would commit us to.

It has been pointed out that material implication emerges as a special case of conditional necessity when the context supplies a modal base B and an ordering source 0 with the following characteristics:[15] w' R_B w iff $w' = w$; $w' \leq_0 w$ for any two w', $w \in W$ whatsoever. B is what Kratzer calls a "totally realistic" modal base; it limits accessibility to just those worlds in which all actual facts are true. 0 is the same trivial ordering source that we appealed to in the discussion of example (3) above. It turns out that for this particular choice of B and 0, (12′) comes out equivalent to (12a). The reader can easily verify this by applying rule (10) and (12′) with R_B and $\leq_0$ as specified.

But there are different parameters that the context might supply, and this means that the predicted truth condition for (12′) will not in general reduce to (12a). It will always entail (12a), though. This follows from the assumption that the invisible necessity operator tolerates only realistic modal bases and ordering sources. (In order to prove this, you have to know that every realistic modal base determines a reflexive accessibility relation, and every realistic ordering source determines an ordering under which the actual world is as close to the ideal as any world. These are immediate consequences

---

[15] Kratzer (1981b, 68f)

of the definition of "realistic."[16] Our prediction is thus that (12) will generally express a stronger claim than the conventional formalization (12a). It will entail (12a), but not necessarily be entailed by it.

For instance, we predict that (12) might in a suitable context be used to talk about a male disposition for aggressive behavior, i.e., in the rough sense of: "in view of men␣s inherent aggressive tendencies, men who own donkeys cannot help beating them." This reading is based on a circumstantial modal base of a particular sort: accessible from the actual world are just those worlds in which men␣s psyches are as they are in fact. The ordering source is again supposed to be the trivial one which ranks all worlds equally close to the ideal. Evaluated w.r.t. these parameters, (12′) would be true iff it is true in all worlds where men␣s psyches are as in fact that all men beat all donkeys they own.

Whether or not (12) has a reading like this is a hard question to decide by intuitive judgment. But it is fair to say, I think, that the prediction we have arrived at is at least not in conflict with intuition. Perhaps the following example might be used to lend some support to the prediction that donkey sentences, just like other conditionals, can be assertions about worlds other than just the actual one. Consider (14), a variant of Stalnaker's (15) (Stalnaker, 1968, 100):

(14)   If a communist superpower enters the Vietnam conflict, the United States will attack it with nuclear weapons.
(15)   If the Chinese enter the Vietnam conflict, the United States will use nuclear weapons.

Stalnaker uses (15) to argue against the material-implication theory of if-then. He writes (loc. cit.):

> ... the following piece of reasoning is an obvious *non sequitur*: ␣I firmly believe that the Chinese will-stay out of the conflict; *therefore* I believe that the statement is true.␣ The falsity of the antecedent is never sufficient to affirm a conditional, even an indicative conditional.

By an analogous reasoning, one can argue that (14) is not naturally read as equivalent to (14a):

14   a. $\forall x((x$ is a comm. superpower $\wedge$ x enters the Vietnam conflict) $\rightarrow$ the U.S. will attack x with nuclear weapons).

(14a) is automatically true if no communist superpower enters the Vietnam conflict. Yet, one would not assent to (14) on the mere grounds of one␣s

---

[16] Kratzer (1981b, 44)

belief that all communist superpowers will remain uninvolved. Rather, the truth of (14) seems to require that the U.S. be disposed to act in a certain way. If that disposition is lacking, then the sentence counts as false, no matter what the communist superpowers will in fact do. This dispositional reading is predicted by the analysis of conditionals I am subscribing to. However, this argument can be challenged, just like Stalnaker's argument concerning (15) has been challenged: Our unwillingness to assent to conditionals on the mere grounds of falsehood of the antecedent may well have a Gricean explanation and is not in itself compelling evidence for the presence of a necessity operator. So I must emphasize, once again, that the ultimate justification for the invisible operator analysis of bare conditionals is not that it predicts more adequate truth conditions for bare conditionals than any other analysis, but that it combines with a uniform semantics for "if"-clauses and moreover provides a natural explanation for the apparent universality of indefinites in "if"-clauses.

4.3 A note on generic and other restrictive indefinites.

It appears that sometimes an indefinite all by itself does the job of an "if"-clause. Compare (16) with (17):

(16)   If a cat has been exposed to 2,4-D, it $\left\{\begin{array}{l}\text{(a) can go blind.}\\\text{(b) often goes blind.}\\\text{(c) always goes blind.}\\\text{(d) goes blind.}\end{array}\right\}$

(17)   A cat that has been exposed to 2,4-D $\left\{\begin{array}{l}\text{(a) can go blind.}\\\text{(b) often goes blind.}\\\text{(c) always goes blind.}\\\text{(d) goes blind.}\end{array}\right\}$

We have said that the "if"-clause in (16) constitutes a restrictive term for the visible or invisible operator in the matrix clause. (17) suggests that the full range of operators restrictable by "if"-clauses also accept mere indefinite NPs as their restrictive terms. The so-called "generic" use of the indefinite (cf. Chapter I, sections 1.6 and 2) is a special case of this: it is an indefinite restricting an invisible operator, as in (17d)[17]. I will use the more general term "restrictive use of the indefinite" to cover all cases where an indefinite exhaustively constitutes the restrictive term of an operator, visible or not.

---

[17] The idea that generic indefinites are semantically akin to "if" clauses is not new. See, e.g., Katz (1972), Lawler (1973).

When indefinites are construed as restrictive in this sense, we have logical forms that look like the following:

(17) c'.

$$
\begin{array}{c}
\text{S} \\
\text{always}_1 \quad \text{NP}_1 \quad\quad\quad \text{S} \\
\quad \text{NP}_1 \quad \bar{\text{S}} \quad\quad \exists \quad \text{S} \\
\text{a cat} \quad \text{that } e_1 \text{ has been} \quad\quad e_1 \text{ goes blind} \\
\text{exposed to 2,4-D}
\end{array}
$$
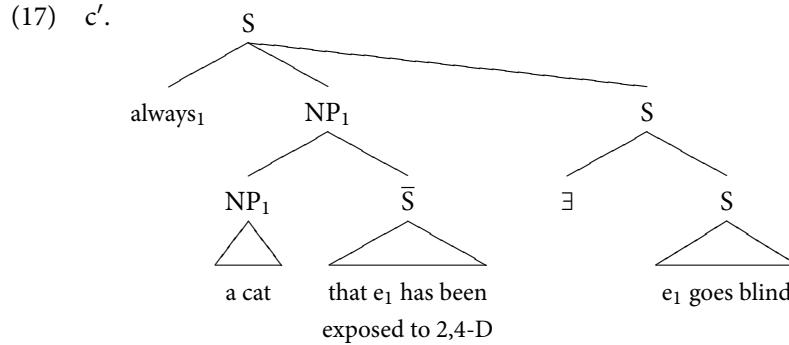
In order to generate such logical forms, we need to spell out and/or modify our construal rules in an appropriate way. I will leave this undoubtedly important task undealt with here. More precisely, I will leave it open how the indefinite comes to occupy its second position in the tripartite operator-headed structure. Assuming that some rule or other puts it there, it will follow automatically that Operator Indexing copies its index onto the operator, and the interpretation of the structure will proceed by the same rules of interpretation that are already available to us: The indefinite will get its usual propositional interpretation, and the operator will as usual apply to its two argument formulas, without regard to their syntactic categories. We thus predict the truth condition of (17c') to amount to:

$$\forall x_1((cat(x_1) \wedge \text{has-been-exposed-to-2,4-D}(x_1)) \rightarrow \text{goes-blind}(x_1)),$$

which seems adequate.

What do we predict about the meaning of "generic" indefinites, e.g., about (17d)? As a first guess, we would tend to assume that the invisible operator involved here is the very same invisible necessity operator that figures in bare conditionals. So (17d) would have a logical form identical to (17c'), except that "$\Box_1$" would appear in the place of "always$_1$." The interpretation would be determined by rule (10) of section 4.1 above, i.e., (17) would be taken to express human necessity with respect to a contextually supplied modal base and ordering source. The resulting truth condition would be just like that of the conditional (16d), involving universal quantification over accessible and maximally ideal possible worlds and over cats exposed to 2,4-D in those worlds. This result is at least roughly adequate: it captures the genericity of (17d) and the paraphrasability of (16d) in terms of (17d).

There is a substantial body of literature on generic indefinites[18], and some of the observations one finds there fit quite well with the hypothesized presence of a necessity operator; others, however, appear to suggest that this necessity operator differs in puzzling ways from the one in bare conditionals.

Nunberg and Pan (1975), for instance, come to the conclusion that a generic statement "An F is G" means not just that all or most Fs happen to be G, but rather that G is a property that holds of Fs "in virtue of [their] classmembership" in the class of Fs (p.415). For example, (17d) implies that cats go blind *because* of their being cats that were exposed to 2,4-D. This kind of causal connection between being F and being G is a paradigm case of what Nunberg and Pan mean when they say that G holds "in virtue of" F-membership. In addition to causal connections their characterization is intended to cover, e.g., the sort of "criterial" connection involved in cases where an object⊠s being G is a necessary condition for the applicability of the predicate F, as in "A true Christian is forgiving." It is not hard to relate Nunberg and Pan⊠s proposal to our hypothesis that generic indefinites restrict a necessity operator: causal and criterial connections like the ones just alluded to are special cases of conditional necessity. Quite generally, it seems fair to say that for G to apply "in virtue of classmembership" in the class of Fs is roughly the same thing as for F-hood to "necessitate" G-hood.

Now we have seen that, as a result of the operator⊠s dependence upon contextually supplied parameters, there are many different types of necessity, and so we have to ask ourselves whether generic indefinites can be read in that many different ways or whether they combine only with a limited range of modality types. In my discussion of bare conditionals, I presumed that the invisible necessity operator involved there accepted only realistic modal bases and ordering sources. It followed from this that every bare conditional entailed universal quantification in the actual world. (Cf. 4.2 above; e.g., (12a) was said to be entailed by any admissible interpretation of (12′).) Is the same limitation to realistic modal bases and ordering sources found with sentences where an indefinite restricts the invisible necessity operator? Apparently not. If it were, then "An F is G" should always entail "Every F is G." But there seems to be a generally agreed upon judgment that, for instance, the existence of a single cat that did not go blind after exposure to 2,4-D does not suffice to falsify (17d), and that "a"-generics contrast in this respect with "every"-NPs.[19] I am not sure which of various conceivable conclusions to

---

[18] See, e.g., Lawler (1973), Nunberg and Pan (1975), and Carlson (1977), as well as references cited in those works.

[19] See especially Nunberg and Pan (1975) and Carlson (1977).

draw from this judgment. Perhaps the combination of "□" with a restrictive indefinite tends to invite stereotypical, rather than realistic, ordering sources. A "stereotypical " ordering source ranks worlds in terms of their closeness to an ideal of "normality" of some sort. This opens up the possibility that the actual world w_0 is not as close to the ideal as some other accessible world w, say, because w_0 contains a few abnormally chemical-resistant cats and w does not. Then (17d) could be true in w_0 even though some abnormal cats of w_0 were exposed to 2,4-D without going blind.

There is obviously more to be said here.[20] My stopping at this point should make it obvious that I am not attempting to defend seriously a theory of generic indefinites in this dissertation. But I hope to have suggested an at least prima facie plausible treatment within the present theory, and to have given some indication of the direction in which refinements might proceed.

5 The Behavior of Indefinites with Respect to Constraints on Scope and Anaphora

Existing studies of structural constraints on the scope possibilities of quantifying NPs, such as Rodman (1976), May (1977), and Reinhart (1976), take it for granted that indefinites are quantifying NPs. Since I am rejecting this assumption, it becomes necessary to reanalyze the data involving indefinites that were supposed to be accounted for by constraints affecting all quantifying NPs, including indefinites. I hope to show that these data can still be explained, without any need to complicate the constraints that have been proposed.

This section is divided into four parts. The first two are preliminary: they clarify the concepts and recapitulate some constraints that can be motivated without taking a controversial stand on indefinites. The third part argues that no further assumptions are needed to accommodate most facts about indefinites within a variable-analysis of indefinites. The fourth part addresses some additional facts concerning scope possibilities for indefinites and has to end with the conclusion that those facts are as problematic for my theory as they have been for the quantificational analysis of indefinites.

---

[20] In particular, something should be said about the relative merits of the approach suggested here and a proposal that G. Carlson has offered in his dissertation (Carlson, 1977, Chapter V, section 2.4.1), according to which generic singular indefinites denote kinds. I must leave this for another occasion.

5.1 The notions of binding and anaphoric relatedness

Let us first fix the terminology. Both binding and anaphora are marked by coindexing in logical form, but they can be told apart as follows:[21]

(i) X *binds* Y iff:
X is an operator which has i as one of its selection indices, and Y is an NP which has i as its referential index, and X c-commands Y.

(ii) X and Y are *anaphorically* related to each other iff:
X and Y are both NPs which have i as their referential index and no operator binds one but not the other.

Notice that (i) and (ii) jointly imply that certain coindexings will mark neither binding nor anaphora; I will say more about such "vacuous" coindexings below.

Strictly speaking, this terminology will never permit us to say that an NP binds a pronoun, regardless of whether the NP is a quantifying one, as in (1), or not, as in (2).

(1) Every man arrived in a car that he owned.

(2) A man arrived in a car that he owned.

For one thing, the terminology applies only to logical forms, not to sentences. And if we consider a logical form that represents (1), we cannot even ask whether the NP "every man" binds the pronoun "he," because there is no NP "every man." Still, I will want to say that, in some sense, "every man" binds "he" in (1) (under the intended reading, that is), whereas "a man" does not bind "he" in (2) (again, under the intended reading). let us therefore introduce the following derivative terminology:

---

[21] The terminology I use here, especially my use of the term "binding," is very different from the terminology that one finds, e.g., in Chomsky's writings. Chomsky (1981) distinguishes between "argument-binding" and "non-argument-binding." What I call "binding" is presumably always a case of "non-argument-binding" in Chomsky's sense, but not vice versa. What I call "anaphoric relatedness" includes cases of "non-argument-binding," of "argument-binding," and of no "binding" at all, in Chomsky's terminology. Note that "anaphoric relatedness," unlike both Chomskyan kinds of binding, does not require either element to c-command the other.

It should be clear that my choice of terminology is in no way intended to compete with or replace Chomsky's. I have no reason to doubt that concepts like those employed by Chomsky play a role in conditions on the wellformedness of logical forms and other syntactic levels of representation.

> If T is a text, T′ is a logical form for T, T contains the NPs X and Y,
> and Z is the determiner of X, then we say that "X binds Y (under
> the reading of T represented by T′)" iff Z binds Y in T′.

It is in this derivative sense that my analysis denies that "a man" binds "he"
in (2), whereas a quantificational analysis of indefinites would assert that it
does. We can also speak of two NPs in a text as "anaphorically related (under
a certain reading)" in a derivative sense, meaning thereby that the two NPs
are anaphorically related in logical form. A consequence of this terminology
is that any two NPs that stand in a binding relation are also anaphorically
related, but not vice versa. (Notice that I take the identification of an NP in
a text with the "same" NP in a logical form for the text to be unproblematic.
The latter need neither be in the same place nor dominate the same material
as the former, since it or a part of it might have been moved by a rule of
construal.)

Before I proceed to the constraints to which binding and anaphora
are subject, let me make a remark on the status of (i) and (ii) in the
theory. (i) and (ii) are *not* to be taken as *definitions* of "binding" and
"anaphoric relatedness." If I were to define those notions, I would do it in
terms of the semantic interpretations that formulas and logical form have
been assigned, not in terms of structural configurations. I would start
by defining a "vacuous coindexing" as a case of identical indices where
an equivalent semantic interpretation would have resulted if the indices
had been different.[22] "Binding" and "anaphoric relatedness" could then be
defined as two types of non-vacuous coindexings, one involving a selectional
and a referential index, the other one two referential indices. Given such
definitions, (i) and (ii) would be derivable as theorems about binding and
anaphoric relatedness.

The point of this remark on the status of (i) and (ii) is the following:
Our intuitions about binding and anaphora relations in a given utterance are
not intuitions about something in addition to, and logically independent of,
the truth conditions of the utterance. Therefore, the particular formulations
of (i) and (ii) are in some sense forced upon us by the semantics we have
chosen for logical forms. We could not simply modify them (say, simplify

---

[22] To be concrete, I have a definition like this in mind:

Defition: Let ϕ be a formula, i a referential index, and M the set of all occurrences of
variables with index i in ϕ. Let N be a subset of M. Then the elements of N are *coindexed
vacuously* with the elements of M−N iff, for some referential index j ≠ i, ϕ has the same
satisfaction conditions as ϕ', where ϕ' results from ϕ by replacing every element of N by
a variable with index j.

(ii) by dropping the condition that "no operator bind one but not the other"), unless we make simultaneous drastic revisions in the rules of interpretation — maybe revisions to the point where a quantifier is no longer interpreted as a quantifier.

In the light of these remarks, let me point out that I have not added anything to my theory by formulating (i) and (ii): All the empirical predictions the theory makes now, it already made without (i) and (ii). This will remain so even after the various principles to be talked about in the next subsections have been added to the theory, because none of those principles employs the concepts of binding and anaphoric relatedness (they only refer to such purely structural concepts as "scope" and "coindexing"). So (i) and (ii) have no purpose but to introduce some more vocabulary in terms of which to express some of those predictions, because this is convenient and facilitates the discussion and comparison of predictions from this and alternative theories. When it comes to considerations of simplicity and naturalness of the principles that the theory advocated here employs, (i) and (ii) are to be ignored — they are not needed for the theory to work.

5.2 Constraints on scope and coindexing

Any grammar of English has to account somehow for the fact that it is not always possible to read a pair of an NP and a pronoun as anaphorically related. For instance, it is impossible to read the pairs of italicized NPs in the following texts as anaphorically related (and of course it is a fortiori impossible to read them as related by binding):

(3) $\left\{ \begin{array}{l} \textit{Every man} \\ \textit{A man} \\ \textit{John} \end{array} \right\}$ saw *him*.

(4) *He* saw a snake near $\left\{ \begin{array}{l} \textit{everyone} \\ \textit{a friend of mine} \\ \textit{John} \end{array} \right\}$

(5) When *every man* is angry *he* slams the door.

(6) *Every soldier* has a gun. *He* will shoot.

As (3) and (4) show, some of the relevant constraints apply indiscriminately to universal NPs, indefinites, and names. Others do not, as the following analogues to (5) and (6) show, in which anaphoric readings are possible.

(5) a. When $\left\{ \begin{array}{l} \textit{somebody} \\ \textit{John} \end{array} \right\}$ is angry *he* slams the door.

(6) a. $\left\{ \begin{array}{l} \textit{Someone} \\ \textit{John} \end{array} \right\}$ has a gun. *He* will shoot.

The point I would like to argue is basically that a simpler account of the observed prohibitions against anaphora can be given if we assume that indefinites do not bind their anaphors than if we assume they do. Regardless of one's analysis of indefinites, one is led to assume that anaphoric relations between NPs and pronouns can get affected by constraints of at least two different sorts: (a) constraints on identity of referential indices, of which there appear to be two, known as the Disjoint Reference Rule (DRR) and the Noncoreference Rule (NCR);[23] (b) constraints on scope, see below. DRR and NCR can be thought of either as conditions on the applicability of NP-indexing, or as filters on its output. At any rate, they should apply before any other rules of construal have been executed.[24] DRR says: "Do not coindex NP and NP′ if NP c-commands NP′, NP′ is not a reflexive or reciprocal, and neither tense nor a specified subject intervenes between NP and NP′." NCR says: "Do not coindex NP and NP′ if NP c-commands NP′ and NP′ is not a pronoun." These two constraints account for the impossibility of anaphoric relations in (3) and (4), since identity of referential indices is a necessary condition for anaphora. Neither DRR nor NCR discriminates between different kinds of non-pronominal NPs, in particular not between universal NPs and names.

The constraint whose effects we observe in (5) and (6) is of a rather different nature. It is not a constraint on the *relation* between two NPs at all, but concerns only one NP, more specifically: the possible scopes of that one NP. It is only indirectly that this scope constraint limits the anaphoric relations into which the affected NP can enter. Before I formulate the constraint, we need some conceptual clarification.

Scope is primarily a property of nodes in logical form, but it can also be attributed, in a derivative sense, to constituents of sentences under a certain reading. The following definition determines a scope for every node in logical form, including nodes that are not quantifiers or other operators. The scope of a node X in logical form consists of X together with all the nodes that

---

[23] The vast linguistic literature on these constraints includes Wasow (1972), Lasnik (1976), and Reinhart (1976), to mention only a few. The practice of distinguishing between DRR and NCR in the way that I assume here originated with Lasnik (1976).

[24] They should do so because we want to rule out coindexings like this:

He$_i$ saw [everyone$_i$'s mother]$_j$

If NP-Prefixing were to apply before NCR and DRR here, it would create a configuration that nothing rules out:

[everyone$_i$'s mother]$_j$: he$_i$ saw e$_j$

X c-commands.[25] Given this definition of scope, and given theorem (i) about binding, it is obvious that a necessary condition for a quantifier's binding an NP is that the latter be in the scope of the former. It also follows (with the terms "binding," "anaphora," and "scope" understood in their derivative sense) that an NP whose determiner is a quantifier can neither bind nor be anaphorically related to anything outside its scope. This is how constraints on scope possibilities come to indirectly constrain anaphoric possibilities as well. An example may help to illuminate this point: Consider sentence (7):

(7)   A woman who saw every man disliked him.

Suppose that, for reasons to be addressed shortly, the maximal possible scope of "every" under any grammatical reading for (7) does not exceed the relative clause; in other words, any structures that look like (7a) are ruled out as logical forms for (7), and only those of the form (7b) are sanctioned.

(7)   a.*

---

(7)   b.

```
                               S
                    ┌──────────┴──────────┐
                  NPⱼ                      S
            ┌──────┴──────┐          ┌─────┴─────┐
          NPⱼ            S̄        eⱼ disliked himₖ
          ╱╲          ┌────┴────┐
     a woman      whoⱼ          S
                      ┌──────────┼──────────┐
                  everyᵢ        NPᵢ          S
                               ╱╲      ┌─────┴─────┐
                            __  man    ∃           S
                                                   ╱╲
                                              eⱼ saw eᵢ
```
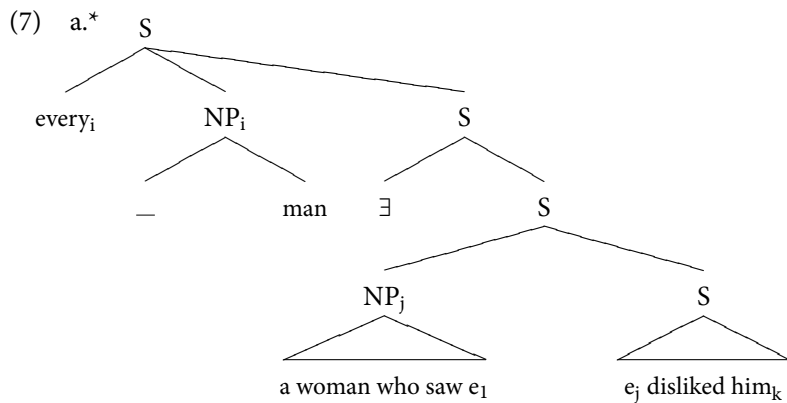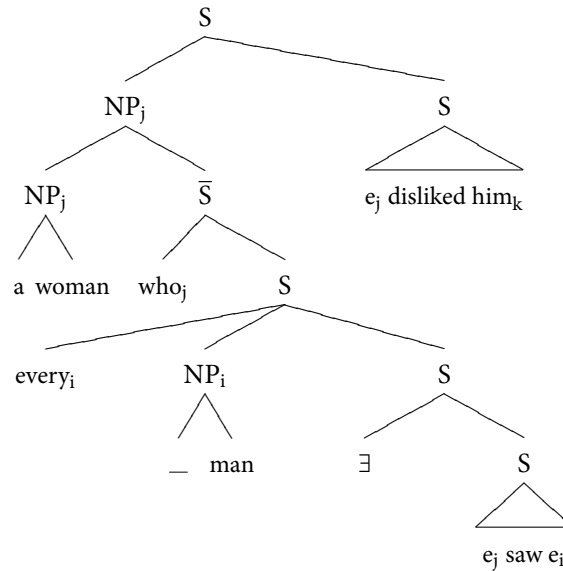
The assumption is that the relevant constraint that allows (7b), but disallows (7a), pays attention only to scope-matters, not to coindexings. In particular, it pays no attention to whether $i = k$ or $i \neq k$. (In fact, it would apply in a completely analogous fashion to a sentence that did not contain the pronoun "him" at all, say: "A woman who saw every man laughed.") Now in (7b), "him_k" is not in the scope of "$every_i$"" and therefore it cannot possibly be bound by it, not even if $i = k$. Furthermore, "him_k" cannot possibly be related anaphorically to [__ $man]_i$, not even if $i = k$; this is so because [__ $man]_i$ is bound by the quantifier "$every_i$," which does not also bind "him_k," a situation that is incompatible in principle with anaphoric relatedness (cf. (ii)). There is no constraint against coindexing "every man" with "him," but the coindexing is necessarily vacuous.

I have yet to specify the constraint(s) on scope that supposedly apply to examples like (7). The basic fact seems to be that quantifier scope is clause-bound, i.e., that the maximal scope for a quantifier is the smallest S which contains it in surface structure.[26] The only exceptions to this are quantifiers in the complement clauses of certain propositional attitude verbs, which can

---

[26] I cannot do justice to the question whether this is indeed the proper generalization to be drawn from the many facts that have been considered in the literature on structural constraints on quantifier scope. Cf. Rodman (1976), Reinhart (1976), and May (1977), to mention only a few works that are relevant here. I follow the last-mentioned most closely in the sketchy exposition that my present purposes necessitate. See also note 2.

apparently take wider scope than the matrix verb. Examples of this sort are (8) and (9).

(8)   John thinks that everyone is crazy.
(9)   Mary has permitted us to invite everybody.

I am going to disregard this qualification. Otherwise, i.e., in relative clauses, if-clauses, adverbial clauses, etc., the clause-boundedness of quantifiers is pretty much exceptionless. Examples like (5) and (7) are typical in this respect. Given our system of construal rules, the constraint can be implemented as a condition on the applicability of NP-Prefixing, which I formulate tentatively as follows: *Do not adjoin an NP any higher than to the lowest S in which it originates.* I will refer to this constraint as the "Scope Constraint" in subsequent discussion.

As it stands, this constraint applies to NPs of any kind. We have only considered examples of "every"-NPs so far, and have found them to obey the Scope Constraint. But does it generalize to names as well? As we examine the relevant examples, it turns out that they do not provide evidence either way: their behavior is compatible with the assumption that names are restricted by the Scope Constraint, as well as with the assumption that they are not. Consider (10), the analogue of (7):

(10)   A woman who saw John disliked him.

Potential logical forms could look either like (10a) or like (10b). (Note that proper names, being non-pronominal, do undergo NP-Prefixing.)

(10)   a.
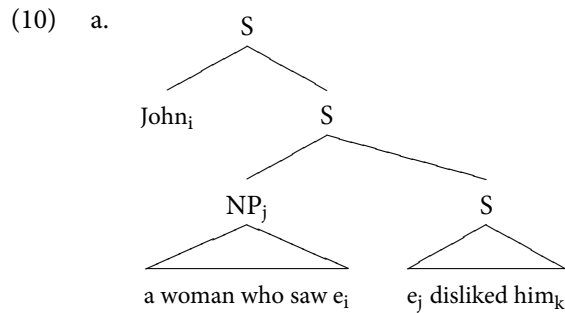
(10) b.
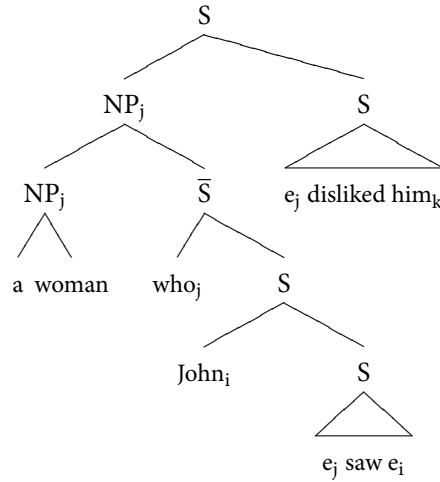
If the Scope Constraint applies to names, (10a) should be ruled out, otherwise both should be permitted. However, there is no way we can tell, since (10a) and (10b) have equivalent interpretations. In order to substantiate this, I have to make explicit my intended interpretation for names: like indefinite NPs, name-NPs are semantically open sentences. For instance, $[John]_i$ corresponds to the open sentence "$x_i = John$." More generally, an indexed NP $[\alpha]_i$, where α is a proper name whose bearer is a, is true w.r.t. a variable assignment iff the value assigned to the i-th variable is identical to a. The rest follows from what I have said earlier: The interpretation of the S-constituent that dominates a name-NP is calculated by conjoining the interpretation of the name-NP with the interpretation of its sister S. Given these assumptions, (10a) and (10b) receive equivalent interpretation. Notice that, in either structure, we may have i = k, in which case "$John_i$" and "him_k" qualify as anaphorically related by our definition. This conforms to our intuitive judgment that (10) has a reading where "John" and "him" corefer. It can be shown that the equivalence of (10a) and (10b) is not a coincidence about this particular sentence, but that the scope of a proper name never has semantic significance.[27]

---

[27] This would seem false, considering an example like (1):

(1) $Everyone_1$ who likes $John_2$ is sad.

Depending on whether the scope of "$John_2$" is construed as inside the relative clause or as encompassing the entire sentence, its logical forms correspond to the following two formulas, respectively, according to my theory:

(1) a. $x_2 = John \; \& \; \forall x_1(x_1 \text{ likes } x_2 \rightarrow x_1 \text{ is sad})$

To sum up, the Scope Constraint mayor may not pertain to proper names. For all we know, the maximally general formulation that I have given is not contradicted by any facts about proper names.

Aside from "every"-NPs, is there any more positive evidence for the Scope Constraint? Yes, it clearly affects "no"-NPs, as the following examples show:

(11) A woman that no man liked $\left\{ \begin{array}{l} \text{came in.} \\ \text{talked to him.} \end{array} \right\}$

(12) When no dog bit him John $\left\{ \begin{array}{l} \text{screamed.} \\ \text{kicked it.} \end{array} \right\}$

    a. *When no dog bit him did John $\left\{ \begin{array}{l} \text{scream.} \\ \text{kick it.} \end{array} \right\}$

Like in the case of "every"-NPs, the limiting effect of the Scope Constraint on scope-possibilities can be observed independently of its indirect role in constraining binding and anaphora relations. For later reference, I would like to point out some more facts that I take to follow from the Scope Constraint together with certain assumptions about constituent structure. Consider the following contrasts:[28]

(14)   a. No woman returned to the clubhouse when she had finished the race.
      b. When she had finished the race, no woman returned to the clubhouse.
(15)   a. Everyone should see the doctor if he is sick.
      b. If he is sick, everyone should see the doctor.

In each pair of sentences, we have a "no"- or "every"-NP in the matrix sentence and a pronoun in an adverbial clause. The former can apparently

---

(1) b. $\forall x_1((x_2 = \text{John} \& x_1 \text{ likes } x_2) \rightarrow x_1 \text{ is sad})$

(1a) and (1b) are by no means equivalent: under a variable assignment which makes "x_2 = John" false, (1a) will be false and (1b) will be true. However, it will later emerge that this consideration is ultimately irrelevant. I will propose that (1), regardless of its reading, is subject to a "felicity condition" which ensures that it will only be felicitous with respect to such variable assignments as verify "x_2 = John." It is understood that the question of truth or falsity arises only under the precondition of felicity. (1a) and (1b) do agree in truth value w.r.t. any variable assignment that meets the felicity condition, i.e., verifies "x_2 = John." This result generalizes to the claim I made in the text, insofar as that claim is understood as follows: Whenever two readings of a text differ only w.r.t. the scope of a proper name, they have identical felicity conditions and, where they are felicitous, identical truth conditions. See section 6 and Chapter III, section 5, for detail.

[28] Examples (14a)-(14b) are from Fodor and Sag (1982).

bind the latter if the adverbial clause is final, but not if it is initial. The Scope Constraint predicts this, provided that we assume that final adverbial clauses are attached underneath the S-node which immediately dominates the "no"- or "every"-NP, whereas preposed ones are attached higher up. More concretely, I will assume that final adverbial clauses are generated in position (a), and perhaps sometimes (b), and initial ones in (c).[29]
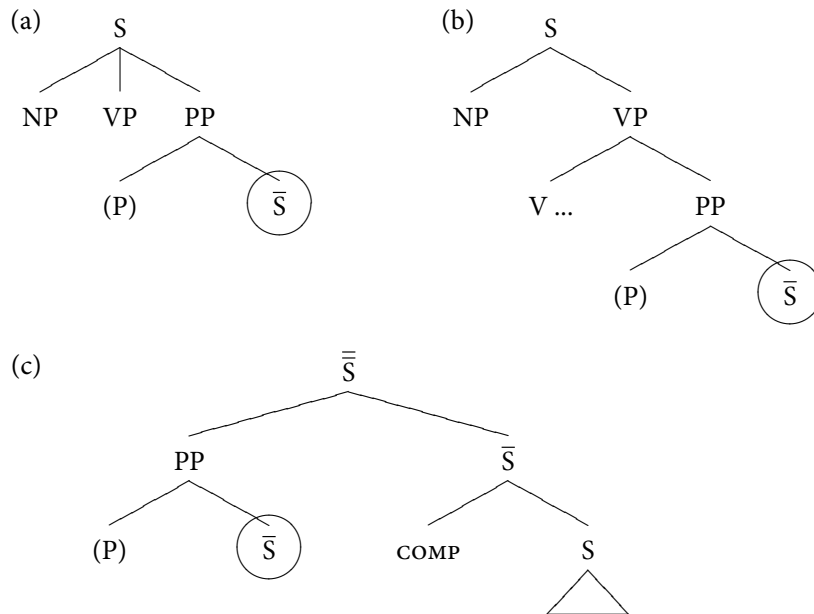
(Cf. Reinhart (1976) for evidence from NCR-facts in favor of such structures.)

It must be admitted that the Scope Constraint and the two constraints on coindexing (DRR and NCR) do not jointly suffice to account for everything that has been studied in the way of constraints on binding. In particular, they do not account for a class of cases in which a pronoun may not be bound by a quantifier, although the quantifier's scope would seem to be wide enough to include it and coindexing does not seem to violate DRR or NCR. These are the so-called "weak crossover" phenomena (using that term in its recently acquired wide sense). (16a) and (17a) are examples:

---

[29] If (c) is correct, the structures for conditionals that I have been and will be drawing are simplifications, which hopefully do not affect the points I intend to make.

(16) a. *His* mother likes *everyone*.
      b. *His* mother likes *John*.
      c. Someone likes everyone.

(17) a. The fact that *he* lost bothered *everyone*.
      b. The fact that *he* lost bothered *John*.
      c. Something bothered everyone.

The pronouns "his" and "he" in (16a) and (17a) cannot be read as bound by "everyone." Yet, the scope of "everyone" is able in principle to extend over the site of those pronouns, as the existence of $\forall$-$\exists$-readings for (16c) and (17c) shows. Neither NCR nor DRR can be responsible for the impermissibility of binding in the (a)-cases, because then the (b)cases should not permit cataphoric readings either.

Weak crossover cases are usually dealt with by positing another constraint on coindexing, which, unlike DRR and NCR, applies specifically to quantifying NPs, saying something like: "NP and NP′, where NP is quantifying, cannot be coindexed unless NP c-commands NP′ in surface structure," or perhaps: "NP and NP′, where NP is quantifying, cannot be coindexed unless NP is to the left of NP′ in surface structure." Leaving it open whether c-command or leftness is the relevant factor, I will refer to this as the "Weak Crossover Constraint."

The existence of the Weak Crossover Constraint as a separate stipulation over and beyond DRR, NCR, and the Scope Constraint is somewhat dissatisfying from a theoretical point of view, since it leads to a great deal of redundancy: Many coindexings that are vacuous because of the Scope Constraint anyway get ruled out additionally by the Weak Crossover Constraint. But I do not have a more satisfying solution to offer.

In the context of the present dissertation, we will have to attend to the question whether quantifyingness is indeed the distinguishing factor that determines whether an NP is or is not subject to the Weak Crossover Constraint, and in particular to the question how this affects the choice between a quantificational and a variable analysis of indefinites.

5.3 *Coindexing and scope constraints in view of the variable analysis of indefinites.*

Indefinites show a behavior that is different both from the behavior of "every"-NPs and "no"-NPs on the one hand, and from that of names on the other. For instance, (18), where an anaphoric reading is possible, contrasts with (19), where it is not:

(18)   If $\left\{ \begin{array}{l} \textit{somebody} \\ \textit{John} \end{array} \right\}$ commits a crime, *he* must go to jail.

(19)   If $\left\{ \begin{array}{l} \textit{everyone} \\ \textit{no one} \end{array} \right\}$ commits a crime, *he* must go to jail.

Here we see the indefinite pattern with the name. But in other cases, it patterns with the "every"-NPs and "no"-NPs: (20) permits anaphora, whereas (21) does not:

(20)   Whenever *he* didn't bother her, Mary ignored *John*.

(21)   Whenever *he* didn't bother her, Mary ignored $\left\{ \begin{array}{l} \textit{someone.} \\ \textit{everyone.} \\ \textit{no one.} \end{array} \right\}$

These facts are just what we predict if we combine our analysis of in definites with what has been presented in the preceding subsection. Here is why:

Needless to say, we expect indefinites to obey Disjoint Reference, and moreover, since they are non-pronominal, Noncoreference. This is generally accepted, and I need not argue for it. (See (3) and (4) in section 5.2 above for confirming examples.) More interestingly, we expect indefinites to obey the Scope Constraint, since we have formulated the latter in a maximally general fashion. let us see how it affects the indefinites in examples (18) and (21). For (18), we generate only logical forms like (18a); (18b) being ruled out by the Scope Constraint:

(18)   a.
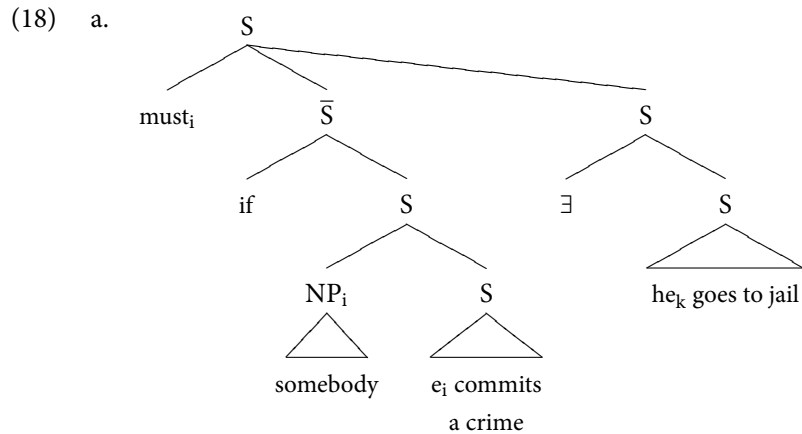
(18)   b.*
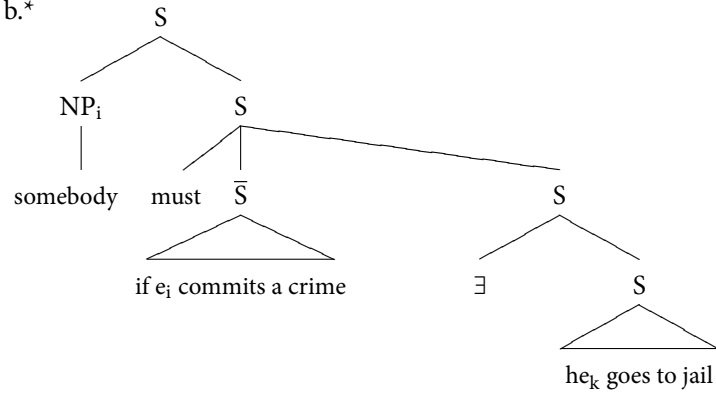
The crucial point is that in (18a), although "he_k" is outside the scope of "$somebody_i$," an anaphoric relation may obtain between them. All it takes is that i = k; given i = k, the conditions for anaphoric relatedness will be met, since the only operator that binds one of them (i .e., "must.") will also bind the other.

Note that essentially the same state of affairs would have arisen, had the sentence contained "John" instead of "somebody." The logical form would have been isomorphic to (18a), and the definition for anaphoric relatedness satisfied under the same conditions. Had the sentence contained "everyone," however, things would have been different: "every" would have been moved out, leaving behind "_-one." The latter could have been coindexed with "he." but only vacuously, since "every," whose scope is confined inside the if-clause, would have bound "one," but not "he." A general lesson can be drawn from this comparison: Indefinites can be anaphorically related with things outside of their scope; in this respect, they are like names and unlike quantifying NPs. This explains the contrast between (18) and (19), and it explains the following contrasts in an analogous way. In each pair, the (a)-variants permit anaphora between the italicized NPs, whereas the (b)-variants do not.

(22) $\left\{ \begin{array}{l} \text{(a) } \textit{John/one of the soldiers} \\ \text{(b) } \textit{every soldier/no soldier} \end{array} \right\}$ has a gun. *He* will shoot.

(23)   A woman who knew $\left\{ \begin{array}{l} \text{(a) } \textit{John/a magician} \\ \text{(b) } \textit{every magician/no magician} \end{array} \right\}$ brought *him* to the party.

Now what about (20) versus (21)? Here it looks like indefinites share some property with quantifying NPs that names lack. Because of the Scope Constraint, the "someone"-variant of (21) can only have logical forms like (21a), and none like (21b).

(21) a.

```
                        S̿
            ┌───────────┼──────────────────┐
        -ever PP                           S̄
              │                     ┌───────┴───────┐
              S̄                    ∃_k              S̄
          ┌───┴────┐                                │
        when-      S                                S
              ┌────┴────┐                    ┌───────┴────────┐
       he_i didn't bother her             NP_k               S
                                            │          ┌──────┴──────┐
                                         someone    Mary ignored e_k
```

(21) b.*

```
                    S
              ┌─────┴──────┐
        someone_k          S̿
                     ┌──────┴──────┐
               whenever he_i didn't bother
                  her, Mary ignored e_k
```
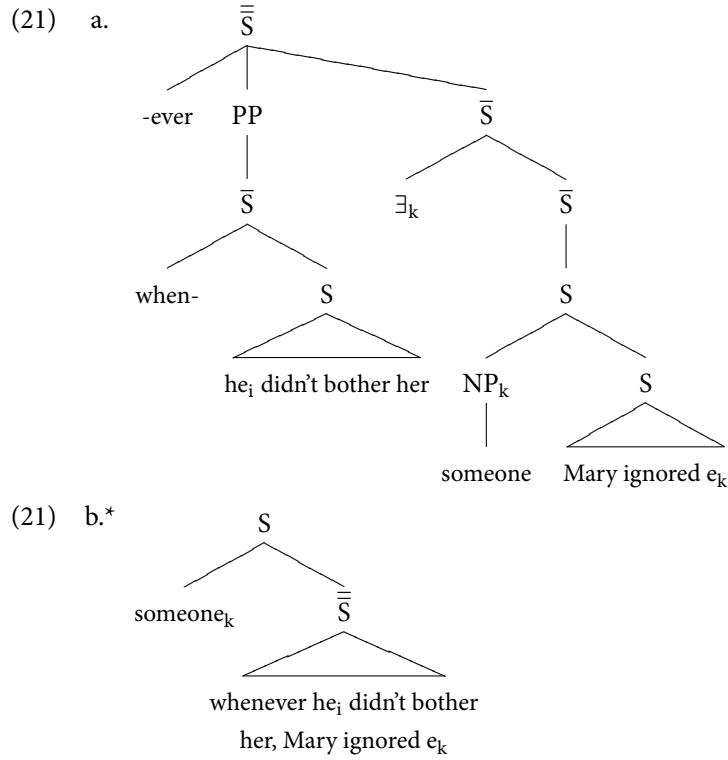
"Whenever" has here been analyzed as an amalgam of the universal Q-adverb "always" (realized as "-ever") and a "when"-clause that restricts it. The logical form is tripartite, as required for quantifiers, and the nuclear scope of "-ever" is subject to Existential Closure. (Actually, "-ever" quantifies over time in this sentence.) Given that Existentlal Closure has applied, we are forced by Quantifier Indexing to bind "someone_k" within the nuclear scope of "-ever," and this means that "someone_k" and "he_i" cannot possibly be anaphorically related, not even if i = k.

For comparison, suppose now we had had "John_k" in the place of "someone_k." Then the logical form would have looked just like (21a) with one crucial difference: "John_k" would not have been coindexed with the existential quantifier next to it, because names are definite and therefore not eligible for Quantifier Indexing. With a quantifying NP, e.g., "everyone," in the place of "someone," the logical form would have looked different from either the case with an indefinite or the case with a name: instead of the bipartite S [NP_k: Mary ignored e_k], there would have been a tripartite one, [every_k: NP_k: Mary ignored e_k]. Anaphora between NP_k and "he_i" would

have been impossible, like in the "someone"-variant, but not because of the intervening ∃, but because of the intervening "every."

The example we have just analyzed illustrates an important difference between proper names and indefinites: Whereas the scope of a proper name has no semantic significance, the scope of an indefinite does. This puts indefinites on a par with quantifying NPs, but it does not mean that they are quantifying.

There seems to be something problematic about the explanation that I just offered for the contrast (20) versus (21): It relies crucially on the presence of a quantifier (here: "-ever") relative to which the adverbial and the matrix clause function as restrictive term and nuclear scope, respectively. Had there been no quantifier, Existential Closure would not have applied, and then, what would have prevented anaphora between the indefinite and the earlier pronoun? let us look at an example where we actually have a preposed adverbial-clause that does not restrict a quantifier. Adverbial clauses with temporal "as" can be of this sort.

(24)   As Mary mentioned *his* name, we saw a *magician*.

"As" presumably means "at the time at which... ," and therefore the logical form should look like (24a). This is interpreted as a conjunction of the adverbial clause and the matrix clause.[30]

(24)   a.
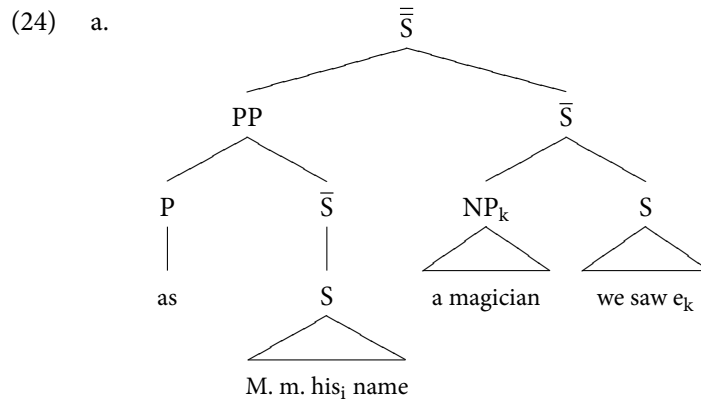
If i=k, the definition for anaphoric relatedness is met in (24a). Yet, the facts are the same in (24) as in (21): anaphora is not possible.

The problem disappears if we appeal to the Novelty Condition: Since the indefinite is to the right of the pronoun, coindexing is not permissible. If this

---

[30] Presumably, the content of the adverbial clause has the status of a presupposition. I ignore this here.

is the correct explanation for (24), then we actually have a double explanation for (21): Aside from the fact that, as I pointed out, coindexing between "he" and "someone" in (21a) is doomed to vacuity, it moreover violates the Novelty Condition.

Examples like (21) and (24) are not the only ones in which indefinites show the same behavior as quantifying NPs. There is a whole variety of examples in which this is the case, and that is why studies of anaphora and scope constraints have been so relatively successful in predicting the behavior of indefinites while analyzing them as quantifiers. Consider, for instance, the following pair:

(25)   Mary took every cat that $\left\{ \begin{array}{l} \text{nobody} \\ \text{somebody} \end{array} \right\}$ liked.

Under the assumption that indefinites are existential quantifiers, (25) illustrates that their scope, just like the scope of other quantifiers, is limited to the relative clause (a special case of the Scope Constraint). Just as the "nobody"-variant of (25) can never mean that nobody was such that Mary took every cat he liked, the "somebody"-variant lacks a reading according to which somebody was such that Mary took every cat he liked. This very judgment turns out to be predicted in my theory, albeit with recourse to different assumptions: By the Scope Constraint, the NP "somebody" is adjoined to the relative S. Therefore, Quantifier Indexing copies its index onto "every," which is the lowest c-commanding quantifier. The resulting interpretation is that for every pair of a cat and somebody that liked it, Mary took the cat in that pair. This happens to be equivalent to: For every cat such that there was somebody that liked it, Mary took it. The excluded reading, where "somebody" would have been bound by the existential quantifier which closes the text in which (25) occurs, is ruled out because of the Scope Constraint.

Let us draw a preliminary summary. To the extent that the indefinites in the examples above show the behavior that one would expect of an existential quantifier obeying constraints on quantifier scope, this behavior is equally well explained in the theory I am advocating, which holds that indefinites do not quantify, and that scope constraints are not limited to *quantifying* NPs. This claim was substantiated by reanalyzing examples (21), (24), and (25), and if it is accepted, the variable-analysis of indefinites would appear to be at least as capable as the quantifier-analysis to combine with a simple and observationally adequate set of constraints on anaphora and scope. Beyond that, however, I have explained facts which are not covered under a quantificational analysis of indefinites, viz., facts that attest to less limited anaphoric relationships for indefinites than there are for undisputedly quantified NPs

(cf. (18), (22), and (23)). This puts the theory advocated here not only on a par with, but at an advantage over, its competitor — unless, of course, the facts in question can be shown to admit of a different explanation (e.g., along the lines of one of the approaches in Chapter I).

But we have yet to see whether we come to such satisfying findings when we examine the behavior of indefinites with respect to the Weak Crossover Constraint. Here we witness the indefinites patterning with the quantifiers, not with the names: The "somebody"-variants of (26) and (27) do not allow cataphoric readings any more than the "everybody"-variants.

(26) *His* mother likes $\left\{ \begin{array}{l} \textit{somebody.} \\ \textit{everybody.} \\ \textit{John.} \end{array} \right\}$ $= (16a)$ $= (16b)$

(27) The fact that *he* lost bothered $\left\{ \begin{array}{l} \textit{somebody.} \\ \textit{everybody.} \\ \textit{John.} \end{array} \right\}$ $= (17a)$ $= (17b)$

This is not to be expected from the point of view of a variable analysis of indefinites, if the Weak Crossover Constraint is indeed limited to *quantifying* NPs. I took it to be so in my exposition above, following conventional wisdom. However, alternative formulations can be found. Perhaps Weak Crossover applies specifically to [-*definite*] NPs, where that class is understood as including NPs with the indefinite article as well as quantifying NPs. It may be arbitrary and ad hoc to count quantifying NPs as indefinite rather than definite; but recall that it is at least consistent with a hypothesis entertained above (in section 2), viz., that quantifying NPs leave behind a [-definite] remnant when their quantifiers are removed by Quantifier Construal, and that this is why they impose a selectional index on that quantifier, as predicted by Quantifier Indexing. At any rate, it is natural enough to assume that whatever is not [+definite] is [-definite], and if so, then to make quantifying NPs [-definite] is not any more costly a stipulation than not to.

If we take the Weak Crossover Constraint to refer to indefiniteness, rather than quantifyingness, then the facts of (26) and (27) are accounted for, although of course not in a particularly insightful way. So it is fair to say that the weak crossover phenomena with indefinites do not force us to return to a quantificational analysis of indefinites; they are equally well described under a variable analysis.

5.4   *The exceptional behavior of "specific" indefinites.*

Up to now, I have limited my attention to those data which show indefinites obeying the Scope Constraint. These are also the data that used to be taken as support for the quantificational analysis of indefinites. All the while it was known that some indefinites behave differently. Compare (25) with (28):

(25)   Mary took every cat that somebody liked.

(28)   Mary took every cat that a friend of hers, whose birthday it was, liked.

While the scope of "somebody" seems confined within the relative clause, as predicted (see above), the more elaborate indefinite that replaces it in (28) not only can, but must be, read with wider scope: (28) does not leave open the possibility that, for every cat that Mary took, a different friend of hers liked it.

What explains the difference between (25) and (28)? The indefinite in (28) has what is often called a *specific* reading. Before trying to say what a specific reading is — a question to which conflicting answers have been given — let me mention some facts that are generally agreed to hold of the phenomenon: Any indefinite NP can in principle be read specifically, but for some this is much less preferred than for others. Factors that increase the availability of a specific reading are (according to Fodor and Sag (1982)): descriptive richness of the predicate in the NP; use of the modifiers "certain" and "particular" after the indefinite article; left-dislocated, topicalized, and (to a lesser degree) subject position; and co-occurrence with a non-restrictive relative. The last-mentioned factor not only favors, but forces, specificity. A specific reading is also forced if indefinite "this" is used instead of the indefinite article.[31] Our example (28), as opposed to (25), involves several of the specificity-favoring factors mentioned.

But what *is* specificity? The one thing that is agreed on is that the specific reading always entails the widest-scope existential reading. More precisely: If a sentence [X-[a $\overline{N}$]-Y] is interpreted under the specific reading of [a $\overline{N}$], then it entails $\exists x(x$ is an $\overline{N} \wedge [X\text{-}x\text{-}Y])$. This provides a criterion for distinguishing the specific reading from any narrow-scope readings that there may be, but it does not say whether and how the specific reading differs from the widest-scope existential reading. And in fact, some people think that that is just what the specific reading is: a widest-scope existential reading. (Under my assumptions, this would of course mean that the indefinite is *bound* by a widest scope existential quantifier, not that it *is* one.) But others disagree: For them, specific indefinites are referring expressions,

---

[31] See Prince (1981) for facts about indefinite "this."

sort of like deictic pronouns, except that the hearer is not supposed to be able to identify the referent. This is the position that Fodor and Sag, among others, take.

Let us return to our question: What explains the fact that (28) has a reading which — to put it neutrally — entails the widest-scope existential reading, whereas (25) does not permit — or at least very highly disfavors — such a reading? Or, more generally, what explains the fact that specific indefinites appear to violate scope constraints? The simplest answer that comes to mind in the context of the approach here developed would seem to be this: Indefinites, or maybe non-quantifying NPs in general, are permitted to violate the Scope Constraint, though only as a non-preferred option. They will normally obey it, but overriding factors can force extra-wide scope in certain cases. Such factors could be of various sorts: Descriptive richness of the indefinite can make one reading less likely than another. Words like "certain" might simply be marked, as part of their idiosyncratic lexical properties, as favoring wide scope. Certain syntactic positions, such as topicalized and subject position, are notorious for favoring relatively wide scopes anyway, even with quantifying NPs, and might for the same reason — whatever it may be — ease violations of the Scope Constraint for indefinites.

That the presence of an appositive relative not only favors, but forces, a widest-scope reading would presumably fall out from an analysis of appositive relatives, which one can expect to predict (a) that an appositive relative is interpreted as conjoined to the unembedded sentence it occurs in; and (b) that the relative pronoun must be anaphoric to the NP to which the appositive relative is attached. It follows from (a) and (b) that example (28), for instance, is interpretable only when the indefinite is given matrix-clause scope.

Unfortunately, such a simple answer will not do. The point of Fodor and Sag (1982) is precisely that a mere relaxation of scope constraints for indefinites will not suffice to predict correctly the behavior of specific indefinites. Although the proposal they argue against is different from the one I just sketched in that it presumes a quantificational, rather than a variable, analysis of indefinites, it turns out that at least one of their central arguments carries over.[32] Let me recapitulate that argument briefly.

---

[32] Their second central argument, which I do not think is applicable under the assumptions of a variable analysis of indefinites, has to do with VP-Deletion. If I were to reproduce that argument and to show how one might respond to it in the context of the present treatment of indefinites, I would first have to talk a fair amount about VP-Deletion in general. I must postpone this to another occasion.

Fodor and Sag observe that the following sentences have only two, not three, readings:

(29)   Each teacher overheard a rumor that a student of mine had been called before the dean.

(30)   Each teacher thinks that for a student of mine to be called before the dean would be preposterous.

Under the assumption that (29) contains three expressions with semantically significant scope — "each teacher," "rumor," and "a student of mine" — and that "a student of mine" can violate scope constraints, one should expect three different scope constellations to be possible:

(i)   each teacher – rumor – a student of mine;

(ii)   each teacher – a student of mine – rumor;

(iii) a student of mine – each teacher – rumor.

Reading (ii), however, is missing. Analogously, (30) lacks a reading where "a student of mine" takes narrower scope than "each teacher," but wider scope than "be preposterous." The generalization appears to be that indefinites which violate scope constraints always must take the very widest possible scope. This is not predicted by a theory that simply allows for violations of scope constraints by indefinites, regardless of whether that theory treats indefinites as quantifiers or as variables in the manner proposed here.

Fodor and Sag go on to suggest that the missing intermediate scope phenomenon is predicted if, as they advocate, specific indefinites are referring expressions and as such do not participate in scope relations at all. In the context of the present theory, this proposal might be approximated by saying that specific indefinites are actually [+definite] from the point of view of construal and interpretation rules. It would then follow, as I pointed out with regard to proper names in section 5.2, that specific indefinites are unaffected by Quantifier Indexing, that they are associated with certain felicity conditions (see section 6.2 below), and that as a result, the scope of a specific indefinite is never semantically significant. There would be no need to appeal to violations of scope constraints to predict the seeming wide-scope readings for (28), (29), and (30); it would suffice to say that the indefinites in those sentences may be marked [+definite]. Note that this assumption would also yield the welcome prediction that specific indefinites need not obey the Weak Crossover Constraint, a fact that one could not explain by merely relaxing the scope constraints. Consider Wasow's (1972) example (31), which permits backwards anaphora:

(31)   The woman *he* loved betrayed *a man I knew*.

However, one clearly feels that a point is being missed in analyzing specific indefinites as [+definite]. After all, they are *in*definite in *some* respect. At least their felicity conditions differ systematically from those of genuine definites. We would thus have to conclude that definiteness is not, after all, a single binary feature. Rather, there would seem to be definiteness with respect to the construal component on the one hand, and definiteness with respect to felicity conditions on the other.

Perhaps we had better consider specific indefinites as [-definite], as originally assumed, and introduce a further construal rule that raises an indefinite (perhaps: any non-quantifying NP) into a position dominated immediately by the T (text) node. We could stipulate that this rule precedes the Weak Crossover Constraint and is not subject to the Scope Constraint. This is ad hoc, but would work.

Leaving the matter unresolved, let me conclude that the exceptional behavior of specific indefinites remains as puzzling under the assumptions of this dissertation as it is from the point of view of the standard quantificational analysis of indefinites.

To make matters worse, we find that some of the exceptional behavior of specific indefinites is shared by generic indefinites. As Postal (1970) and Wasow (1972) have noted, we get backwards anaphora, i.e., Weak Crossover violations, and even concomitant Scope Constraint violations, in generic sentences like (32).

(32)   If *he* has an ugly wife, *a man* should find a mistress.

We could definitely not account for this by giving indefinites a marked option to be [+definite], since indefiniteness must be crucially assumed in order to predict the generic reading (see section 4.3 above). (32) might call for another ad hoc construal rule, again superceding the Scope and Weak Crossover Constraints, which positions an indefinite next to the generic operator it restricts.

6   Are Non-Pronominal Definites Variables Too?

I have not said much about definites other than personal pronouns so far, and indeed, there are some problems lurking there. We expect that a definite description like "the cat" should differ minimally from its indefinite analogue "a cat," i.e., that it should behave alike up to those properties that depend on the feature of definiteness. This expectation arises not only from an aesthetic preference for symmetry, but also from arguments that present the

standard quantificational analysis of definite descriptions with challenges parallel to those which led us to abandon the quantificational analysis of indefinites. However, the assumption that definite descriptions are like their indefinite counterparts up to the feature [+definite] does not obviously lead to the right predictions. Some improvement will be made by adding suitable stipulations concerning the felicity conditions of definite descriptions. But certain problems will remain open and become manageable only in Chapter III.

6.1 Definite descriptions are not quantifying.

After all that I have said, the reader will hardly be surprised that I reject Russell's (1905) analysis of definite descriptions as quantifying NPs along with his quantificational analysis of indefinites. I have several reasons for this, most of which have already been alluded to and need only brief repetition here:

(a) It is a consequence of Russell's analysis that a definite description which does not contain any bound variables has a unique referent (if it occurs in a true sentence and not under the scope of negation, that is). This prediction is not borne out in certain examples of anaphorically used definite descriptions, especially examples like (1).

   (1)   If a man beats a donkey, the donkey kicks him.

   "The donkey" in (1) does not refer to any particular donkey any more than a pronoun "it" in its place would. The only way of reconciling such examples with a Russellian analysis of definite descriptions seems to be by arguing that the problematic definite descriptions do, despite appearances, contain bound variables. (Cf. Cooper's (1979) theory of pronouns and my discussion of it in Chapter I, sections 1.4 and 2.3.) In the context of the current theory, we might save Russell's analysis by saying that "the donkey" in (1) is to be disambiguated as "the donkey which is identical to it$_i$," where i is the index of "a donkey." (In this manner, you can of course analyze any bound-variable pronoun as a definite description as well.)

(b) As far as constraints on anaphora and scope go (see section 5), definite descriptions appear to be able to be anaphorically related with pronouns outside of their scope, as indicated by the non-vacuous coindexing of the two italicized NPs in (2).

   (2)   [Every boy who likes [*his*$_1$ *mother*$_2$]$_1$] visits *her*$_2$ for Christmas.

The scope of "[his$_1$ mother]$_2$" in (2) cannot possibly be wider than the restrictive term of "every," since "his$_1$" is to be bound by "every." So "her$_2$" must be outside the scope of "[his$_1$ mother]$_2$." If definite descriptions are not quantifying, this will not prevent the two from being anaphorically related, but if they are, then the coindexing must be vacuous (see section 5 for why this is so). Yet, there is clearly a reading under which "[his$_1$ mother]$_2$" and "her$_2$" are perceived as anaphorically related, and so we are compelled to conclude that definite descriptions must not be quantifying.

Note that one might escape this conclusion by arguing as follows: Not everything that we intuitively perceive as an anaphoric relation is necessarily a matter of non-vacuous coindexing in logical form. There may well be semantic relations between NPs that are superficially indistinguishable from anaphoric relatedness in the technical sense, yet may hold between NPs with different, or only vacuously identical, indices. For instance, Kripke (cf. Chapter I, section 1.3 above) would maintain that for two NPs to be used as referring to the same speaker's referent does not require that there be any semantic relation (such as containing the same variable) between them. But in the example at hand, sameness of speaker's reference cannot be appealed to, as neither NP refers at all, and I do not know what other pseudo-anaphoric relation could plausibly be appealed to. I admit that we are not strictly compelled by our intuitions about (2) to diagnose an anaphoric relation between "his mother" and "her" in the technical sense of "anaphoric relatedness." But the burden of proposing an alternative diagnosis is on those who, because they consider definite descriptions as quantifying, must deny that the two are anaphorically related in this technical sense.

(c) Given the treatment for indefinites that I have developed, treating definite descriptions as quantifying NPs would make definite and indefinite descriptions very different from each other. This seems wrong in view of the fact that many natural languages do not have definite and indefinite articles and use the same article-less form where English has either "a …" or "the …" It could of course be the case that in such languages, all NPs not marked for definiteness are ambiguous between a [-definite] non-quantifier on the one hand and a quantifier with a certain force on the other. But a more plausible hypothesis is that indefinites and their definite counterparts are alike in every respect except for the feature [+-definite] and their behavior w.r.t. rules and principles that make reference to that feature, and that in a language with NPs unmarked for definiteness, those NPs are

ambiguous only with respect to the feature [+-definite] and can freely be treated as either [+definite] or [-definite] by the relevant rules and principles.

6.2 *The felicity conditions of definites with descriptive content*

In view of the above arguments, I assume that definite descriptions have no quantificational force, but contain an essential free variable, just like their indefinite counterparts. Under this assumption, a minimal pair of simple texts like (3) and (4) receive almost identical logical forms.

(3)   A cat is at the door.
(4)   The cat is at the door.
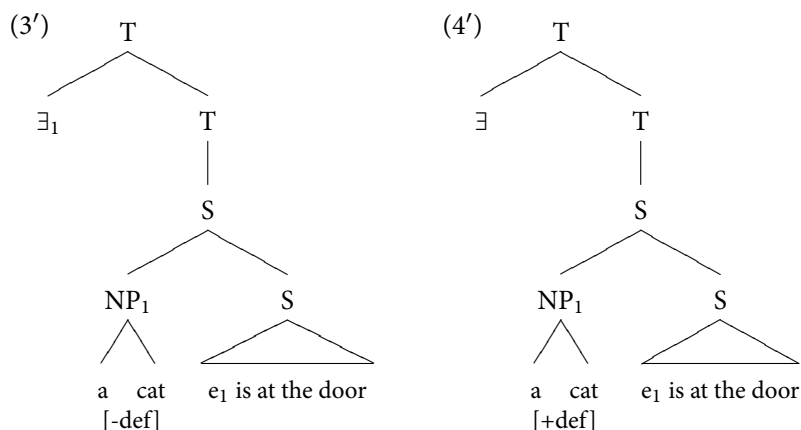
The only difference results from Quantifier Indexing: "a cat" gets existentially bound, because it is indefinite, while "the cat" remains free. The interpretations associated with (3′) and (4′) correspond to a closed and an open formula, respectively:

(3)   a. $\exists x_1(cat(x_1) \wedge at\text{-}the\text{-}door(x_1))$
(4)   a. $cat(x_1) \wedge at\text{-}the\text{-}door(x_1)$

(3a) is a satisfactory representation of the meaning of (3), but (4a) does not so obviously represent the meaning of (4), as we will see. Recall from section 3.3 above that a text with free variables in it requires for its felicitous utterance a context which determines an assignment of values to its free variables, and its truth value on a particular occasion of utterance depends on that

contextually supplied assignment. So if (4), under the reading represented by (4′), is uttered in a context which determines Felix as the value of the 1st variable, then it is true if Felix is a cat and at the door, and false otherwise.

But there is something wrong with this analysis. What we seem to be predicting is that by uttering (4), the speaker asserts about a contextually determined individual that it is a cat and it is at the door. But intuitively, the description "cat" is not part of what is asserted about the referent, but rather serves to narrow down the range of things that can felicitously be referred to. So it is not possible to read (4) as a claim — a false claim, as it were — that a certain non-cat is a cat and at the door. The two predicates, "cat" and "at the door," play different roles in our understanding of (4), a fact that is obscured by the symmetry of a formula like (4a). As our analysis stands, it actually predicts equivalent interpretations for (4), (5), and (6), clearly contrary to intuitive judgment.

(5)   [The entity at the door]$_1$ is a cat.
(6)   It$_1$ is a cat and it$_1$ is at the door.

Nevertheless, I want to maintain that (4′), hence (4a), is correct as a representation of one aspect of the meaning of (4), i.e., the purely truth-conditional aspect.

It is the aspect of felicity conditions that we have not given adequate attention so far. With an utterance of (4′), it is simply not enough that the context supply some value or other for the 1st variable. (4′) carries moreover a so-called presupposition, viz., the presupposition that the value of the 1st variable is a cat. Formula (4b) expresses this presupposition:

(4)   b. cat(x_1)

Presuppositions amount to felicity conditions in the following way: If an utterance is to be felicitous in a situation, its presupposition has to be presupposed to be true by the discourse-participants in this situation. Because of this, presuppositions frequently aid in the disambiguation of ambiguous and context-dependent sentences: of all the potential readings the sentence might have, only those come into consideration whose presuppositions are in fact presupposed in the context at hand. In the case of (4), it works like this: In principle, (4′) is hopelessly ambiguous and might mean as many different things as there are entities in the universe, since anyone of these entities is a candidate for value assignment to the 1st variable. But because of the presupposition (4b), an utterance of (4′) can only be felicitous if the value assigned to the 1st variable is something that is already presupposed to be a cat. In most ordinary contexts, this will mean that only cats qualify, and the ambiguity is greatly reduced.

Generalizing from the example of (4), I am proposing that the present theory be augmented by the following assumption: Definites contrast with indefinites in yet another respect, aside from their different behaviors w.r.t. Quantifier Indexing and the Novelty Condition: In definites, the descriptive content of the NP is presupposed, whereas in indefinites it is (merely) asserted.

What I have just said about the felicity conditions of definite descriptions is a little different from the widespread assumption that definite descriptions presuppose existence and uniqueness, and a brief comparison may be called for. While I take (4) to presuppose (4b), the reader may be accustomed to thinking of (4) as presupposing (4c).

(4)   c. $\exists x(cat(x) \wedge \forall y(cat(y) \rightarrow y{=}x))$

(4c) says that, first, there is a cat, and second, there is only one. As far as the first part of this is concerned, it is of course entailed by (4b): the value of x_1 cannot be a cat unless there is one. So I am in agreement with the familiar idea that a definite description presupposes existence.[33] But what about uniqueness, the second part of (4c)? (4b) is clearly compatible with there being any number of cats in addition to the one associated with x_1. It would seem that in order to exclude such excess cats, I would have to amend (4b) to something like (4d):

(4)   d. $cat(x_1) \wedge \forall y(cat(y) \rightarrow y{=}x_1)$

However, I want to argue that no such amendment is called for, because the uniqueness requirement is already implicit in the felicity conditions I have stated so far, and (4d) would merely duplicate it.

In what sense is the uniqueness requirement implicit in our current felicity condition? Put yourself in the role of one who witnesses an utterance of the monosentential text (4) and is trying to understand it. Your task is to identify the intended referent of "the cat." Suppose you have no contextual clues as to the speaker's intentions. All you have is the general knowledge of the felicity conditions I have formulated so far, and of course of the grammar. So you know three things: (a) The intended reading of the utterance is represented by a logical form with the structure of (4′), in which "the cat" has some index or other, call it i. (b) The context furnishes a unique value

---

[33] "Existence" is here used in the sense of being an element of the domain of individuals. Whether that implies existence in the actual world, or whether one can also refer to and quantify over possible individuals that do not actually exist, may be left open here. That question is independent of whether the presupposition of a definite description should be represented as in (4c), or rather as in (4b).

for the variable number i. (c) That value is a cat. (Recall that (b) derives from
the felicity condition formulated in section 3.3 above, and (c) derives from
the added assumption that definite descriptions presuppose their descriptive
content, i.e., that the context-supplied value for the variable must satisfy a
formula like (4b) in this case.) How can you identify the intended referent
on the basis of just these three assumptions? Well, if you are lucky enough
to find that there is exactly one cat, (a) through (c) will imply that that must
be the one. This is how the existence of a unique cat serves as a sufficient
condition for your success in understanding (4).

Is uniqueness also a necessary condition for your success? Apparently
not. You may be aware of the existence of several cats and still be able to
eliminate all but one as unlikely candidates, simply by using some common
sense and by taking the speaker to be reasonably rational and cooperative.
In such a case, there need not be a unique cat; there just needs to be a unique
most likely cat. This appears to be a weaker requirement than would ensue
from a presupposition like (4c) or (4d). However, it is generally understood
by those who posit uniqueness presuppositions in formulations like (4c)
that we must not read this in the sense that there is literally only one cat
in existence. Rather, we are to interpret (4c) as requiring merely that there
be a unique "relevant" cat, or a unique "most salient" cat, or a unique cat
that is the "most likely" referent, or something of this sort. If we take these
qualifications into account, then we can indeed conclude that the listener
who reasons on the basis of (a), (b), and (c) will succeed in understanding
the speaker's utterance of (4) *if and only if* there is a "unique" cat, in an
appropriate sense of "unique."

What I have tried to show here is this: Given only the assumptions
that have already been introduced — i.e., the general felicity condition of
section 3.3, plus the stipulation that definites presuppose their descriptive
content — it can be deduced that some sort of uniqueness requirement will
normally have to be met if communication by means of definite descriptions
is to succeed. So there is no need for an additional stipulation that definite
descriptions carry a uniqueness presupposition. Under the view I am
taking here, one could in principle utter (4) in the presence of several
(equally likely) cats without violating any felicity conditions. But while it
would be "felicitous" in the technical sense of the theory, such an utterance
would suffer from unresolvable ambiguity and hence be unacceptable, which
means, in a natural non-technical sense of that word, "infelicitous."

I ought to note that it is not only "the"-NPs that presuppose their
descriptive content, but non-pronominal definites in general. This includes
examples like "his father," "that cat," "this dog," and "John." For instance,
when a text has a logical form that includes a free occurrence of [his$_j$

father]$_i$, [that cat]$_i$, [this dog]$_i$, or [John]$_i$, then it will carry a presupposition "father-of-$x_j(x_i)$," "cat(x_i)," "dog(x_i)," or "x_i = John," respectively. Perhaps one should even think of pronouns as having rudimentary descriptive content in virtue of their gender. One should then modify the present system so that pronouns undergo NP-Prefixing too, and that a text that contains, say, a free occurrence of "she$_i$" presupposes "female(x_i)." I will continue to ignore gender in this dissertation.

Another question arises in this connection: If "the ζ," "that ζ," and "this ζ" are all three identical with respect to definiteness, descriptive content, and presupposition, then what accounts for the fact that they are not completely interchangeable in every context? (Similarly, why is "she" generally not replaceable by "the female one?") Apparently, something additional must be said about the felicity conditions associated with definites that would take the specific choice of determiner into account. I will not have much to say on this issue in the present dissertation, but will return to it briefly in Chapter III, section 5.3.

6.3   *Non-referring definites and a projection problem for felicity conditions.*

My discussion in the preceding subsection was limited to definites that are free in the logical form of the text they occur in, and there was a good reason for this limitation: What I said about the felicity conditions for free definites does not carry over in any straightforward way to those that are bound. Consider an anaphoric use of "the cat," as in the text (7), construed with the logical form (7′):

(7)   If a dog barks at a cat, the cat always meows.

(7′)

```
                    T
                   / \
                  ∃   T
                      |
                      S
                    / | \
            always_{1,2}  S̄        S
                        /  \      /  \
            if a dog_2 barks at a cat_1   ∃ the cat_1 meows
```
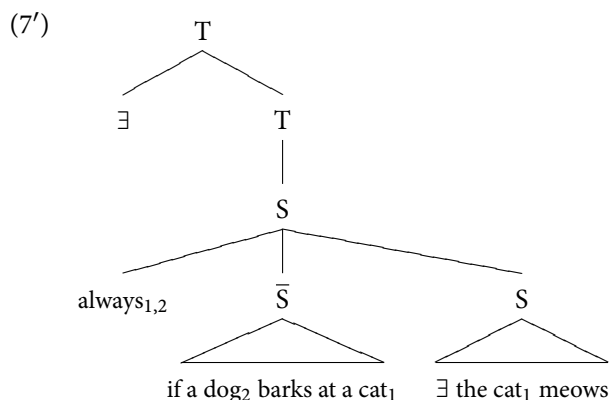
Because "the cat" is anaphoric to a preceding indefinite here, it is bound by the universal quantifier. What does this mean for the felicity conditions that apply to an utterance of $(7')$? For one thing, $(7')$ does not call for the speech context to supply a value for the 1st variable. This is not only predicted by what I said about felicity conditions in section 3.3, but it also accords with our intuition that (7) can be understood without any help from the context. A fortiori, $(7')$ does not presuppose that the value of the 1st variable is a cat. For if it presupposed that, then it could only be uttered felicitously in a context where the discourse participants already presuppose that the value of the 1st variable is a cat. But how could they possibly presuppose that when the 1st variable is not linked to any particular value at all by the context they are in? There is just no sense in which the utterance of $(7')$ as a whole presupposes the truth of "$cat(x_1)$."

The situation we have observed here is not an unfamiliar one in the study of presuppositions: A presupposition that originates with a particular element (here: with the definite NP "the cat_1") may or may not end up being a presupposition of the maximal construction in which that element is embedded. Sometimes, as in the text (4), it does get inherited all the way up the tree; other times, as in (7), it does not. The problem of predicting in a systematic way how the presuppositions of larger constituents depend on the presuppositions of their parts is known as the "projection problem for presuppositions."[34] Generalizing slightly, we may speak of the "projection problem for felicity conditions" to refer to the problem of how the felicity conditions for larger constituents depend on the felicity conditions associated with their parts, where we include under "felicity conditions" both the ones that arise as a matter of presuppositions and the ones that do not.

One of the most successful theories of presupposition projection, especially as far as success in predicting the presuppositions of quantified sentences is concerned, has been developed in a series of papers by Karttunen and by Karttunen and Peters.[35] According to that theory, quantifiers act as "filters." This means that a presupposition which originates inside the scope of a quantifier will not be inherited as it stands by the quantified formula, but it will not disappear without a trace either; rather, the quantified formula will have a presupposition that is calculated in a certain systematic way from

[34] So named by Langendoen and Savin (1971). There are at least two significantly different approaches that have been taken towards a solution of the problem, one originating with Karttunen (1973), the other with Gazdar (1979).

[35] See Karttunen (1973), Karttunen and Peters (1975, 1979).

the presuppositions of its constituent formulas. For example, a universal quantifier is subject to roughly the following rule.[36]

(8) If ψ^2 is the nuclear scope of a universal quantifier Q in the structure

$$\phi$$
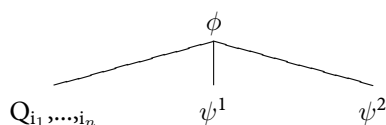
$$Q_{i_1,\dots,i_n} \qquad \psi^1 \qquad \psi^2$$

and ψ^2 carries the presupposition χ, then ϕ will carry the presupposition

$$[Q_{i_1,\dots,i_n}, \psi^1, \chi]$$

Applied to our example (7′), this predicts that (7′) presupposes (9):

(9) $\forall x_1 x_2((\text{dog}(x_2) \wedge \text{cat}(x_1) \wedge \text{bark-at}(x_2,x_1)) \rightarrow \text{cat}(x_1))$,

given that the nuclear scope of "always$_{1,2}$" in (7′), in which "the cat$_1$" is still free, presupposes (4b) "cat(x_1)." (9) is of course a tautology, and is therefore empirically indistinguishable from no presupposition at all; the felicity condition it imposes on an utterance context for (7′) is one that will be trivially met by any context.

What is the point of having a rule like (8) if it only predicts a tautological presupposition? We need to look at another example to see that (8) can do some real work for us. Take (7″), with a structure identical to that of (7′), but a different index on "the cat." (7″) is generated by the construal component as another one of (7)'s infinitely many logical forms.

---

[36] This formulation of the rule fails to deal with cases where the restrictive term, i.e., ψ^1, carries presuppositions too.

(7″)

```
                      T
                   ╱     ╲
                  ∃        T
                           │
                           S
                     ╱     │      ╲
              always₁,₂    S̄          S
                         ╱   ╲      ╱    ╲
              if a dog₂ barks at a cat₁   ∃ the cat₂ meows
```
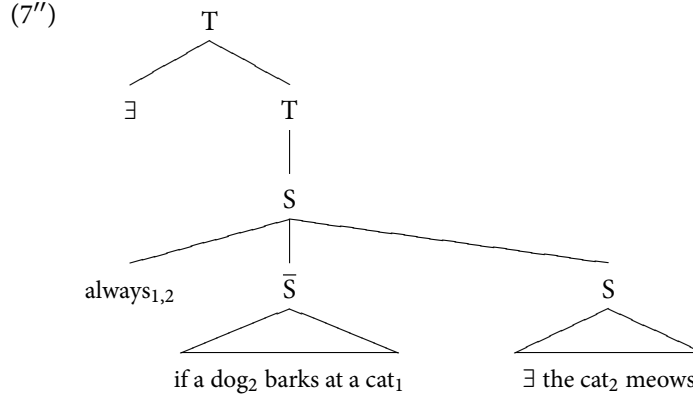
The truth conditions associated with (7″) amount to something that is clearly not an intuitively available reading with any utterance of (7), viz., that every dog that barks at a cat is a cat and meows. Our theory had better yield the prediction that "a dog" is not a possible antecedent for an anaphoric use of "the cat" and that (7″) is ruled out. It turns out that (8) serves to provide us with that prediction. According to (8), (7″) presupposes something silly:

(10) $\forall x_1, x_2 ((dog(x_2) \wedge cat(x_1) \wedge \text{bark-at}(x_2, x_1)) \rightarrow cat(x_2))$

This means that (7″) cannot be a felicitous reading for (7) unless the discourse participants take it for granted that dogs are cats. Since they ordinarily would not, (7″) will not be available and alternative disambiguations of (7) will receive preference.

Up to this point, everything is working fine: the projection of presuppositions from the definite NPs they originate with, to larger and larger constituents, and up to the complete text, may be a rather complicated matter, but it appears to follow rules which have been well described. However, let me point out a problem: There is no satisfying account of presupposition projection in *existentially* quantified formulas. Karttunen and Peters (1979) note explicitly that their theory assigns insufficiently weak presuppositions to existential sentences, and other people have also run into this problem.[37] In our current theory, the problem manifests itself in the following sort of example:

---

[37] Karttunen and Peters (1979, 53) discuss the example "Someone managed to succeed George V on the throne of England." The problem there is how the implicature associated with the verb "manage," i.e., that there was some difficulty to overcome, gets projected from the open sentence "x managed..." to the existentially closed sentence "Someone managed...." Note that this problem arises quite independently of whether a presupposition or implicature, and what sort of implicature, conventional or

(11)   A dog saw a cat. The cat meowed.

(11′)
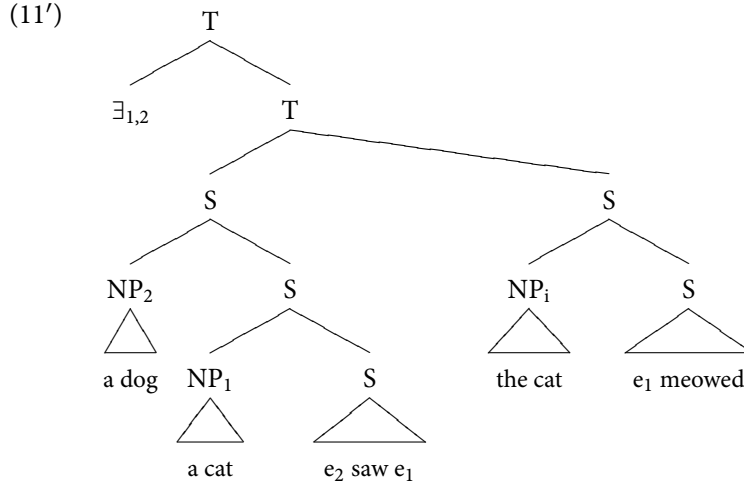
As far as the construal component goes, i could be any number in (11′), including 1 or 2. Intuitively, i = 1 represents an available reading, but i = 2 does not, just like in the analogous case of (7′) versus (7″). We would hope that we could again appeal to the presuppositions that (11′) receives for the various choices of i to explain why i = 1 is okay and i = 2 is out. But what filtering properties should we posit for the existential quantifier to make that kind of explanation work?

A natural first guess would be that rule (8) above is not idiosyncratic to the special case of the universal quantifier, but falls under a rule schema for quantifiers in general. This would lead us to expect that the existential quantifier filters presuppositions from its (nuclear) scope as follows:

(12)   If ψ is the scope of $\exists_{i_1,\dots,i_n}$ and ψ carries the presupposition χ, then $[\exists_{i_1,\dots,i_n} \psi]$ carries the presupposition $[\exists_{i_1,\dots,i_n} \chi]$

Suppose further that the lower T-node in (11′), which is interpreted as the conjunction of its daughters, receives its presupposition according to the rule that is standardly assumed to capture the filtering properties of the connective "and," viz., (13) (cf. Karttunen (1973) and elsewhere):

(13)   If ϕ presupposes χ and ψ presupposes ξ, then $[\phi \wedge \psi]$ presupposes $[\chi \wedge [\phi \to \xi]]$.

---

conversational, is involved. Lerner and Zimmermann (1987) draw attention to the same problem, as it manifests itself within their treatment of presupposition projection.

Applied to (11$'$), this would mean that the lower T-node is associated with presupposition (14), and the topmost, existentially closed T-node, with presupposition (15).

(14) $(\text{dog}(x_2) \wedge \text{cat}(x_1) \wedge \text{saw}(x_2,x_1)) \rightarrow \text{cat}(x_i)$
(15) $\exists x_1,x_2 \, ((\text{dog}(x_2) \wedge \text{cat}(x_1) \wedge \text{saw}(x_2,x_1)) \rightarrow \text{cat}(x_i))$

But (15) is so weak that its truth, even for i = 2, will be taken for granted in most speech contexts. Thus (15) could not be appealed to in an explanation for the unavailability of an i = 2 reading.

So we are led to abandon (12) and come up with an alternative hypothesis about presupposition projection under existential quantification. It looks like a better guess would be that $[\exists_{i_1,...,i_n} \psi]$ carries the universal presupposition $[\forall_{i_1,...,i_n} \chi]$ (ψ and χ as in (12)). Under this hypothesis, (11$'$) would presuppose (16) rather than (15).

(16) $\forall x_1,x_2 ((\text{dog}(x_2) \wedge \text{cat}(x_1) \wedge \text{saw} \, (x_2,x_1)) \rightarrow \text{cat}(x_1))$

(16) is indeed tautological for i= 1 and silly for i = 2, and hence it is the sort of presupposition that would help us to predict the contrast in acceptability for the i = 1 versus the i = 2 readings. However, there are other examples that seem to suggest that existential statements do not in general presuppose something quite as strong as a universal proposition of the sort of (16). For instance, example (17) may well be used with the coindexing indicated:

(17) A dog$_2$ barked at a cat$_1$. The little coward$_1$ ran away.

This does not require a context in which it is presupposed that every cat that was barked at by a dog is a little coward (nor does it introduce such an assumption as an implicature if used in a context that did not contain it as a presupposition beforehand). If this intuition is correct, then we would be overshooting the mark by amending (12) so that it predicts (16) instead of (15). But even if the hypothesis that existential formulas carry universal presuppositions in the manner of (16) were correct factually, it would raise serious problems for a theory that attempts to predict the filtering properties of the various quantifiers from general principles. I cannot be more specific about these problems here, but must refer the reader to the literature on presupposition projection (cf. the references in note 37).

What I hope to have conveyed in this section is that (a) the present approach to definite descriptions is contingent upon some theory of presupposition projection for its proper functioning, and (b) it is not a trivial task to devise such a theory. I will suggest in the course of the next chapter (Chapter III, section 2.4) that the revised theory of semantic interpretation developed

there opens up a much more natural line of approach in this area. A detailed elaboration of that line of approach and its consequences for examples like (17) is not included in this dissertation, but will be provided elsewhere.[38]

6.4 Problems with narrow-scope definites

As definites, definite descriptions should not undergo Quantifier Indexing. But this assumption creates problems as soon as we look at definite descriptions with bound variable pronouns inside them. Consider the following two sentences, both of which contain a definite description in the nuclear scope of "every."

(18) Every man saw the dog.
(19) Every man saw the dog that barked at him.

With (19), consider the reading where "him" is coindexed with "every man." The predicted logical forms for (18) and (19) are essentially alike.

(18′/19′)
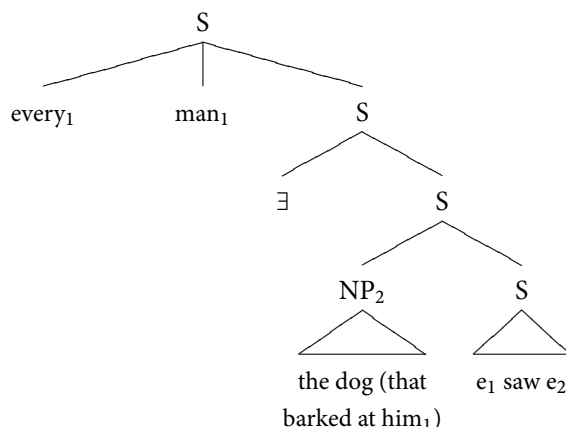
NP_2 remains unbound and is thus predicted to refer to a contextually determined individual, unless the sentence is part of a larger text in which NP_2 has an indefinite antecedent — in that case, it will be existentially bound from outside the sentence. Either way, the predicted reading does not permit that each of the men might have seen a different dog. For (18), this prediction seems to be acceptable: We do tend to read (18) as saying that every man saw the same, context-supplied or previously mentioned dog. But with (19), there is absolutely no implication that all men saw the same dog. We would

---

[38] See Heim (in preparation) [Heim (1983b)]

normally imagine that different dogs barked at different men, and that "the dog that barked at him" need not refer to the same dog for any two men. This intuition is in conflict with the reading that we generate by assigning (19) a logical form like (19′).

The same problem arises with definites in the restrictive term, rather than the nuclear scope, of a quantifier. Consider (20):

(20)   Every man who saw the dog that barked at him fainted.

If definites do not undergo Quantifier Indexing, then the definite description in (20) will remain free instead of getting coindexed with "every." The resulting reading again does not permit the dogs to vary with the men, as they intuitively can.

What can we do to correct this inadequacy? It comes to mind that perhaps definite descriptions sometimes do undergo Quantifier Indexing. If the ones in (19) and (20) did, then those sentences would receive interpretations in which the dogs can vary with the men as desired. In fact, (19) and (20) would then receive precisely the truth conditions of (21) and (22), respectively.

(21)   Every man saw a dog that barked at him.
(22)   Every man who saw a dog that barked at him fainted.

Of course, we would like to predict *some* difference in the meanings of (19) and (20) on the one hand and (21) and (22) on the other. This would presumably have to be a difference on the level of presuppositions. For instance, (19) would apparently have to be assigned a presupposition that for every man there was a dog that barked at him, whereas (21) would lack any such presupposition. The precise implications for presupposition theory remain to be determined.

Definite descriptions in the scope of negation also seem to sometimes behave as though they underwent Quantifier Indexing, despite their definiteness. (23), for instance, creates a paradox if we insist that Quantifier Indexing is limited to indefinites.

(23)   Mary didn't have lunch with the king of France, because France doesn't have a king.

If "the king of France" remains free in the logical form of (23), then we predict that a felicitous utterance of (23) will require the context to furnish a referent for it and the "because"-clause will therefore have to be false. If, on the other hand, we permitted Quantifier Indexing to bind "the king of France" to the existential quantifier that Existential Closure has inserted under the negation operator, we could predict a consistent interpretation.

So it looks like there is quite compelling evidence against the formulation of Quantifier Indexing that I have proposed, according to which the crucial parameter for the applicability of Quantifier Indexing is the feature [±definite]. But what else could it be? Should we assume that only pronouns and proper names are exempt from Quantifier Indexing whereas everything else undergoes it? If that turns out to be the correct generalization, then what does depend on definiteness as the crucial parameter? Only the Novelty Condition and the assignment of presuppositions? These questions are obviously central to a dissertation that purports to elucidate the nature of definiteness. We seem to have arrived at a point where we must re-examine the answers we were heading toward in the earlier sections of this chapter, when we had not extended our attention to nonpronominal definites. By the end of the next chapter, we will be in a better position to do so than now. For the time being, let us put the problems we have raised in this section aside and continue to assume that Quantifier Indexing applies exclusively to indefinites.

7 On So-called "Discourse Referents" and Their Lifespans

Indefinites and their ability to serve as antecedents for anaphora have been my central concern in both this and the preceding chapter. They were also central to some research that Lauri Karttunen did in the late sixties, and part of which was later published under the title "Discourse referents" (Karttunen, 1976).[39] Discourse referents are not referents. But there is more to be said about them than what they are not, and I think that the theory I have begun to develop in this chapter puts me in a better position than other theories have been to reconstruct the concept in the spirit of Karttunen's work. It also puts me in a position to explain some of Karttunen's observations concerning the conditions under which indefinites establish discourse referents and the conditions which determine whether these are "stable" or "short term" discourse referents, although not all of Karttunen's findings will be accounted for.

The closest that Karttunen comes to saying what discourse referents are is this (1976, 366):

> Let us say that the appearance of an indefinite noun phrase establishes a 'discourse referent' just in case it justifies the occurrence of a coreferential pronoun or a definite noun phrase later in the text.

---

[39] See also Karttunen (1968b,a).

It is clear from Karttunen's own comments that discourse referents are not individuals and that to establish a discourse referent does not necessarily mean to refer to anything. In the context of our theory, I propose that we identify discourse referents simply with numbers: to establish a discourse referent means to carry a referential index in logical form. This may be a somewhat odd identification, since the term "discourse referent" certainly sounds like it should apply to something outside the linguistic representation. But as far as the generalizations which the term is designed to help capture are concerned, I believe that referential indices are best suited to play that role in the present theory.

At first sight, an identification of discourse referents with referential indices seems to contradict Karttunen, who says explicitly (1976, 367):

> According to the standard theory, referential indices are merely formal indicators of coreference with no further semantic significance. They are not meant to imply the existence of discourse referents in our sense.

Apparently, he thinks that there cannot be a discourse referent for every referential index, because that would mean that every NP establishes a discourse referent. (After all, every NP has a referential index.) Then even "a car" in the following text would be predicted to establish a discourse referent, a prediction which — according to Karttunen — must be false because it is clearly unacceptable to continue with a coreferential pronoun:

(1) Bill doesn't have a car_i. *It_i is black.

So how can we identify discourse referents with referential indices without substantially altering Karttunen's use of the notion?

The answer lies in the following stipulation, which defines what we mean by a discourse referent's "lifespan":

> *If NP_i is bound by an operator $0_{...i...}$, then the discourse referent i that NP_i introduces ceases to exist outside the scope of $0_{...i...}$. If NP_i is free, then the discourse referent i lives on throughout the entire text.*

Applied to example (1), this implies that the discourse referent i, which gets established by the NP "a car," is "dead" by the time the pronoun "it" comes along. The reason is that Existential Closure applies in the scope of negation (see section 2 above) and that the logical form for (1) looks as follows:

(1′)

```
                    T
                 ╱      ╲
              ∃          T
                      ╱      ╲_____
                    S                            ╲
                 ╱    ╲                            S
               not      S                         ╱╲
                     ╱     ╲                      ╱  ╲
                  (∃ᵢ)       S                   itᵢ is black
                         ╱      ╲
                      NPᵢ         S
                      ╱╲          ╱╲
                     ╱  ╲        ╱  ╲
                    a   car     Bill has eᵢ
```
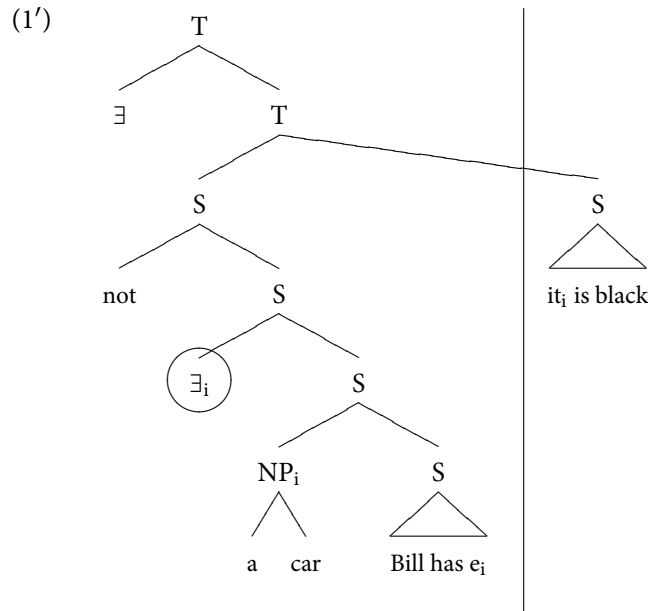
The vertical line marks the end of the scope of the encircled quantifier, and since i is a selectional index on that quantifier, the discourse referent i does not survive beyond that line.

So there is away in which Karttunen and I disagree about (1): Karttunen says that "a car" does not establish any discourse referent at all, whereas I say it establishes an extremely short-lived one. But there is no empirical difference between these two claims: either one predicts that "it" cannot be anaphoric to "a car." We agree on the point that is essential to the notion of a discourse referent: an anaphoric pronoun is possible as long as the discourse referent established by the antecedent lives on.

This essential connection between the lifespan of a discourse referent and the possibility of anaphoric relations is simply stipulated in Karttunen's work; it is part of the definition of "discourse referent," if I understand him correctly. Given the present reconstruction of the notion, it need not be stipulated, but can be deduced from the definition of "lifespan" together with the theorem (ii) concerning "anaphoric relatedness" of section 5.1 above. Both the definition and the theorem say something about configurations of the following type:

$$...\underbrace{[0_{..i..}....NP_i........]}_{\text{scope of } 0_{..i..}}....NP_j....$$

(ii) said that NP_i and NP_j could not be anaphorically related in this configuration, not even if i = j. The definition of "lifespan" says that the discourse referent i has no chance of surviving until NP_j comes along in such a configuration. So it must be a necessary condition for anaphoric relatedness that each NP be within the lifespan of the discourse referent of the other. Provided that the two NPs are coindexed, this is also a sufficient condition for anaphoric relatedness.

I did not introduce the notion of a discourse referent into my theory because it contributes anything of substance. I rather think that any significant generalizations about discourse referents and their lifespans are expressed just as easily in terms of the concepts I have been using so far, especially anaphoric relatedness. I even think that "discourse referent" is a misleading term, aside from being superfluous, because reference has nothing to do with it. The only reason I have attempted to reconstruct the notion within my framework is that it enables me to compare Karttunen's findings to the predictions of my theory.

After surveying a number of environments in which indefinites may occur (negation, quantifiers, various types of sentence-embedding verbs), Karttunen summarizes (1976, 383):

> ... in general, an indefinite NP establishes a permanent discourse referent just in case the quantifier associated with it is attached to a sentence that is asserted, implied, or presupposed to be true, and there are no higher quantifiers involved.

By a "permanent" discourse referent, Karttunen means one that exists from its introduction throughout the rest of the text. In my theory, a permanent discourse referent amounts to a referential index which is either free in the text or bound by a quantifier whose scope encompasses the entire text. It follows immediately that, as far as my theory predicts, permanent discourse referents get established in the following two cases: (a) when a deictic definite is uttered, and (b) when an indefinite is uttered that has wider scope than any quantifier or operator in the sentence of which it is a part.

Let us concentrate on case (b), since Karttunen's generalization is concerned specifically with indefinites. It is a consequence of our rules of construal that all indefinites end up bound by some operator or other. This may be either an operator that is realized syntactically, or an existential quantifier that has been created by Existential Closure.[40] In any case, it will be the lowest operator that c-commands the indefinite in logical form. If and only if no operator with scope over the matrix sentence in which

---

[40] Or, of course, the invisible necessity operator that is associated with bare conditionals.

the indefinite occurs or with narrower scope intervenes will the indefinite be bound by a quantifier with text scope (which is always an existential quantifier) and thus establish a permanent discourse referent. This is what my theory predicts , at any rate.

Does this prediction coincide empirically with Karttunen' s, as quoted above? We cannot tell right away. There are cases where they clearly agree, and other cases where it depends on how we analyze certain phenomena that I have not talked about so far. Among the former cases are those where negation or a "higher quantifier" interfere with the permanence of a discourse referent. We have already considered an example with negation taking scope over the indefinite:

(1)   Bill doesn't have a car$_i$. It$_i$ is black.

As was apparent from the logical form (1′) (see above), "a car$_i$" is bound by an existential quantifier with narrower scope than the negation operator. It is thus not bound by a text-scope quantifier, and therefore does not establish a permanent discourse referent. In Karttunen's words, the quantifier "associated with" the NP "a car" (we would say: the quantifier that "binds" it) "is attached to" (we would say: "has scope over") a sentence that is under the scope of negation. Since a negated sentence is neither asserted, nor implied, nor presupposed to be true, the quoted generalization implies that the indefinite will not establish a permanent discourse referent.

For an example with a "higher quantifier," take Karttunen's (2):

(2)   Harvey courts a girl$_i$ at every convention. She$_i$ is very pretty.

The reading we are interested in is the one where "every convention" is given wider scope than "a girl," i.e., where the universal quantifier is "higher" than the indefinite. Under this reading, the pronoun "she" cannot be understood as anaphoric to "a girl," and this shows that "a girl" has not established a permanent discourse referent. Karttunen subsumes this observation under the generalization that any higher quantifier will prevent an indefinite from establishing a permanent discourse referent (see quote above). It is consistent with this generalization that the alternative reading, i.e., the one with "every convention" taking narrower scope than "a girl," does permit "she" to be anaphoric to "a girl": in that case, no higher quantifier intervenes and a permanent discourse referent can get established.

The same predictions about (2) follow from my theory: "a girl" must be bound by the text-scope quantifier if it is to establish a permanent discourse referent. But it will only be bound by the text-scope quantifier if it is read with wider scope than "every." Otherwise, it will get bound by the quantifier which Existential Closure introduces inside the nuclear scope of "every" and

whose scope is too small to include the subsequent "she," let alone the entire text.

So far, we have found Karttunen's generalization concerning the conditions under which an indefinite establishes a permanent discourse referent in agreement with the predictions of my theory. Notice that my theory subsumes under a common explanation what Karttunen presents as two unrelated facts: Indefinites under negation fail to establish permanent discourse referents, and so do indefinites under higher quantifiers. From the point of view of my theory, the reason is the same in the two types of cases: both negation operators and higher quantifiers trigger Existential Closure in their (nuclear) scopes and thus make any indefinites that occur under them unavailable for binding by the text-scope existential quantifier. Karttunen, on the other hand, offers no common explanation for the lifespan-limiting effect of negation and higher quantifiers. In fact, he offers no explanation at all for why the conditions under which permanent discourse referents get established should be what he observes them to be. His paper draws attention to the facts and offers some generalizations, but it stops there. I have gone farther than that by deducing at least some of his observations from deeper principles.

But we have yet to look at a number of other environments that Karttunen lists as obstacles to the establishment of a permanent discourse referent. How, for instance, can I account for the impossibility of anaphora in the following examples (all from Karttunen, op. cit.)?

(3)   Bill can make a kite$_i$. It$_i$ has a long string.
(4)   John wants to catch a fish$_i$. Can you see it$_i$ from here?[41]

For Karttunen, these fall under the generalization that for a permanent discourse referent to get established, the indefinite must occur in a clause that is asserted, implied, or presupposed. Modals (e.g., "can") and non-factive propositional attitude verbs (e.g., "want") are interpreted in such a way that the speaker does not commit herself to the truth of the complement clause in any way, neither by asserting it, nor by implying or presupposing it. That is why, according to Karttunen, indefinites under the scope of "can" or "want" fail to establish permanent discourse referents and why the subsequent pronouns in the above examples cannot be anaphoric to the indefinites.

---

[41] With each of these two examples, we are considering only one of the two available readings, i.e., the one where the indefinite has narrowest scope (is "nonspecific," as Karttunen says). For the wide scope readings, anaphora is possible, indicating that a permanent discourse referent has been established.

My explanation is this: "can" and "want" are operators. Hence they trigger Existential Closure inside their nuclear scopes. Hence any indefinites under the scope of "can" or "want" will get bound within the sentences and be inaccessible to binding by the text-scope existential quantifier. Hence there will not be a permanent discourse referent. This is of course once again the same explanation that I applied to examples (1) and (2), where negation and quantifiers were involved instead of modals or propositional attitude verbs.

In what sense are modals and propositional attitude verbs operators? As for modals, I refer the reader back to section 4. Propositional attitude verbs are similar (at least as long as we idealize away from the fact that cointensional expressions are not always interchangeable salva veritate in propositional-attitude-contexts, which we will). The verb "want," for instance, can be analyzed as indicated in the following paraphrase of "John wants to catch a fish": "For every world w such that w is compatible with what John wants: John catches a fish in w." Let us assume that the logical form looks roughly like this:

(4′)

```
                         S
                  _____/ _____
                 /               \
              NP₁                 S
               |            _____/ _____
             John         /               \
                        e₁                 VP
                                     _____/ _____
                                    /               \
                                   V                 S
                                   |           _____/ _____
                                 want        /               \
                                            ∃₂                S
                                                       _____/ _____
                                                      /               \
                                                    NP₂                S
                                                    /\               /  \
                                                   /__\             /____\
                                                  a fish       PRO₁ to catch e₂
```
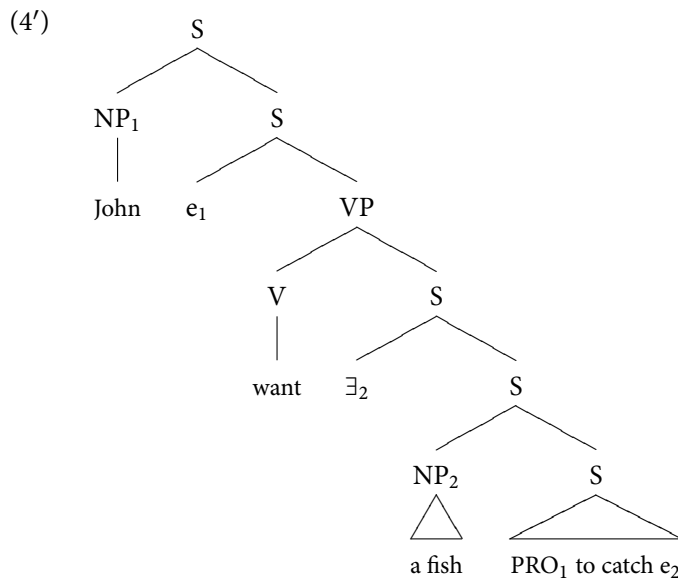
Let us further assume that "want" qualifies as an operator, and that its complement is its nuclear scope, hence a domain for Existential Closure. (There is no overt restrictive term.) The intended interpretation is approximately that a sentential constituent of the form:

```
              S
            /   \
          NP     VP
          |     /  \
          α   want   X
```
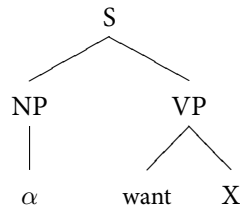
is true iff X is true in every world that is compatible with what the individual denoted by α wants. Whatever the details of this treatment of propositional attitude verbs, their status as operators and the resulting applicability of Existential Closure will ensure that any indefinites inside their complements are unable to enter into anaphoric relations with NPs outside of them.

If this is the correct explanation for why "a fish" does not establish a permanent discourse referent in (4), then how come "a car" in (5) (again an example from Karttunen, op. cit.) does establish one?

(5)   John knew that Mary had a car$_i$. But he had never seen it$_i$.

Here "it" can be anaphoric to "a car." Karttunen observes that the relevant difference between (5) and (4) is that "know" is a factive verb, meaning that its complement is understood as presupposed to be true, whereas "want" is non-factive.   Indefinites in presupposed clauses establish permanent discourse referents, while indefinites in non-presupposed clauses do not.

Why should this be so if my theory is correct? As it stands, this theory predicts no difference between (4) and (5). Certainly, "know" is as much an operator as "want" (its analysis being roughly): "for every world which is compatible with what α knows: ..."), and the logical form of the first sentence of (5) is in all relevant respects like (4′). I will return to this issue in Chapter III, section 5.3.   For the time being, I have to concede that Karttunen's generalization is here ahead of my theory in the amount of facts it covers.

There are more facts among those discussed by Karttunen that I do not yet have an explanation for, especially facts concerning "short-term" discourse referents, as Karttunen calls them. The quote I gave above concerns only the conditions under which *permanent* discourse referents get established. But what about discourse referents with shorter lifespans? Karttunen has examples like (6) and (7), where one of the two pronouns admits of an anaphoric relation with the indefinite, whereas the other one does not.

(6)   You must write a letter$_i$ to your parents and mail it$_i$ right away. They are expecting it$_i$.
(7)   John wants to catch a fish$_i$ and eat it$_i$ for supper. Do you see it$_i$ over there?

The behavior of these examples is predictable on the basis of our theorem on anaphoric relatedness: In each case, the indefinite occurs in the scope of the operator ("must" and "want," respectively), and the first pronoun is also in the scope of that operator, while the second pronoun is outside. From this it follows that the first, but not the second, pronoun can be anaphoric to the indefinite. Or, expressed in the terminology we have adopted from Karttunen, it follows that the discourse referent established by the indefinite remains alive until the end of the scope of the modal or verb, but not beyond. (6) and (7) thus present no problem for us; they conform to our stipulation (see above) that the lifespan of a discourse referent i coincides with the scope of the operator which binds NP_i (if any). However, Karttunen points out a number of examples where this stipulation fails and where the lifespan of a discourse referent is longer than we would expect. For instance:

(8) You must send a letter$_i$ to your parents. It$_i$ has to be sent by airmail.
(9) Mary wants to marry a rich man$_i$. He$_i$ must be a banker.
(10) Harvey courts a girl$_i$ at every convention. She$_i$ always comes to the banquet with him.

In each of these, you are to read the first sentence in such a way that the operator (i.e., "must," "wants," and "every," respectively) takes wider scope than the indefinite. In accordance with what we have said so far, the discourse referent established by the indefinite in each example should live only inside the scope of that operator, and therefore the subsequent pronoun should not admit of an anaphoric interpretation anymore than it did in the parallel examples (3), (4), and (2). But in fact, it does admit of an anaphoric interpretation, which suggests that the discourse referent has lived on in spite of the limited scope of the operator.

   In comparing examples like (8), (9), and (10) to their minimally different counterparts like (3), (4), and (2), respectively, Karttunen points out that in each of the former, but in none of the latter, the second sentence (the one with the pronoun in it) is itself under an operator, and that the operator in the second sentence is in some sense "of the same type" as the operator in the first sentence. In (8), for instance, the "must" of the first sentence is matched by a "has to" in the second sentence. In (9), a "must" (which here is understood as expressing buletic necessity) follows up on the "wants." In (10), the universal quantifier "always" succeeds the universal quantifier "every." In each case, the anaphoric pronoun is in the scope of that second operator, and that is what makes anaphora possible after all. As Karttunen puts it: the lifespan of a short-term discourse referent which has been established inside the scope of a quantifier "may be extended [beyond the scope of that quantifier — I.H.] by flagging every successive sentence with a quantifier of the same type" (op.

cit., p. 377). My theory cannot predict this observation, and I will have to leave it unaccounted for in the present chapter; but see Chapter III, section 5, for a first attempt at an explanation.

I hope that this section has not only clarified the relation between Karttunen's work on indefinites and anaphora and mine, but has also convinced the reader that insight is gained by reducing the concept of a discourse referent and its lifespan to certain formal properties of logical form, rather than using it as an unanalyzed concept as Karttunen did.

By translating Karttunen's observations as to when indefinites establish discourse referents and how long-lived they are into my own theory, I have been able to do more than reformulate them: A substantial core of his generalizations have turned out to follow from principles that I had already built into my system of construal rules. The remaining generalizations, which I could not explain, show that more work is to be done, of course, and some of that will hopefully be done by the end of the dissertation.

Aside from explaining some facts that Karttunen could only describe, I hope to have dispelled some of the mystery that the notion of a discourse referent is likely to have presented, especially to philosophically minded semanticists: If they are not referents, what are they? What is their relation to referents where there are any? Such questions are difficult if it is presumed that discourse referents, although not referents, are individuals of some sort. But under the current reconstruction — or elimination, if you like — of the concept, they are no more puzzling than the question: What is the meaning of a variable?

8   Concluding Remarks Towards a Theory of Definiteness

The title of this dissertation is "The Semantics of Definite and Indefinite Noun Phrases" and is meant to raise two related, but distinct, questions:

(a) What are the semantic characteristics, if any, that definites and indefinites share with each other, but not with other NPs?

(b) What is the semantic interpretation of the definite-indefinite contrast?

Question (a) can of course only be posed in this way if it is assumed that some NPs are neither definite nor indefinite. If we, for instance, call "a cat" indefinite, "the cat" and "it" definite, and "every cat" and "no cat" neither, than (a) amounts to a question that I have been preoccupied with throughout the first and much of the current chapter. Taking issue with the prevailing view among logical semanticists that definites and indefinites have nothing essential in common that would set them apart from "every"- and "no"-NPs,

I have proposed that the major semantic subdivision of NPs into quantifying NPs and quantifier-free (or "variable-like") NPs runs as follows:

QUANTIFYING

NO	YES
a cat	every cat
the cat	no cat
it	

In arguing for this way of drawing the line, I have concentrated mostly on the indefinite, whose reclassification from the right column into the left column is particularly objectionable at first sight. Chapter I was devoted to exploring the problematic consequences of classifying "a cat" as quantifying, and in the present chapter I have laid out some consequences of the alternative view that "a cat" does not quantify, most notably, the desirable consequence that we receive a straightforward treatment for donkey sentences.

What about question (b), the question about the semantic contrast that makes definites and indefinites distinct from each other? This question was not under consideration in Chapter I, and it was not a focus of attention at the outset of the current chapter. However, as soon as we had chosen to answer question (a) in the way we did, i.e., by grouping "a cat," "the cat," and "it" together as non-quantifiers, we were confronted with the problem of drawing distinctions *within* that group. To distinguish between different non-quantifiers is prima facie not as easy as it is to distinguish between different quantifiers, where you can simply appeal to differences in quantificational force. When there is no quantificational force, you have to come up with distinctive semantic features of some other kind. There are two that suggest themselves, and I have exploited both for one purpose or another in the course of this chapter: One is "descriptiveness," i.e., the presence or absence of descriptive content which distinguishes full NPs like "a cat" and "the cat" from pronominal ones like "it." The other one is definiteness, which sets "the cat" and "it" apart from "a cat." So definiteness has come to play a role in the theory, and this has happened as an indirect consequence of the decision to regard definites and indefinites as alike in their lack of quantificational force.

We are thus finding ourselves in the midst of designing a theory in which the definite-indefinite distinction plays a crucial role. But we have not until now attended explicitly to the question how this distinction is reflected in the semantic interpretation of definite and indefinite NPs, i.e., to question (b). Drawing together various assumptions that have been introduced in the course of this chapter, we arrive at the following preliminary answer:

Definiteness amounts to a cluster of three properties: (i) Definites do not undergo the construal rule of Operator Indexing. (ii) Definites are not subject to the Novelty Condition. (iii) Definites presuppose their descriptive content (if any). Indefinites are characterized by the opposite property in each of these three respects: they undergo Operator Indexing, are subject to the Novelty Condition, and do not presuppose their descriptive content. The semantic contrast between definites and indefinites is a product of these three characteristic differences, which jointly affect felicity and truth conditions in a more or less indirect way.

By associating the feature [+definite] with the conjunction of properties (i) through (iii), and the feature [-definite] with the conjunction of their negations, we have started to develop what may be called a "theory of definiteness." In its present state, this theory leaves various questions open that would seem to be relevant. In particular, one should like to know whether there are types of NPs that are neither definite nor indefinite. Note that this is certainly a logical possibility, since (i) through (iii) are mutually independent properties (none of them entailed by the other two). One group of NPs that might be unspecified for definiteness are the quantifying NPs, though we have entertained the idea that they are [-definite] (cf. sections 2 and 5.3). So it is consistent with our assumptions, if not very well motivated at this point, that quantifyingness and definiteness crossclassify NPs as follows:

		QUANTIFYING	
		NO	YES
DEFINITE	NO	a cat	every cat no cat
	YES	the cat it	

Leaving aside quantifying NPs, it is still by no means a matter of course that the remaining NP-types, i.e., the non-quantifying ones, should divide up exhaustively into definites and indefinites, i.e., that every non-quantifying NP should either have all three of the properties (i) through (iii), or all three of the opposite properties. In languages other than English, say Latin, we seem to have a third option: articleless NPs that may function as either definite or indefinite. If we include this third option, we are still far from the full range of conceivable possibilities. So suppose we were to hypothesize that every (nonquantifying) NP that exists in a natural language falls into one of these three groups:

(a) **definites** = those that exhibit (i), (ii), and (iii);

(b) **indefinites** = those that exhibit the opposites of (i) through (iii);

(c) "**ambiguous**" = those which have definite as well as indefinite occurrences, i.e., occurrences that have all of (i) through (iii) and occurrences with all of the three opposite properties.

This would be a hypothesis about a substantive language universal. It could be falsified, and in fact, we have already looked at prima facie falsifying evidence from English definite descriptions. Recall our discussion in section 6.4, where we had the impression that sometimes definite descriptions undergo Quantifier Indexing,while still exhibiting definite behavior w.r.t. felicity conditions.[42]

Let us disregard those cases for the time being, although they are of course important and must be dealt with eventually. (I will return to them in Chapter III, section 5.) We may then speculate that the hypothesis is basically true, i.e., that the properties (i) through (iii) are in fact an inseparable cluster. Given that state of affairs, we are tempted to look more deeply into the question whether (i), (ii), and (iii) should really be stipulated as three independent properties of definites. Is there not some unique, more basic principle from which all three of them could be derived? This question will be one of the motivating concerns behind the next chapter. There I will develop a rather different and much stronger version of the Novelty Condition, and will attempt to relate all three components of the definite-indefinite contrast to this new condition.

---

[42] "Specific" indefinites, as discussed in section 5.4 above, might constitute another group of NPs that fall outside the three options (a)-(c), thus falsifying the hypothesized universal.
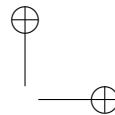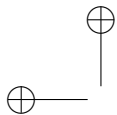
Chapter III

DEFINITENESS IN FILE CHANGE SEMANTICS

In this chapter, I will recast the analysis of indefinites, pronouns, and quantifiers that I developed in the preceding chapter in a somewhat different conceptual framework. The main difference will consist of an altered division of labor between "grammar" in a narrow sense and semantics. The construal component of the grammar will become simpler and less important, while an enriched theory of semantic interpretation will take over some of the burden that used to be carried by the construal component. In particular, I will eliminate Existential Closure, Quantifier Indexing, and the Novelty Condition from the construal component and will instead build a theory of definiteness into the principles which interpret logical forms, i.e., which associate them with felicity and truth conditions.

1 Informative Discourse and File-Keeping

1.1 Introduction

Suppose we look at language as a means of conveying information from a speaker to an audience. What is the specific contribution that definite NPs on the one hand, and indefinite NPs on the other, make towards this purpose, the transmission of information? Let us approach this question with a naive mind, as though we had forgotten that we have already been in the midst of formulating an analysis of definites and indefinites in a particular theoretical framework.

Imagine that A is talking, while B is listening and trying to understand what A says as they go along. For "understand," we could say: extract and retain the information that A's utterances contain. Metaphorically speaking, B's task is to construct and update a file which, at any point in the
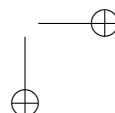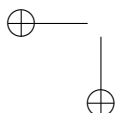
conversation, contains all the information that A has conveyed up to that point.[1] A is uttering the following text:

(1) (a) A woman was bitten by a dog. (b) She hit him with a paddle. (c) It broke in half. (d) The dog ran away.

At the beginning of A's utterance, B's file is empty. (We will soon see that in reality, no conversation starts with a completely clean slate like this, but let us pretend for now that this one does.) After A has uttered (a), B takes two new cards and gives each a number, say, 1 and 2. On card 1, B writes: "is a woman" and "was bitten by 2." On card 2, B writes: "is a dog" and "bit 1." So now B's file consists of these two cards, with these things written on them. Next, A utters (b). B takes another new card, numbers it "3," and writes on it: "is a paddle" and "was used by 1 to hit 2." Moreover, B updates card 1 by adding on it: "hit 2 with 3," and card 2, by adding: "was hit by 1 with 3." Then utterance (c) comes along, and B updates the file by adding on card 3: "broke in half." Then comes (d), and B adds on card 2: "ran away."

   With this illustration in mind, let us repeat the question: What are the particular roles that indefinite and definite NPs respectively play in helping B to extract the information contained in A's utterances? Apparently, B treats the two types of NPs according to the following rule: *For every indefinite, start a new card; for every definite, update a suitable old card.* (The second half of this rule is of course vague because it does not say how to determine suitability. We will have to put up with this vagueness for the time being, but will eventually clarify some of the issues involved in suitability decisions.) This rule, metaphoric and imprecise though it is, contains in a nutshell the theory of definiteness that I am going to adopt. As I will argue, the association of indefiniteness with a "new card" and definiteness with an "old card" is not just one among other contrasts that correlate with the definite-indefinite distinction, but the most basic contrast, from which others can be made to follow. Before this theory can take shape, we have to clarify what the metaphorical file is a metaphor for, and how it fits into a truth-conditional semantics of the sort we have been assuming.

1.2   *How files relate to facts, and how utterances change them*

A file can be evaluated as to whether it corresponds to reality or misrepresents it. For instance, if a file contains a card on which it says: "is a brother

---

[1] This metaphor must have been used many times before. Its usefulness for the present purposes was brought to my attention by Angelika Kratzer.

of Mary's," but Mary is in fact an only child, then that file is not consistent with the facts. It is "false," as we will say.

In order to establish the truth of a file, we must find a sequence of individuals that *satisfies* it. A sequence of individuals satisfies a file if the first individual in the sequence fits the description on card number 1 in the file, the second individual in the sequence fits the description on card number 2, and so on, for each number that is the number of some card in the file. Consider, for example, the file F, which consists of two cards, numbered 1 and 2, where on 1 it says: "is a woman" and "was bitten by 2," and on 2 it says: "is a dog" and "bit 1." (F is the file that B had constructed after A's first sentence in the example above.) Consider further a sequence a_N whose first member, a_1, is a woman, whose second member, a_2, is a dog, and a_2 bit a_1. (What the remaining members of a_N are like is irrelevant.) a_N satisfies F. On the other hand, any sequence whose first member is not a woman, or whose second member is not a dog, or whose second member did not bite its first member, fails to satisfy F.

In order for a file to be consistent with the facts, there has to be at least one sequence that satisfies it. We will call a file that is consistent with the facts "true," and one that is not "false." So F is true iff there is a sequence whose first and second members are a woman and a dog that bit her. Such a sequence of course exists if and only if there is a woman who was bitten by a dog. F is false just in case no dog bit any woman.

Let us look again at our sample conversation between A and B. We can distinguish five different stages in this conversation, each of which is characterized by a different file. (I am now using "file" in the sense of "momentary state of a file." So a conversation is associated with a sequence of distinct files, rather than with one file whose properties change.) At the initial stage, before anything has been said, the file is F_0; after A has uttered sentence (1a), it is F_1; and so on up to F_4, which obtains after the utterance of (1d) has been completed. Depending on what reality is like, each of these files specifies a particular set of sequences which includes all and only those sequences that satisfy it:

F_0 A^n (i.e., the set of all sequences whatsoever)
F_1 {a_N: a_1 is a woman, a_2 is a dog, and a_2 bit a_1}
F_2 {a_N: a_1 is a woman, a_2 is a dog, a_3 is a paddle, a_2 bit a_1, and a_1 hit a_2 with a_3}

I will call the sets on the right the "satisfaction sets" of the files on the left and refer to them as $\text{SAT}(F_0)$, $\text{SAT}(F_1)$, etc.

Since we are thinking of files as recording information that has been conveyed by utterances, there has to be an intimate connection between

what an utterance means and how it causes the file to change. Roughly, each utterance leads to a file with a smaller satisfaction set than the previous file, as all those sequences that are "incompatible" with the utterance get eliminated. For example, the change from F_1 to F_2, which is caused by A's uttering: "She hit him with a paddle," amounts to discarding from the satisfaction set all those sequences which fail to have a paddle as their third member and/or fail to be such that the first member hits the second with the third. They are discarded because the utterance is not compatible with such sequences. Let us try to characterize this connection between utterance meaning and file change in somewhat more precise terms.

I just said that sequences of a certain sort (e.g., those whose third members are not paddles) were "incompatible" with the meaning of A's utterance of sentence (1b). What does "incompatible" mean in this context? It means that the sequences in question fail to satisfy sentence (1b) under the reading with which A uttered it. Recall the relation of satisfaction between sequences of individuals and disambiguated sentences of English that was defined in section 3 of Chapter II above. Suppose A uttered the text (1) under a reading which corresponds to the following logical form:

(1′)
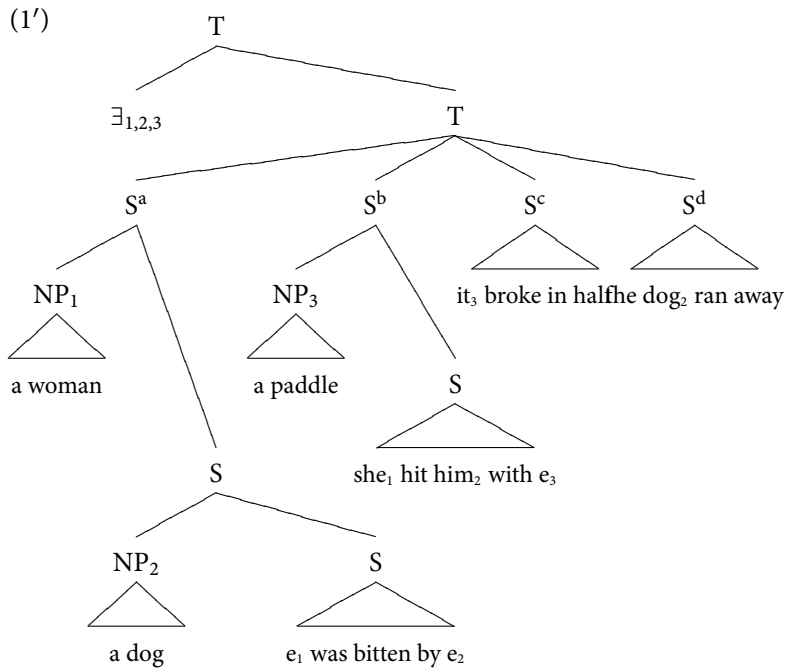
For the whole formula (1′), as well as for each of its subformulas, we can calculate a satisfaction condition on the basis of Ch. II, section 3.2. Here we

are particularly interested in the subformula dominated by S^b, which is the logical form of sentence (1b). We calculate:

$$a_N \text{ } \underline{\text{SAT}} \text{ } S^b \text{ iff } a_3 \text{ is a paddle and } a_1 \text{ hit } a_2 \text{ with } a_3.$$

Given this, we can describe the change from F_1 to F_2 as follows:

$$\text{SAT}(F_2) = \text{SAT}(F_1) \cap \{a_N: a_N \text{ } \underline{\text{SAT}} \text{ } S^b\}$$

In general terms, the satisfaction condition of an utterance relates in the following way to the file change which that utterance brings about:

(A) If sentence S is uttered under the reading represented by logical form ϕ and F is the file that obtains at that stage of the conversation at which the utterance occurs, and F' is the file that obtains after the utterance, then the following relation holds between F and F':

$$\text{SAT}(F') = \text{SAT}(F) \cap \{a_N: a_N \text{ } \underline{\text{SAT}} \text{ } \phi\}$$

With principle (A), we have basically clarified where truth-conditional semantics fits into a view of conversation as a series of file changes.

We are now ready to reconsider some of the principal assumptions which figured in the analysis of indefinites that I outlined in Chapter II. We will start with some reflections on the status of the Novelty Condition (section 2) and then proceed to Existential Closure and Quantifier Indexing (sections 3 and 4). But before all that, I should comment a little on the relationship between my "files" and certain other notions that have been employed in the philosophy of language, linguistics, and Artificial Intelligence research.

1.3   File cards as discourse referents

I am not the first person to have viewed conmunication as somehow analogous to file-keeping. Karttunen's paper on "Discourse referents," for example, starts out with the following words (Karttunen, 1976, 364):

> Consider a device designed to read a text in some natural language, interpret it, and store the content in some manner, say, for the purpose of being able to answer questions about it. To accomplish this task, the machine will have to fulfill at least the following basic requirement. It has to be able to build a file that consists of records of all the individuals, that is, events, objects, etc., mentioned in the text and, for each individual, record whatever is said about it. Of course, for the time being at least, it seems that such a text interpreter is not a practical idea, but this should not discourage us from studying in abstract what kind of

> capabilities the machine would have to possess, provided that our study provides us with some insight into natural language in general.
>
> In this paper, I intend to discuss one particular feature a text interpreter must have: that it must be able to recognize when a novel individual is mentioned in the input text and to store it along with its characterization for future reference.

The problem that Karttunen has formulated here is also one of the main problems that will occupy us in this chapter: What are the conditions under which a new card is added to the file? But whereas Karttunen speaks of file cards merely as a didactic device to motivate his readers, and soon abandons the metaphor in favor of his newly created theoretical concept of a "discourse referent," I use "file" and "file card" themselves as theoretical concepts. I think that the difference is merely terminological: what Karttunen calls "discourse referents" are, I suggest, nothing more and nothing less than file cards. Some people might disagree with this identification and maintain that discourse referents are something beyond file cards, that they are what the file cards describe. But such a distinction gains us nothing and creates puzzling questions: File cards usually describe more than one thing equally well. For example, if a card just says "is a cat" on it, then this description fits one cat as well as another.

So if discourse referents were literally the things that file cards describe, then we could not assume there to be a one-to-one correspondence between file cards and discourse referents. But Karttunen clearly makes this assumption, e.g., when he speaks of an indefinite NP as introducing *a* discourse referent, not a *set* of discourse referents. Karttunen is of course well aware of such puzzles and therefore takes care to distinguish discourse referents from the actual cats and cars and fish that happen to fit the descriptions on file cards. Discourse referents are supposed to be of a more abstract nature than that. I certainly agree that discourse referents must not be confused with the things that the file cards describe. But I see no reason to distinguish them from the file cards. Rather, I find that identifying them with file cards does away with questions as to their ontological status that are at best uninteresting and at worst confusing.

So let us assume that Karttunen's discourse referents are the same as our file cards. But did I not already propose a different reconstruction of Karttunen's notion in Chapter II, section 7? There I argued that discourse referents were best identified with referential indices. The two proposals seem to contradict each other, but it turns out that they do not. Referential indices are numbers, and file cards can also be identified with numbers. That is the way we have been referring to them anyway: "card number 1," "card number 2," etc. More importantly, there is a systematic correspondence

between referential indices and file-card numbers: the referential index of the NP uttered determines which file card is affected (i.e., introduced or updated) by the utterance. This correspondence was already apparent in our informal treatment of examples. (Recall how text (1), uttered under the reading represented by logical form (1′), was said to affect the file.) For the general case, it follows from principle (A). Therefore, it makes no difference whether we think of an NP's discourse referent as its referential index or as the card affected by its utterance: they always correspond to the same number.

In Chapter II, I equated discourse referents with referential indices, because of all the concepts that were available in my theory then, referential indices were best suited to play the role which Karttunen designed his discourse referents to play. Now that I have made new concepts available, we might as well equate discourse referents with file cards. Technically, it comes to the same thing, at least for now (but see section 2.2, where we will loosen the connection between referential indices and file cards in a certain way). Intuitively, file cards are better suited to the role of discourse referents than referential indices. The notion of a discourse referent's "lifespan" is grasped much more easily, once we think in terms of file cards: it is simply the period that starts with a card's addition to the file and ends (if ever) with that card's elimination from the file. Adding and eliminating cards are most natural operations in the context of file-keeping. (Their precise impact on the formal properties of files will be clarified in due course.)

Not surprisingly, Karttunen's work has felt congenial to Artificial Intelligence researchers, whose goal is precisely to design the automatic text-interpreter that Karttunen evokes in the introduction to his paper. Webber, for instance, uses a notion of "discourse model" that appears to be equivalent to my notion of a file. She calls file cards "discourse entities." Consider the following passage from the introduction to her thesis (1978, 21):

> Informally, a discourse model may be described as the set of entities 'naturally evoked' by a discourse and linked together by the relations they participate in. These I will call discourse entities. (I can see no basic difference between what I am calling 'discourse entities' and what Karttunen [1976] has called 'discourse referents.' My alternate terminology rests on wanting to keep 'referent' a separate technical term.)

While I share Webber's discomfort with Karttunen's terminology, I find that her alternative gets us from the frying pan into the fire. "Entities," just like "referents," makes one imagine the "things in the real world" that Webber clearly says must not be confused with discourse entities in her sense (see op. cit., p. 28, n. 3 and elsewhere). There are remarks in Webber's

thesis which unmistakably set discourse entities apart from anything like individuals, e.g., the remark (op. cit., p. 24) that "discourse entities are basically no more than hooks for descriptions." But she also uses rather misleading ways of talking: She ascribes to discourse entities such properties as being-a-T-shirt and having-been-mentioned-by-the-speaker (see, e.g., p. 36 and elsewhere throughout the thesis), which cannot be applied in a literal sense to them. I hope that my terminology of "files" and "file cards" is less laden with misleading associations, but I want it to be understood that the concepts behind those terms and the recognition of (at least some of) their theoretical importance are by no means without precedent in the linguistics and Artificial Intelligence literature.

1.4   Files as common ground

This section places the notion of a file in relation to ideas that have grown out of work concerned with such topics as presupposition, context-dependency, and illocutionary force. I will refer particularly to Grice's notion of "common ground," better known as "pragmatic" or "speaker's presupposition" in the sense of Stalnaker (1974) and to Stalnaker's theory of assertion (1979).

Every utterance occurs in a context. A context has all sorts of properties: e.g., it is located at a particular time, in a particular place, in a particular possible world, and it involves a certain speaker and certain addressees, to mention only some of the parameters along which contexts can differ from one another. Another parameter is the "common ground" or "speaker's presupposition," which Stalnaker explicates briefly in the following passage (1979, 321):[2]

> [...] the concept of speaker presupposition. This, I want to suggest, is the central concept needed to characterize speech contexts. Roughly speaking, the presuppositions of a speaker are the propositions whose truth he takes for granted as part of the background of the conversation. A proposition is presupposed if the speaker is disposed to act as if he assumes or believes that the proposition is true, and as if he assumes or believes that his audience assumes or believes that it is true as well. Presuppositions are what is taken by the speaker to be the COMMON GROUND of the participants in the conversation, what is treated as their COMMON KNOWLEDGE or MUTUAL KNOWLEDGE. The propositions presupposed in the intended sense need not really be common or

---

[2] For more detail on this concept, see Stalnaker (1973, 1974). According to Stalnaker (1979, 321), Grice spoke of assumptions with common ground status in this sense in the William James Lectures (Grice, 1967).

> mutual knowledge; the speaker need not even believe them. He may presuppose any proposition that he finds convenient to assume for the purpose of the conversation, provided he is prepared to assume that his audience will assume it along with him.

Strictly speaking, each person has his or her own presuppositions in this sense. Nevertheless, I will talk as if all participants of a conversation presuppose the same things in any one context. This will permit me to speak simply of the common ground *of a context*, which I take to be identical to each participant's presuppositions in that context. (In Stalnaker's terms, this means that I am considering only "nondefective" contexts.)

Stalnaker construes a speaker's presuppositions (or the common ground) as a set of possible worlds, viz., the set of all and only those possible worlds which are compatible with everything that the speaker presupposes. He calls this set the "context set" of the context under consideration. That is where I have to depart from him. I propose that the common ground of a context be identified with what I have been calling the "file" of that context. As we will see, files cannot be construed as sets of possible worlds, although each file determines such a set.

How does it make sense to identify common grounds with files? Well, a common ground divides the set of all possible states of affairs into those that are compatible with it and those that are not. So does a file: We have already shown how to evaluate files in terms of truth and falsity with respect to a given world. (See section 1.2. There I spoke of truth and falsity simpliciter, i.e., truth and falsity w.r.t. the actual world. But it is obvious that we can speak in the same way of truth and falsity w.r.t. any other world.) As I have defined it, a file is true in a world iff it has a non-empty satisfaction set in that world. So each file determines a set of possible worlds consisting of all and only the possible worlds that are compatible with it: it is the set of all the worlds in which the file has a non-empty satisfaction set. (Again, I simplified matters in section 1.2 by speaking of a file's satisfaction set simpliciter, rather than its satisfaction set in a certain world.)

One might construe files as sets of pairs of a possible world and a sequence in the following way:

$$F = \{\langle w, a_N \rangle : a_N \text{ satisfies } F \text{ in } w\}$$

For example, the file we have been referring to as F1 amounts to the following set of pairs:

$$F_1 = \{\langle w, a_N \rangle : a_1 \text{ is a woman in } w, \text{ and } a_2 \text{ is a dog in } w, \text{ and } a_2 \text{ bit } a_1 \text{ in } w\}$$

For any given world w, such a set of pairs determines a satisfaction set:

$$\text{S\textsc{at}}_w(F) = \{a_N : \langle w, a_N \rangle \in F\}$$

Also, the set of pairs determines a set of worlds which includes just those worlds in which the file is true; call this the "world set" of the file:

$$W(F) = \{w : \text{for some } a_N, \langle w, a_N \rangle \in F\}$$

But we must not identify the file with its world set, because two different files sometimes have identical world sets, for example F_1 above and F_1' below:

$$F_1' = \{\langle w, a_N \rangle : a_1 \text{ is a woman in } w \text{ and was bitten by a dog in } w\}$$

So the relation between Stalnaker's notion of common ground and mine is this: Stalnaker identifies the common ground with what I would call the common ground's or file's world set. From my point of view, he thus restricts himself to one aspect of the common ground, i.e., he characterizes it only incompletely. From his point of view, on the other hand, my files do not just represent common grounds, but add to them some kind of internal structure that plays no role in their evaluation w.r.t. truth and falsity. In order to decide whether a common ground is true or false in a given world, we need to know only what its world set consists of; we need not know with which of the different common grounds with that world set we are dealing. However, if our task is not to evaluate the truth of a common ground, but to predict how the common ground of a context will change as a result of what is uttered in that context, then the additional differentiation becomes relevant: Two common grounds with identical world sets, in which identical utterances occur, may develop into two new common grounds which are not only not identical, but do not even have identical world sets anymore.

This brings us to the next issue that Stalnaker discusses in his paper: How do common grounds change from one context to the next in the course of a conversation, and to what extent is that change determined by the meanings of the utterances that occur in the conversation? Despite the fact that Stalnaker construes common grounds as sets of possible worlds, whereas we do not, his findings carryover perfectly. The following passage contains the essential points (1979, 323):

> Now how does an assertion change the context? There are two ways, the second of which, I will suggest, should be an essential component of the analysis of assertion. I will mention the first just to set it apart from the second: The fact that a speaker is speaking, saying the words he is saying in the way he is saying them, is a fact that is usually accessible to everyone present. Such observed facts can be expected to change the presumed common background knowledge of the speaker and his audience in the same way that any obviously observable change in the

physical surroundings of the conversation will change the presumed common knowledge. If a goat walked into the room, it would normally be presupposed, from that point, that there was a goat in the room. And the fact that this was presupposed might be exploited in the conversation, as when someone asks, "how did *that* thing get in here?" assuming that others will know what he is talking about. In the same way, when I speak, I presuppose that others know I am speaking, even if I do not assume that anyone knew I was going to speak before I did. This fact, too can be exploited in the conversation, as when Daniels says "I am bald," taking for granted that his audience can figure out who is being said to be bald.

I mention this commonplace way that assertions change the context in order to make clear that the context on which an assertion has its ESSENTIAL effect is not defined by what is presupposed before the speaker begins to speak, but will include any information which the speaker assumes his audience can infer from the performance of his speech act.

Once the context is adjusted to accommodate the information that the particular utterance has produced, how does the CONTENT of an assertion alter the context? My suggestion is a very simple one: To make an assertion is to reduce the context set in a particular way, provided that there are no objections from the other participants in the conversation. The particular way in which the context set is reduced is that all of the possible situations incompatible with what is said are eliminated. To put it a slightly different way, the essential effect of an assertion is to change the presuppositions of the participants in the conversation by adding the content of what is asserted to what is presupposed. This effect is avoided only if the assertion is rejected.

What Stalnaker says here, especially in the last paragraph, is, very roughly speaking, tantamount to what I formulated as principle (A) above. But in many ways, Stalnaker is more accurate:

(a) He makes it clear that the context-changing effect he describes is particular to assertions, and he does not commit himself to any predictions about utterances of other illocutionary forces. The same qualification should be added to everything that I have said so far, or will say in the sections to come, about the file-changing effects of "utterances."

(b) He acknowledges that the common ground also gets changed by other factors than the meanings of assertions. I have so far created the impression that my files evolve strictly as a function of what is being said. This is not how I want the notion to be understood in the future. Rather, I assume that the file is influenced by just the kinds of factors that Stalnaker (in the first paragraph of the quote) mentions as influencing the common ground. For example, I take it that the appearance of a goat changes the file, although it is of course not obvious from what I have said so far how exactly it changes it. Note in this

connection that it was completely unrealistic for me to pretend that the first utterance of my sample conversation (1) took its effect on a context whose file was "empty." (An "empty" file is one whose satisfaction set is A^n for every possible world.) Any number of things must have been common knowledge between the participants A and B before A started talking. Nevertheless, I will sometimes make such unrealistic assumptions to keep things simple.

(c) Stalnaker not only restricts himself to the context change effected by assertive utterances, but provides for exceptions even among those: The context change he describes is always contingent upon the willingness of the audience to accept the assertion. If they reject it, the common ground will not be cut down to make it compatible with the assertion. (This is not to say that no change at all will take place.[3]) This further qualification applies mutatis mutandis to my principle (A) as well.

Taking heed of all these points, one might reformulate principle (A) as follows:

(A′) Suppose a sentence S is uttered with assertive force under the reading represented by ϕ. Suppose further that c is the context with respect to which the assertion of S is evaluated, and that F_c is the common ground (=file) at c. Then, unless the assertion is rejected, it will bring out a context c′ which differs from c as follows:

(i) $F_{c'} \subseteq F_c \cap \{\langle w, a_N \rangle: a_N \underline{\text{SAT}}_w \phi\}$

I described c as the context "with respect to which the assertion of S is evaluated," rather than simply as the context "in which the assertion takes place," or as the context "prior to the assertion." This way, I have respected

---

[3] The following dialogue provides a particularly impressive example of file change in response to a rejected assertion:

A: A man jumped out from the crowd and fell in front of the car.
B: He didn't jump, he was pushed.

Examples like this were first discussed by Strawson (1952). As we will see, the acceptability of a pronoun, as in B's answer above, must be taken to indicate that A's utterance has caused a card to get introduced, though A has not succeeded in establishing the intended entry on that card. I am not sure what exactly is going on in such dialogues. Some of the information that A's utterance contains does not make it into the file, but that is only half the story. The other half is that information which was not expressed at all does get into the file. B seems to treat A's utterance as though A had said something like: "That event that you and I both witnessed the other day involved a man jumping out from the crowd and falling in front of a car." I will have to leave this phenomenon unexplained.

Stalnaker's remark (in the second paragraph of the quote) that the essential effect of an assertion is calculated with respect to a common ground that already includes information such as who the speaker is, although such information was not available prior to the utterance. Note also that I cautiously wrote "⊆" instead of "=", so as to avoid the claim that the common-ground-changing effect of an assertion is exhaustively described in terms of its "essential" effect, as Stalnaker calls it.

Another difference between (A′) and my earlier formulation of (A) is that the change is described in terms of files construed as sets of pairs and that the satisfaction relation between sequences and formulas is relativized to a world. With (A), I was concentrating on a fixed world and was considering only the change of the file's satisfaction set w.r.t. that world. The present formulation takes the world parameter into account, so as to facilitate comparison with Stalnaker's paper. The last line of (A′) of course entails, for any given world w, that the satisfaction sets of F_c and $F_{c'}$ differ as follows:

$$\text{SAT}_w(F_{c'}) \subseteq \text{SAT}_w(F_c) \cap \{a_N: a_N \ \underline{\text{SAT}}_w \ \phi\}$$

To what extent does (A′) agree with Stalnaker's account of how an assertion affects the common ground? For Stalnaker, the last line of (A′) would have to read as follows (where C_c and $C_{c'}$ are the context sets of c and c′, respectively):

(ii) $C_{c'} \subseteq C_c \cap \{w: \phi$ is true w.r.t. c and $w\}$

(i) and (ii) implement the same basic idea: that the new common ground is determined essentially by intersecting the old common ground with the set of those "parameters" which make the assertion true. The difference is just in the nature of the relevant parameters: worlds for Stalnaker, pairs of worlds and sequences for me.

The empirical consequences of choosing between (i) and (ii) cannot be evaluated independently of specific assumptions about the semantics of English, especially assumptions as to which English expressions are semantically variables. The remaining sections of this chapter will clarify the issues that are involved here. Right now, I can only say this much:

Suppose we limit our attention to assertions whose truth does not depend on the sequence at all, but merely on the possible world. This is certainly the case when the logical form ϕ of the assertion contains no free variables and we thus have:

If there is some a_N such that $a_N \ \underline{\text{SAT}} \ \phi$, then for all a_N, $a_N \ \underline{\text{SAT}}_w \ \phi$.

In this case, (i) entails (iii):[4]

$$W(F_{c'}) \subseteq W(F_c) \cap \{w: \phi \text{ is true w.r.t. } w\}$$

Since the world set of the file at a context is always the same as the context set at that context, (iii) is equivalent to (ii). (To be precise, I also take it for granted here that the truth of ϕ is not context dependent.) In other words: my theory agrees with Stalnaker's on the question of how assertions of closed sentences contribute to the evolution of the world set, alias the context set.

In the remainder of this chapter, I will continue to work with the simplifications that I used before: I will ignore the world parameter, assume an extensional semantics, and look at a file's actual satisfaction set rather than the entire set of world-sequence pairs. I will also continue to say "utterance" when "assertion" would be accurate and disregard other qualifications that have been mentioned. Hopefully, this will not lead to any confusion.

1.5 *File change potential and satisfaction conditions*

Every sentence, or more precisely: every disambiguated sentence, i.e., every logical form, has a characteristic "file change potential." By the "file change potential" of a logical form ϕ, I mean a function from files to files, namely, that function which assigns to every file F the result-file F′ which is brought about by uttering ϕ in a situation in which F obtains. I will use the notation "... $+\phi$" to denote the file change potential of ϕ. So the equation $F + \phi = F'$ means that F′ is the value of the file change potential of ϕ for the argument F. In other words, it means that an utterance of ϕ changes F into F′.

As I have said before, there is an intimate connection between the meaning of an utterance and its impact on the file. This connection was made precise in section 1.2 above by stating principle (A), which relates the satisfaction condition of a logical form in a systematic way to the file change that this logical form triggers. (A) can also be expressed as follows now:

(A) $\text{SAT}(F{+}\phi) = \text{SAT}(F) \cap \{a_N: a_N \underline{\text{SAT}} \phi\}$

This formulation brings out explicitly that principle (A) states the systematic connection between a logical form's satisfaction condition and its file change potential.

---

[4] Proof: Consider an arbitrary $w \in W(F_c')$. Then, by definition of $W(F_c')$, there is an a_N such that $\langle w, a_N \rangle \in F_c'$. Then, by (i), $\langle w, a_N \rangle \in F_c$ and $a_N \underline{\text{SAT}}_w \phi$. Hence, $w \in W(F_c)$ and ϕ is true w.r.t. w.

In the semantic theory that we have been assuming up to this point, the satisfaction conditions of logical forms play a central role: They are directly assigned to logical forms by the recursive system of interpretation rules. Truth is then defined in terms of them, and file change potentials are also characterized (partially defined) in terms of them, the latter two being secondary concepts relative to the primary concept of satisfaction conditions. But there is no obvious reason why we have to take satisfaction conditions as basic in this way. An alternative that comes to mind is to formulate the rules of interpretation directly in terms of file change potentials.[5] Consider, for example, the simple formula

(2)   She$_1$ hit him$_2$.

As the theory is set up now, we first need to determine the satisfaction condition of (2) before we can know anything about its file change potential. The relevant rule of interpretation, in this case rule (i) (repeated here from Ch. II, section 3.2), yields the satisfaction condition (3).

(i) Let ϕ be an atomic formula, consisting of an n-place predicate ζ and an n-tuple of variables whose indices are $i_1,...,i_n$, respectively. Then, for any sequence $a_N \in A^n$:

$$a_N \text{ \underline{SAT} } \phi \text{ iff } \langle a_{i_1},...,a_{i_n} \rangle \in \text{Ext}(\zeta).$$

(3) a_N <u>SAT</u> (2) iff a_1 hit a_2.

Then we can apply principle (A) to (3) and determine that the file change potential of (2) is as follows:

(4) $\text{SAT}(F + (2)) = \text{SAT}(F) \cap \{a_N: a_1 \text{ hit } a_2\}$.

Now suppose the rules of interpretation were to make direct reference to file change potentials, for instance, instead of (i), we had a rule like (i′):

(i′) Let ϕ be an atomic formula, consisting of an n-place predicate ζ and an n-tuple of variables whose indices are $i_1,...,i_n$, respectively. Then, for any file F:

$$\text{SAT}(F + \phi) = \{a_N \in \text{SAT}(F): \langle a_{i_1},...,a_{i_n} \rangle \in \text{Ext}(\zeta)\}$$

---

[5] There are other authors who have conceived of semantic interpretation as specifying, first and foremost, the change in the context that an expression will bring about. See, e.g., Hamblin (1971), Ballmer (1979), and Smaby (1979).

By applying (i′) to formula (2), we immediately arrive at the characterization of its file change potential that was given in (4). We do not have to take two steps, and we need not invoke principle (A).

It is a routine matter to recast the remaining four interpretation rules of Chapter II, section 3.2, in such a way that they yield directly what would otherwise have resulted from applying principle (A) on the basis of the satisfaction conditions assigned by the original rules. For rule (ii), for instance, we would choose the following substitute:

(ii′) Let ϕ be a cumulative molecular formula with the immediate constituent formulas $\psi^1,...,\psi^n$ (in that order). Then, for any file F:

$$\textsc{Sat}(F + \phi) = \textsc{Sat}((...(F+\psi^1) + \psi^2 ... + \psi^n).$$

(Note that (ii′) reveals the reason why I chose the terminology "cumulative molecular formula": The file change induced by such a formula turns out to be a matter of successively executing the file changes induced by its subformulas, i.e., of "accumulating" in the file the bits of information that each subformula contributes.) I will not spell out the remaining three rules here, as I will devote a separate section (see 4 below) to the file change potentials of operator-headed formulas. It should be plausible, even without a proof at this point, that formulations of the intended sort can be provided.

But what is the point of redesigning semantic interpretation in this way? It apparently is *feasible* to assign file change potentials directly by recursive rules, without going through an intermediate assignment of satisfaction conditions and using principle (A) — but is it also *desirable*?

We are not in a position to answer this question right now, but will return to various aspects of it in the course of the chapter, eventually arriving at a positive answer. For the moment, I have this to say: Ceteris paribus, a theory that employs only the notion of file change potential is preferable to one that uses both satisfaction condition and file change potential. However, we must bear in mind that satisfaction conditions underlie the definition of truth in the current theory, and we will have to come up with an alternative definition of truth before we can really eliminate satisfaction conditions. We will return to this task in section 3 below.

2   Novelty and Familiarity

Traditional grammarians thought a great deal about a question that has more or less disappeared from the semantic literature in which the quantificational analysis of determiners gained such overwhelming popularity: What are the

semantic and pragmatic conditions which determine the choice between a definite and an indefinite noun phrase? The most influential traditional attempt to answer this question was Christophersen's (1939) "familiarity theory of definiteness,"[6] according to which the essential function of definiteness is to signal that the intended referent of an NP is a referent with which the audience is already *familiar* at the current stage of the conversation. An indefinite NP, on the other hand, is used to signal that an as yet *unfamiliar*, i.e., *novel*, referent is being introduced. Starting from this basic idea, Christophersen and later familiarity theorists have attempted to explicate the relevant notion of familiarity and to show how seemingly independent semantic and pragmatic characteristics that correlate with definiteness and indefiniteness can be explained as derivative of the association with familiarity and novelty.

Familiarity theories of definiteness have reappeared in various new versions in the context of modern linguistics and more recently in Artificial Intelligence research.[7] The most notable instance of this is Karttunen's (1968b) paper "What makes definite noun phrases definite?" What makes it particularly notable from my point of view is that Karttunen neither commits himself to the assumption that definite and indefinite NPs are referring expressions, nor does he have to restrict his theory to those uses of definites and indefinites under which they supposedly do refer, leaving the definiteness contrast in nonreferential NPs unaccounted for. Traditional explications of the familiarity/novelty distinction have always sounded as though reference must be taken for granted if the distinction is to apply at all: Whose being familiar or novel, after all, could conceivably be at issue if not the referent's? It is not surprising that interest in the notion of familiarity declined when skepticism about reference as the basic function of NPs grew. Karttunen managed to dissociate familiarity from reference by inventing his "discourse referents" and making them the ones whose familiarity or novelty is at issue in the choice between definites and indefinites. Since discourse referents are not referents, and since an NP may well have a discourse referent

---

[6] See also Jespersen (1949), where the same basic approach to definiteness is taken.

[7] As for early transformational grammar, see, e.g., Baker (1966) and Annear (1968), where "familiarity" was reinterpreted as the formal property of having a coindexed antecedent in deep structure. See also the "location theory" of Hawkins (1978), which replaces "familiarity" by "locatability," but shares the basic spirit of familiarity theories. In artificial intelligence and psychology, familiarity theories have always prevailed over quantificational approaches. See Grosz (1977), Clark (1977), and Clark and Marshall (1981), to mention only a few.

without referring to anything, Karttunen's proposal is not limited to a subset of supposedly referring uses of NPs.

This dissertation may be seen as another revival of the familiarity theory of definiteness. Like Karttunen, I make a major point of dissociating familiarity from reference and of tying the familiar-novel distinction to something more abstract than referents. (In my case, that something will be file cards.) There are many resemblances between Karttunen's theory and mine that they share over and above this point, some of which I will point out as we go along.

What I just said may be a little puzzling in view of the current state of my theory, where the concepts of familiarity and novelty do not play a central role by any means. There is just one reference to the distinction, viz., in the "Novelty Condition," which says that an indefinite must not be coindexed with any NP that precedes it in logical form. This formulation suggests that novelty and familiarity, traditionally felt to be pragmatic concepts, are in the context of my theory construed as purely formal properties of logical forms: an NP is "novel" or "familiar" in a text if it isn't or is coindexed with another NP that precedes it in the text.

There is another respect in which the Novelty Condition deviates from traditional versions of the familiarity theory of definiteness: indefiniteness is a sufficient condition for novelty, but not a necessary one, according to my current analysis. Novel definites do occur, e.g., when a pronoun is used deictically. By contrast, a theory in the proper spirit of Christophersen (1939) would have it that an NP is novel if and only if indefinite, and familiar if and only if definite. It is of course a consequence of my formalistic notion of novelty and familiarity that I cannot strengthen the Novelty Condition to a biconditional: A deictic pronoun is formally "novel" in that it lacks an antecedent in the previous text, while it does presuppose "familiarity" with the referent in the traditional pragmatic sense of the term. Last but not least, my present theory is in an important respect less ambitious than other familiarity theories of definiteness have been. The Novelty Condition is just one of three stipulations in my semantics that applies specifically to indefinites as opposed to definites, the other two being the construal rule of Quantifier Indexing and the assignment of descriptive-content presuppositions. So the correlation between definiteness and familiarity that the Novelty Condition describes is only part of what is needed to predict the contrasting semantic and pragmatic behavior of definites and indefinites. Further, logically independent, stipulations must be made: that only indefinites get bound by the nearest c-commanding quantifier, whereas definites remain free (unless they come to be bound by way of being anaphoric to a bound antecedent), and that only definites

presuppose their descriptive content. I have made no attempt at explaining the latter two differences that correlate with definiteness in terms of the familiarity-difference, and thus cannot claim that familiarity versus novelty is in anyway *the fundamental* distinction that definites versus indefinites serve to express. Therefore, my theory does not seem to be strictly a "familiarity theory of definiteness," at least not if this label is reserved for theories which purport to show that all other systematic contrasts between definites and indefinites are secondary to, if not derivative on, the familiarity-novelty contrast.

In the course of this section, I will revise my theory in several respects, and the result will be much closer in spirit to genuine familiarity theories: I will move away from construing novelty and familiarity in terms of formal properties of logical form. And I will make indefiniteness a necessary, as well as sufficient, condition for novelty. (As for the issue of Quantifier Indexing as an independent part of the semantics of definiteness, see section 4 below.) The idea of a file, as introduced in the previous section, will prove useful in carrying out these revisions.

2.1 Preliminaries

Novelty in the sense of the Novelty Condition is purely a matter of coindexing in logical form. I will therefore speak of "novelty with respect to logical form," and analogously, "familiarity w.r.t. logical form." A different, though not unrelated, sense of novelty and familiarity was alluded to in section 1.1 above, when I formulated the following instruction for file-keeping: "For every indefinite, start a *new* card; for every definite, update a suitable *old* card." In this section, I am going to clarify what this instruction means, and how it relates to novelty versus familiarity w.r.t. logical form.

2.1.1 The domain of a file

As long as we talk about files in the literal sense of the metaphor, it is clear what it means that, e.g., file F_1 contains 2 cards, file F_2 contains 3 cards, and the transition from F_1 to F_2 involves adding a new card. But if we consider files solely in terms of their satisfaction sets, it is impossible to determine how many and which cards they contain. Take the satisfaction sets of F_1 and F_2, which were described as follows in section 1.2 above:

SAT(F_1) = {a_N: a_1 is a woman, a_2 is a dog, and a_2 bit a_1}.

SAT(F_2) = {a_N: a_1 is a woman, a_2 is a dog, a_3 is a paddle, a_2 bit a_1, and a_1 hit a_2 with a_3}.

How could we tell from looking just at these two satisfaction sets that F_2 contains a card number 3 which F_1 lacks? One might think at first that we could tell from the fact that a_3 is mentioned in the specification of $\text{Sat}(F_2)$, but not in the specification of $\text{Sat}(F_1)$. However, this would involve looking at the particular descriptions of the satisfaction sets that we have given here, not just looking at the sets themselves.

Note that the following is also a description of $\text{Sat}(F_1)$:

$\text{Sat}(F_1) = \{a_N: a_1$ is a woman, a_2 is a dog, a_2 bit a_1, and a_3 is identical to itself$\}$.

The impossibility of inferring the number of cards in a file from its satisfaction set alone becomes especially clear if we consider false files. If F_1 and F_2 are both false, for example, then their satisfaction sets equal the empty set and are thus identical. Still, we would like to say that F_2 has more cards than F_1 in it.[8]

In order to be able to say such things, let us henceforth characterize files in terms of two parameters : their satisfaction sets and, in addition, their domains. The domain of F, $\text{Dom}(F)$, is the set that contains every number which is the number of some card in F. For example:

$\text{Dom}(F_1) = \{1,2\}$

$\text{Dom}(F_2) = \{1,2,3\}$

In view of the above discussion, domain and satisfaction sets must be considered separate properties of files that are not interdefinable. I will assume, however, that domain and satisfaction sets constrain each other to some extent, namely as follows:

(B) For every file F, for every $n \notin \text{Dom}(F)$: If a_N and b_N are two sequences that are alike except insofar as $a_n \neq b_n$, then $a_N \in \text{Sat}(F)$ iff $b_N \in \text{Sat}(F)$.

---

[8] I have suggested elsewhere (Heim, forthcoming [Heim (1983a)]) that the satisfaction sets of files be construed as sets of finite, rather than infinite, sequences. This would seem to be a useful thing to do, as it provides one with a straightforward way of reading off the number of cards in a file from its satisfaction set: If $\text{Sat}(F)$ is a set of n-membered sequences, then F is a file of n cards. However, this method is inapplicable with false files, which always have empty satisfaction sets, but should still be distinguished from each other with respect to the number and identity of cards they contain. (If we switch to an intensional semantics, where a file is characterized by its satisfaction sets in all possible worlds, we still have the same problem for necessarily false files.) So we would still need the domain of a file as a separate parameter, even if we worked with finite sequences. I have therefore kept the sequences infinite in this dissertation, which makes for simpler formulations of most interpretation rules.

The intuitive impact of this is that a file cannot impose conditions on the n-th member of its satisfying sequences unless it contains an n-th card. Applied to real files, this is a prohibition against crossreferences to non-existing cards, for instance, against having a card with an entry "was bitten by 2" in a file which does not also contain a card number 2.

Having introduced the notion of a file's domain, we will want to describe the file change potentials of logical forms along two dimensions: in terms of the way the satisfaction set changes, and also in terms of the way the domain changes. Take for instance the interpretation rule that describes the file change potential of an atomic formula. Our current formulation of this rule (i$'$) in section 1.5 specifies only the change in satisfaction sets. Let us therefore complete it in the following way:

(i$''$) Let ϕ be an atomic formula, consisting of an n-place predicate ζ and an n-tuple of variables whose indices are $i_1,...,i_n$, respectively. Then, for any file F,

$$\textsc{Sat}(F + \phi) = \{a_N \in \textsc{Sat}(F): \langle a_{i_1},...,a_{i_n} \rangle \in \text{Ext}(\zeta)\};$$
$$\textsc{Dom}(F + \phi) = \textsc{Dom}(F) \cup \{i_1,...,i_n\}.$$

This formulation expresses the natural idea that the new file must always include cards for all the things that are mentioned in the piece of utterance being evaluated, as well as any cards that may have already been in it before. Rule (ii$'$) can be completed in an equally straightforward way.

(ii$''$) Let ϕ be a cumulative molecular formula with the immediate constituent formulas $\psi^1,...,\psi^n$ (in that order). Then, for any file F,

$$\textsc{Sat}(F + \phi) = \textsc{Sat}((...(F + \psi^1) + \psi^2) ... + \psi^n)$$
$$\textsc{Dom}(F + \phi) = \textsc{Dom}((...(F + \psi^1) + \psi^2) ... + \psi^n)$$

It is easily seen that our initial metaphorical description of file change, as we imagined it to proceed in response to the text about the woman and the dog, conforms to these rules. (Recall that we took the text to be uttered with the reading represented by the logical form (1$'$) displayed in section 1.2 above.)

2.2   *How definiteness affects file change in the current theory*

Given the notion of a file's domain, we can straightforwardly define what it is to "add a card" to the file:

The change from F to F$'$ involves the addition of a card number i iff i $\notin$ $\textsc{Dom}$(F) and i $\in$ $\textsc{Dom}$(F$'$).

We now want to determine what our theory in its current state has to say about the correlation between indefiniteness of a noun phrase uttered and addition of a new card in the resulting file change.

The formulations of rules (i″) and (ii″) that were just given lead to the following predictions: If (a) Dom(F) does not yet contain n, and (b) n is the referential index of an NP in ϕ then the change from F to F + ϕ will involve addition of a new card number n. (Actually, we can make this prediction only for formulas ϕ that are completely analyzable by rules (i″) and (ii″), i.e., do not involve any operators. Formulas with operators have to be disregarded here, since we do not have the pertinent file change rules available yet.)

Now this prediction appears to imply that the conditions under which a new card will get added have nothing to do with definiteness; all that seems to matter is that an NP carry a new referential index, be it a definite or an indefinite NP. However, closer inspection shows that definiteness does play a role though only indirectly. Suppose the logical form ϕ whose utterance we are considering contains NP_n, where NP_n is indefinite. Because it is indefinite, NP_n is subject to the Novelty Condition, which means that n does not occur as the index of any NP that precedes NP_n in the text of which ϕ is part. For all we know so far, we may conclude from this that the file with respect to which the utterance of ϕ is evaluated does not yet contain a card number n. After all, where could such a card have come from if not from the utterance of an earlier NP indexed n? But given that the file does not yet contain a card numbered n, it follows from what we have just determined that the utterance of ϕ will bring it about that a new card, numbered n, gets added.

The reasoning we have just gone through establishes a connection between indefiniteness and adding a new card, and the Novelty Condition is responsible for this connection: Because an indefinite is necessarily novel w.r.t. the logical form of the text it occurs in, its index is not yet in the file prior to its utterance. Therefore, a new card with that index will be added.

Let us define "novelty w.r.t. a file": An occurrence of an NP in a logical form whose index is i is *novel w.r.t. a file F* if i ∉ Dom(F), and *familiar w.r.t. F* if i ∈ Dom(F). We can now sum up the relationship between definiteness, novelty w.r.t. logical form, novelty w.r.t. a file, and adding a new card in the following set of statements:

(i) Indefiniteness is a sufficient condition for novelty w.r.t. logical form. I.e.: If NP_i is an occurrence of an indefinite in the logical form ϕ of a text T, then it must be novel w.r.t. ϕ.

(ii) Novelty w.r.t. logical form is a sufficient and necessary condition for novelty w.r.t. the file. I.e.: If NP_i occurs in the utterance of a sentence S,

where S is uttered as part of the utterance of a text T under the reading ϕ, and F is the file which obtains at the beginning of the utterance of S, then: NP_i is novel w.r.t. ϕ iff NP_i is novel w.r.t. F.

(iii) A new card is added for each NP that is novel w.r.t. the file. I.e.: If NP_i occurs in the utterance of S under the reading ϕ, and F is the file that obtains at the beginning of the utterance of S, then: card number i gets newly added to F if NP_i is novel w.r.t. F.

Statement (i) is simply the Novelty Condition; it is stipulated. The other two are inferred, partly from the rules of interpretation, which characterize the file change potentials of logical forms, and partly from additional assumptions about files, such as the assumption that cards do not get introduced other than by utterances of NPs.

Taken together, (i), (ii), and (iii) predict that file change proceeds as if it were carried out according to the following instruction: "For each indefinite, add a new card." This instruction is related to, but not at all equivalent to, the one I mentioned earlier, which read: "For every indefinite, start a new card; for every definite, update an old card." Contrary to the latter instruction, it does not follow from (i)-(iii) that definites cannot introduce new cards as well; rather, the prediction is that a definite which is novel w.r.t. the file will affect the file in exactly the same way as an indefinite.

So much for the current theory and what *it* predicts about the correlation between definite versus indefinite and familiar versus novel. It is time for some revisions.

2.3   *Deixis and familiarity with respect to the file*

The deictic use and the anaphoric use have at least this in common: Both are among the possible uses of definite NPs, but neither is possible with an indefinite. One possible explanation for this might be that the pragmatics of deixis and anaphora are intrinsically similar, and definiteness correlates with the property they share.

What might that shared property be? In some sense, both deictic reference and anaphoric reference presuppose that the referent be already "familiar" to the audience: In the case of deictic reference, it has attained familiarity by being pointed at, being perceptually prominent, or being otherwise salient. In the case of anaphoric reference, it has been familiarized by previous mention.

The file, as we have been imagining it until now, keeps track of what is familiar by previous mention at any particular point in the conversation: If something has been mentioned before, there will always be a card for it in

the file, a card with the number of the NP which was used for the previous mention. But does the file also reflect what is familiar by contextual salience? So far we have not assumed it does, but let us make the assumption now. Let us assume that anything that is familiar for either reason is represented by a card in the file.[9]

An obvious implication of this assumption is that files must be able to change, and in particular, must be able to have new cards added, without anything being uttered. For instance, if halfway through a conversation between A and B a dog comes running up to them and draws their attention, then that event presumably makes the file increase by a new card. Suppose the file right before the dog's appearance was F, and 7 is a number such that $7 \notin \text{Dom}(F)$. Then the new file, F', might be as follows: $\text{Dom}(F') = \text{Dom}(F) \cup \{7\}$, and for any a_N: $a_N \in \text{Sat}(F')$ iff both $a_N \in \text{Sat}(F)$ and a_7 = the dog that is running up to A and B. So the dog is henceforth represented in the file by card 7. (The number 7 is of course an arbitrary choice. Any other number that is not yet used in F could have been chosen equally well.[10])

Suppose now that, after F has changed into F', A says: "It is going to bite." This sentence has a reading under which "it" carries the referential index 7. Suppose A's utterance gets interpreted under this reading. Then F' will be turned into a new file, F'', which looks like this: $\text{Sat}(F'') = \{a_N: a_N \in \text{Sat}(F')$ and a_7 is going to bite$\}$. In other words: the change from F' to F'' consists of updating card 7, the card that was introduced for the salient dog. The new file, F'', requires for its truth that that same dog is in fact going to bite. In this sense, A's utterance of "It is going to bite" has been read as a claim about

---

[9] The same assumption, except in a different terminology ("discourse referents" instead of "file cards") appears in (Karttunen, 1968b, 16):

> Anything in the immediate environment of the speaker and hearer towards which their attention is directed becomes a discourse referent whether it has been explicitly mentioned or not.

[10] How exactly does the newly added card number 7 represent the salient dog? What entry does that card carry? By describing merely the satisfaction set of F', as I did in the text, I am avoiding this question. Lewis (1979) remarks that an object that rises to salience always does so under a particular "guise." It is natural to assume that a file card describing a salient individual will describe it in terms of the guise under which it is salient. In the case of perceptual salience, as in our example, the guise is a bunch of properties that the discourse participants perceive about the individual, i.e., properties concerning its vital appearance and its location relative to the discourse participants. For many purposes, it will be irrelevant under which guise a salient individual is represented in the file, but see section 2.3 below.

the dog, or, put differently, the pronoun "it" in A's utterance has been read as a deictic pronoun, whose referent is the contextually salient dog.

If we think of deictic reference as mediated by the file in the manner just illustrated, some of our assumptions about the relation between definiteness, novelty w.r.t. the file, and novelty w.r.t. logical form, must change. In particular, we now observe:

(ii′) Novelty w.r.t. logical form is a necessary, but not a sufficient condition for novelty w.r.t. the file.

In other words: It is possible for an NP to be novel w.r.t. the logical form of the text in which it is uttered, yet to be familiar w.r.t. the file that obtains when it is uttered. This is the case with the pronoun "it," read as "it$_7$," that A utters in our example above. 7 is not the index of any NP that has been uttered between A and B before the dog's appearance. (The situation would be different if A and B had already been talking about that dog and, upon its appearance, had recognized it as the one under discussion. In that case, the file would not have been enriched by a new card, but rather by another entry on some old card, maybe: "is running up to us.") So when "it$_7$" is uttered, it is novel w.r.t. logical form, but familiar w.r.t. the file.

As I have been saying. there is no complete correlation between definiteness and familiarity-w.r.t.-logical-form; the latter is merely sufficient for the former. But we may now stipulate that there is a perfect correlation between definiteness and familiarity-w.r.t.-the-file: If deictic definites are to be treated as illustrated, they will typically be familiar w.r.t. the file that prevails when they are uttered. And for anaphoric definites the same is true, simply because anaphoricity means familiarity-w.r.t.-logical-form, which entails familiarity-w.r.t.-the-file. (Of course it is by no means a foregone conclusion that all definites are used either deictically or anaphorically, and we had better look at the remaining uses before generalizing about all definites. But that is a major project. I will say a little more on this topic in section 5 below.) On the other hand, indefinites can be neither deictic nor anaphoric and therefore must be novel both w.r.t. logical form and the file.

Let us stipulate this correlation between definiteness and familiarity-w.r.t.-the-file in the form of a new constraint, the "Novelty-Familiarity-Condition":

> Suppose something is uttered under the reading represented by ϕ, and the file prior to the utterance is F. Then for every NP$_i$ in ϕ, it must be the case that: i ∈ Dom(F) if NP$_i$ is definite, and i ∉ Dom(F) if NP$_i$ is indefinite. Otherwise, the utterance is not felicitous under this reading.

This is not a constraint on the wellformedness of logical forms, but what might be called a "felicity condition": it imposes certain limitations on which readings an utterance admits of w.r.t. a given file.

The old Novelty Condition falls out from the Novelty-Familiarity-Condition and therefore becomes superfluous as a separate stipulation. It falls out because if an indefinite is novel w.r.t. the file, as it has to be to satisfy the Novelty-Familiarity-Condition, then it is a fortiori novel w.r.t. logical form.[11]

The treatment of deictic NPs that I have just adopted is intended to replace the more primitive treatment that I assumed throughout Chapter II and that I implemented in section 3.3 of Chapter II in the definitions of felicity and truth. There I spoke of the context as furnishing values for the free variables in an utterance. Under the present view, it is still correct in some sense that the context furnishes the values for free variables; but it does so indirectly: the context affects the file, and the file — via the Novelty-Familiarity-Condition — imposes felicity conditions on the utterance. There is no longer any need for a felicity condition that directly relates utterance and context, stipulating that the context must determine a referent for each variable free in the text uttered. For suppose i is the index of such a variable. Then i must be the index of a definite NP in the text uttered. (Otherwise, if i occurred only on indefinite and/or quantifying NPs in the text uttered, the rules of construal — in particular, Quantifier Indexing would have ensured that it corresponds to a *bound* variable.) Therefore, by the Novelty-Familiarity-Condition, i must already be in the domain of the file w.r.t. which the text is evaluated. This can only be the case if card i has been established to represent some contextually salient individual.

So the new Novelty-Familiarity-Condition not only replaces the old Novelty Condition, it also replaces the felicity condition of Chapter II on utterances of texts with free variables. But as we drop that felicity condition, how are we to define truth now? Our original definition, (Chapter II, section 3.3) explicitly refers to the contextually supplied values for the free variables, and it is not obvious how we should rewrite it if contexts are no longer seen as supplying values for variables directly. We will reconsider the notion of truth in section 3 below.

What we have accomplished in this section might be described thus: We have found the common denominator for deixis and anaphora: both presuppose familiarity w.r.t. the file. Note two things about this accomplishment: (a) It captures a significant linguistic generalization, viz., that the same

---

[11] Actually, the Novelty-Familiarity-Condition fails to rule out certain vacuous coindexings that would have been ruled out by the Novelty Condition. But this is an irrelevant difference.

linguistic devices are employed for deixis and for anaphora, in English as well as in other natural languages.[12] (b) We could not have achieved this without the concept of a file (or some similarly abstract concept). In particular, the theory of Chapter II did not permit us to give a simple sufficient and necessary condition for definiteness. The only semantic property that all anaphoric and deictic definites had in common, according to that earlier theory, was their being variables — but so were indefinites.

Moreover, the Novelty-Familiarity-Condition has this advantage over the felicity criterion of Chapter II that it replaces: The latter did not sensibly apply to smaller units of utterance than complete texts. (I pointed this out at the end of section 3 of Ch. II above.) The problem was that the old felicity condition always required there to be contextually supplied referents for all variables free in the logical form of the unit of utterance under consideration. But only for complete texts is it actually appropriate to identify free variables with deictically used NPs; in smaller units, e.g., sentences within multisentential texts, there may well be free variables that are bound in the larger environment and therefore not deictic. The Novelty-Familiarity-Condition, on the other hand, applies naturally and with intuitive results to individual sentences, thus reflecting our judgments as to how an already begun discourse may be felicitously continued.

2.4   How to interpret constraints on coindexing

I have adopted the so-called "Noncoreference" and "Disjoint Reference" Rules (NCR and DRR) without worrying about certain puzzles concerning the nature of these rules. The puzzles I have in mind have often been re-marked upon in the literature, and some people have even been so impressed by them as to reject rules like NCR and DRR as incoherent.

First of all, there seem to be blatant counterexamples to NCR and DRR. In statements like (1) and (2), the italicized occurrences of "he" and "John" not only *can* be understood as coreferential — they *must* be so understood if the statements are to have any chance of being true.

  (1)   *He* is *John*.
  (2)   He must be John, because *he* put on *John's* coat.

If NCR and DRR are understood to be rules which, under certain structural conditions, prohibit coreference, then (1) and (2) are undeniable counterex-amples. The same is true if NCR and DRR are rules which require that the

---

[12] Karttunen (op. cit.) makes this point as well.

speaker purport to refer to distinct individuals , or which prohibit that the speaker purports to refer to the same individual. (Interpretations of NCR and DRR along the lines of the latter two formulations have occasionally been proposed in response to counterexamples of a similar sort. These reinterpretations help to avoid some counterexamples, but they do not suffice to deal with (1) or (2), as far as I can see.)

What then is the correct interpretation of NCR and DRR? I think it is the one proposed by Postal (1970), who sees them as prohibitions against presupposed coreference. From the point of view of Postal's proposal, (1) and (2) are no longer counterexamples. Whenever (1) and (2) are uttered, the identity of the referent of "he" with the referent of "John" is ipso facto under debate, and what is under debate is not presupposed. So these examples are consistent with NCR and DRR when the latter are taken to rule out presupposed coreference of "he" and "John."

But it would be premature to consider the problem settled and to move on to other matters. For even if Postal's proposal is correct, it remains to be seen how we can make our theory yield the corresponding predictions. Recall that we have built NCR and DRR into the grammar as constraints on *coindexing*, which related only very indirectly to coreference, or presupposed coreference, or anything of that sort.

Let us first consider this matter in the context of the treatment of deictic reference that we have just abandoned, the treatment of Chapter II. There we said that for an utterance of a logical form ϕ to be felicitous in a context C, C had to furnish a unique individual, $a_{C,i}$, for each i that is the index of a variable free in ϕ. It is clear from this that if ϕ contains two free NPs whose indices are the same, then they must have the same contextually furnished referent. However, nothing seems to ensure that two free NPs with distinct indices will receive distinct referents. Coindexing is only a sufficient, but not a necessary, condition for coreference. This result is welcome insofar as it permits coreference in examples like (1) and (2), even though coindexing is ruled out by NCR and DRR. But it is unwelcome insofar as it also permits coreference in all those cases which intuitively require noncoreference and which gave rise to the invention of NCR and DRR in the first place.

In order to correct this deficiency, we would like to somehow incorporate Postal's notion of "presupposed" identity into the felicity condition that governs the contextual specification of deictic referents. But this is more easily said than done. It would be quite naive to think that we just need to add a clause like (ii) below to the felicity condition:

ϕ is felicitous w.r.t. C only if:

(i) C furnishes a unique individual, $a_{C,i}$, for each index i of a variable free in ϕ; and

(ii) for any i and j: if it is presupposed in C that $a_{C,i} = a_{C,j}$, then i = j.

For what does it mean to presuppose that $a_{C,i} = a_{C,j}$? Either $a_{C,i}$ is in fact $a_{C,j}$, and then the proposition that $a_{C,i} = a_{C,j}$ is the necessarily true proposition. Or else, $a_{C,i}$ is distinct from $a_{C,j}$, in which case the proposition that $a_{C,i} = a_{C,j}$ is necessarily false. But intuitively, Postal's prohibition against presupposed coreference is a prohibition against presupposing a contingent proposition. We have run into a famous philosophical problem here: How can beliefs about identity be contingent?

The problem can be avoided (at least for the purposes at hand) by thinking of contexts as furnishing not simply individuals, but "individuals in guises." For instance, the context C of an utterance of "He$_1$ likes him$_2$" does not just assign Bill to index 1 and Fido to index 2, but it assigns Bill as the one that the speaker is pointing to to 1, and Fido as the one that right now is noisily entering the room to 2. Presupposed coreference in the sense intended by Postal can then be understood as involving a presupposition about the guises: If it is presupposed, for instance, that the speaker is pointing to the same thing that is just entering the room, then these two guises cannot felicitously be associated with distinct referential indices. So if "He$_1$ likes him$_2$" is to be felicitous w.r.t. the context C just described, then there must not be such a presupposition in C.[13]

"Guises" can be naturally thought of as entries on file cards. So if it is correct, as the above discussion would seem to indicate, that we cannot capture the correct relation between coindexing and coreference unless we acknowledge that contextually furnished referents come in guises, then this adds support to the view of deictic reference that we adopted in the preceding section, where we said that the context supplies a file card rather than an object.

---

[13] I have chosen an example with two pronouns here, rather than one with a name and a pronoun as in the initial examples (1) and (2). Proper names (and other non-pronominal definites) raise various additional questions that I prefer to avoid here. I suggested in Chapter II, 6.2 above, rather simple-mindedly, that a proper name [John$_i$] has as its descriptive content the property of being John, which is a property that an individual will either necessarily have or necessarily lack. When this view is spelled out in the present framework, it will lead to the undesirable prediction that there is no felicitous and contingent reading for a sentence like "John is Bill." Presumably, the descriptive contents of proper-name-NPs are not like that. Perhaps [John$_i$] has the property of being called "John" as its descriptive content. This dissertation makes no serious attempt to resolve issues like this one.

Let us now consider the relation between coindexing and coreference in the light of this newly adopted view. First we have to clarify what "coreference" amounts to under this view. Since NPs are now taken to be associated with their referents (if any) only indirectly, via file cards, it is natural to define the locution "referent of an NP" as derivative on the following notion of "referent of a file card,"

Card i in F refers to x iff, for all $a_N \in$ Sat(F), $a_i = x$.

(Note that a file card need not always have a referent according to this definition.) Now suppose NP_i and NP_j are two NPs that occur in a logical form that is evaluated w.r.t. a file F. Then we say that NP_i and NP_j corefer w.r.t. F iff cards i and j in F refer to the same x. It follows that NP_i and NP_j will always corefer when i = j, unless neither NP refers at all. It is not excluded, however, that NP_i and NP_j corefer w.r.t. a file F, even though i ≠ j. For reasons indicated above, we do not want to exclude this case entirely, but we would like to exclude what we have been calling "presupposed coreference."

If we want to distinguish presupposed coreference from mere actual coreference, it will not suffice to look exclusively at the satisfaction set that a file has in the actual world; we also have to consider its satisfaction sets in alternative worlds (cf. section 1.4 above). Suppose a file F entails that two of its cards, say i and j, must correspond to the same individual in any satisfying sequence in any world, i.e.:

(3)   For all w, all a_N: if $a_N \in$ Sat$_w$(F), then $a_i = a_j$.

If a file f with this property obtains in a situation where a logical form with two NPs, NP_i and NP_j, is uttered, then NP_i and NP_j will not only corefer w.r.t. F, but they will do so as a necessary consequence of the information that F already contains. This is the case of presupposed coreference that we want to rule out. To rule it out, we will not make specific provisions to that effect in formulating the felicity conditions for utterances with free NPs. Rather, we will assume that there simply aren't any files with property (3).

More precisely, we will assume that whenever a file with that property arises in the course of a conversation — as it may, e.g., when an identity statement has been asserted and accepted — then it will undergo spontaneous reorganization: cards i and j will be consolidated into a single card which contains all the entries of each of them. Once this reorganization has taken place, only coindexed NPs will be able to refer to the joint referent of the previous cards i and j.

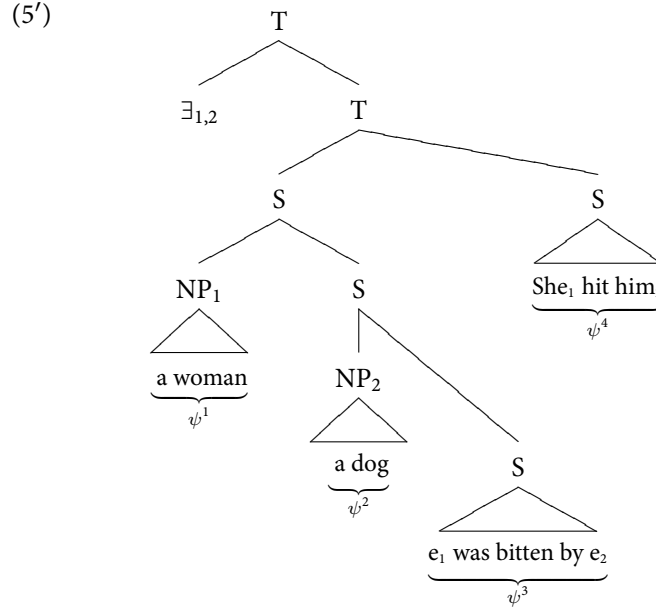*2.5   The Novelty-Familiarity-Condition and the Projection Problem*

Felicity conditions of any sort give rise to a problem analogous to the famous projection problem for presuppositions. I noted this in Chapter II, section 6.3, in connection with my proposal that descriptive definites presuppose their descriptive content. This time, it is the Novelty-Familiarity-Condition that needs to be examined with regard to complex logical forms and the way in which their felicity conditions relate to the felicity conditions of their parts. That the issue is a non-trivial one can be seen from a comparison of the felicity conditions of (4) on the one hand, and (5), a text that contains (4) as a part of it, on the other.

(4)   She$_1$ hit him$_2$.
(5)   A woman$_1$ was bitten by a dog$_2$. She$_1$ hit him$_2$.

Whereas (4) imposes on the file F with respect to which it is evaluated, the requirement that both 1 and 2 be elements of Dom(F), (5) requires no such thing about the file with respect to which it, as a whole, is evaluated; in fact, it requires the opposite, viz., that neither 1 nor 2 be already in the domain of the initial file. This shows that we cannot let the Novelty-Familiarity-Condition apply to molecular formulas in the same way that we apply it to atomic ones. If we did, we would seem to predict that (5) cannot be felicitous w.r.t. any file, because such a file would have to meet contradictory conditions: Its domain could not include either 1 or 2, because 1 and 2 occur as indices of indefinites in (5), but it would also have to include 1 and 2, because 1 and 2 occur as indices of definites in (5). How, then, do complex logical forms inherit the felicity conditions that the Novelty-Familiarity-Condition assigns to their parts?

  The answer I will suggest here is simple: A complex logical form ϕ is felicitous w.r.t. a given file F just in case every elementary step in the construction of F + ϕ from F can be carried out without violating any felicity conditions. By "elementary steps," I mean all the evaluations of atomic formulas that occur along the way from F to F + ϕ. If each of those atomic formulas is felicitous w.r.t. the file that obtains at the moment of its evaluation, then ϕ as a whole counts as felicitous with respect to F.

  For example, the evaluation of (5) contains four elementary steps, corresponding to the four atomic formulas $\psi^1,...,\psi^4$ that its logical form (5′) contains.

(5′)

```
                              T
                        ╱           ╲
                   ∃₁,₂              T
                              ╱              ╲
                            S                  S
                       ╱        ╲            ╱  │  ╲
                    NP₁          S       ⟨She₁ hit him₂⟩
                    ╱╲        ╱      ╲         ψ⁴
              ⟨a woman⟩    NP₂        
                  ψ¹      ╱╲     
                      ⟨a dog⟩      S
                        ψ²      ╱  │  ╲
                           ⟨e₁ was bitten by e₂⟩
                                  ψ³
```

Ignore, for the moment, the existential quantifier, i.e., assume that it has no effect on file change. (We will return to its status in section 3.) Then, by successive applications of rule (ii″) (as formulated in 2.1.1 above), we know that in order to calculate $F + (5′)$ from a given file F, we must calculate the following four files along the way:

$F + \psi^1$

$(F + \psi^1) + \psi^2$

$((F + \psi^1) + \psi^2) + \psi^3$

$(((F + \psi^1) + \psi^2) + \psi^3) + \psi^4 = F + (5′)$

At each of these four steps, the Novelty-Familiarity-Condition has to be met before the calculation can be carried out:

The first step requires that ψ^1 be felicitous w.r.t. F, which means — since ψ^1 contains (here: exhaustively contains) an indefinite indexed 1 — that we must have $1 \notin \text{Dom}(F)$.

The second step requires, analogously, that $2 \notin \text{Dom}(F + \psi^1)$. Since we know (by rule (ii″)) that $\text{Dom}(F + \psi^1) = \text{Dom}(F) \cup \{1\}$, the requirement that $2 \notin \text{Dom}(F + \psi^1)$ can be met only if $2 \notin \text{Dom}(F)$.

For the third step, it is required that ψ^3 be felicitous w.r.t. $(F + \psi^1) + \psi^2$. What this requirement amounts to depends on whether we count the empty

NPs "e_1" and "e_2" as definite, indefinite, or neither. Suppose we count them as definite. Then the Novelty-Familiarity-Condition will apply in such a way as to require that both $1 \in \mathrm{Dom}((F + \psi^1) + \psi^2)$ and $2 \in \mathrm{Dom}(F + \psi^1) + \psi^2)$. These requirements will certainly be met, since we know (again by rule (ii″)) that $\mathrm{Dom}((F + \psi^1) + \psi^2) = \mathrm{Dom}(F + \psi^1) \cup \{2\} = \mathrm{Dom}(F) \cup \{1\} \cup \{2\}$.

Finally, for the fourth step, we must make sure that ψ^4 is felicitous w.r.t. $((F + \psi^1) + \psi^2) + \psi^3$, which it will be if both $1 \in \mathrm{Dom}(F + \psi^1) + \psi^2) + \psi^3)$ and $2 \in \mathrm{Dom}(((F + \psi^1) + \psi^2) + \psi^3)$, again a requirement that cannot fail to be met.

All in all, we have found that for all of the four steps to go through without violating the Novelty-Familiarity-Condition, the starting file F must meet two requirements: $1 \notin \mathrm{Dom}(F)$ and $2 \notin \mathrm{Dom}(F)$. These two requirements can hence be considered to be the felicity conditions that (5′), as a whole, imposes on F.

I have just given an example of how the felicity conditions of a complex utterance can be inferred from the felicity conditions of its atomic parts. Note that this inference could be made without adding any new machinery to the theory. We just added the natural assumption that a complex expression is felicitous if all the elementary steps in its evaluation conform to felicity requirements. Otherwise, we had to rely only on assumptions that were already at our disposal: we needed to know how the Novelty-Familiarity-Condition affects the felicity of atomic formulas, and we had to apply the rules that assign logical forms their file change potentials. Suppose now we were to be equally successful in predicting the felicity conditions of other types of complex formulas on the basis of already available assumptions, and this success were to generalize to other kinds of felicity conditions (not only those related to definiteness). Then we could claim to have a solution to the projection problem, and in fact an optimal solution insofar as it falls out from the rest of the theory at no extra cost.

Of course, we cannot defend any such claims without having dealt with more types of examples, especially examples of presupposition projection in quantified structures, which have been primarily responsible for making the projection problem a non-trivial matter. A more extensive discussion of these matters, which would exceed the scope of this dissertation, will be undertaken in a separate study (Heim, in preparation) [Heim (1983b)].

Continuing with the optimistic assumption that our treatment of example (5) was indeed representative of a general solution to the projection problem, let me draw a conclusion with respect to an issue that was first raised in section 1.5 above. There we faced a choice between two conceptual organizations of the theory, which appeared to be equally feasible: We could either conceive of the recursive rules of interpretation as assigning

satisfaction conditions to logical forms and then characterize other semantic concepts (in particular, the file change potential of a logical form) in terms of those satisfaction conditions. Or else, we could assume that the recursive interpretation rules linked logical forms directly to their *file change potentials*, which might then provide the basis for defining other semantic concepts in terms of them.

In the present section, I tacitly took the theory to be set up in the latter way, and the point I want to make here is that I *had* to make this assumption. For suppose I had been working with a recursive assignment of satisfaction conditions and a principle like (A) (cf. section 1 above) by which to determine the file change potential from the satisfaction condition. This would have given me no way of viewing the file change induced by the complex formula (5) as a sequence of other, more elementary, file changes. File change is always a one-step affair from the perspective of a theory organized in that way. (Only the calculation of the satisfaction condition that is needed to determine the file change may be a matter of many steps.) But without the assumption that (5) induces a sequence of elementary file changes, each of which could be individually subjected to the felicity requirements of the Novelty-Familiarity-Condition, I could not have reasoned as I did when I deduced the felicity conditions of (5) from the felicity conditions of its atomic parts.

In other words, I had to make crucial reference to rules like (ii″) which break down the file changes induced by complex formulas into smaller steps. Without such rules to refer to, the projection problem could not have been taken care of so easily. This gives us a reason to prefer file change potentials over satisfaction conditions when it comes to choosing the semantic concept that figures in the recursive interpretive component. To be sure, it is still conceivable that this preference will be overridden by other considerations that might support the conceptual priority of satisfaction conditions. This leads us to the next section, where we will reexamine the role of the notion of satisfaction as a prerequisite for defining truth and falsity.

3   Truth

We are halfway into revising the theory, and there are various loose ends that need tying up: For one thing, we need to give a new definition for the truth of a formula (or of a text of English under a reading), since the old definition of Chapter II, section 3.3, is no longer applicable: it relies on a treatment of the felicity conditions for deictic pronouns that we have abandoned, and it presumes an assignment of satisfaction conditions to formulas that we are

considering to eliminate. Besides, we have introduced a concept of truth that applies to files, without clarifying whether and how the truth of a file and the truth of the utterances that have contributed to the construction of that file are interrelated. In the process of dealing with these issue, we will also throw some light on the construal rule of Existential Closure.

3.1 Is Existential Closure dispensable?

We have so far kept the system of construal rules precisely as it was developed in Chapter II. In particular, we have been assuming that the logical form of a text is always headed by an existential quantifier with scope all the way across the text, because that is what the obligatory operation of Existential Closure, or more specifically: the second subrule of Existential Closure, brings about. So for example, we took it for granted that the logical form of our sample discourse about the woman and the dog looked as represented under (1′) above (section 1.2), i.e., it was of the form (1a).

(1a)
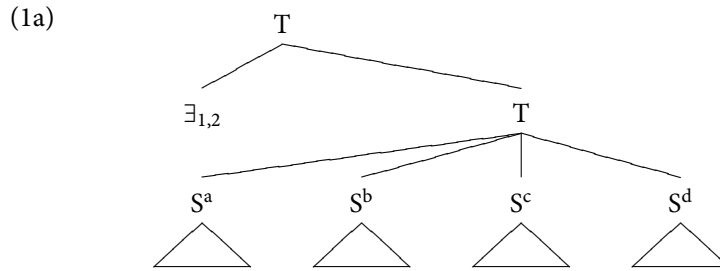
However, when we talked about the way in which the utterance of this text causes the file to change, the initial existential quantifier seemed to play no role at all. Whereas each of the four S-constituents made its impact on the file, changing F_0 into F_1, F_1 into F_2, etc. (see the file-histories in sections 1.2, 2.1.1, or 2.4), the node "$\exists_{1,2,3}$" had no visible influence on the evolution of the file. For all we can tell, it might as well have been missing altogether. Apparently, the file would have evolved identically, had the logical form of the uttered text been (1b).

(1b)
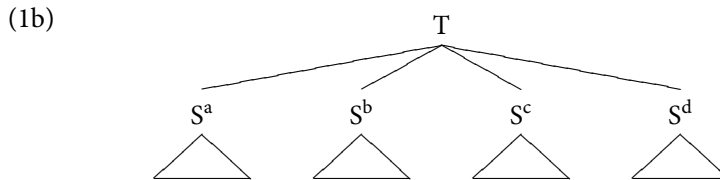
Now (1b) is not a permissible logical form for any text under the assumptions of Chapter II, but suppose we revise those assumptions: Let us simply eliminate the second subrule of Existential Closure. (The second subrule is the one that appends existential quantifiers to texts, as opposed to the first, which applies to the nuclear scopes of quantifiers; we are not yet abolishing the latter here.) Then the logical form of text (1) looks indeed like (1b), and similarly for other texts.

But just because the topmost existential quantifier has not been attributed any file-changing effect does not mean it is altogether superfluous, or does it? We had a good reason (and one independent of anything to do with files) to stipulate it in the first place: We wanted to account for the existential readings of indefinites, and given the variable-analysis for indefinites, that meant we had to supply an existential quantifier to bind them. It seemed obvious that without Existential Closure we would predict hopelessly wrong truth conditions, and would in fact treat indefinites as though they were equivalent to deictic definites, both being represented by variables free in the text. Is that no longer a valid reason for stipulating Existential Closure? It turns out it is not, but it will take us a while to see this. We will have to show that once we have adjusted the definition of truth of a text to the present theory it will predict adequate truth conditions for texts with free indefinites.

It may seem that by permitting indefinites to remain free in their text, as the abolition of Existential Closure of course implies, we can no longer distinguish them from deictic NPs, which are likewise represented by variables free in their text. But keep in mind that the two are not indistinguishable from the point of view of the Novelty-Familiarity-Condition: deictic NPs, being definite, have to be familiar w.r.t. the file, whereas indefinites, even free ones, have to be novel w.r.t. the file. So the texts containing them will, at the very least, differ in their felicity conditions. Now, to be sure, that is not enough: we want to predict a difference in truth conditions as well. How this is accomplished, you will see shortly.

3.2 *A truth criterion for utterances, based on truth values of files*

Let us recall what it means for a file to be true (see section 1.2 above):

F is true iff there is at least one sequence a_N such that $a_N \in \text{SAT}(F)$. Now given that we think of files as recording what has been said in a discourse, we ought to assume that saying something false produces a false file, and saying something true produces (ceteris paribus) a true file. Indeed, we might try to use this relationship between the truth of the utterance and the truth of the

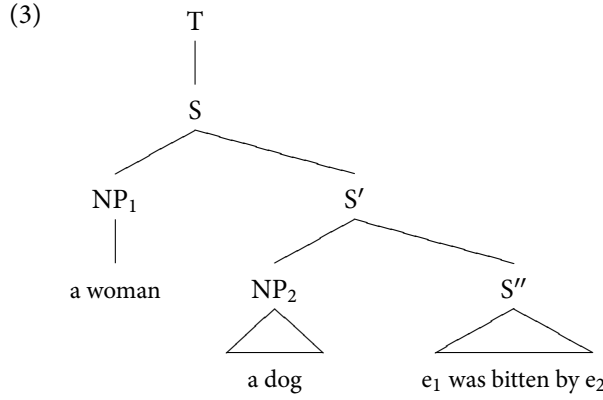resulting file to define the former in terms of the latter. Suppose we stipulated the following:[14]

(C) A formula ϕ is true w.r.t. a file F if F + ϕ is true, and false w.r.t. F if F is true and F + ϕ is false.

(C) is of course not literally a truth criterion for utterances, but it implies one. Suppose an expression X of English is uttered under the reading represented by ϕ, and F is the file that obtains prior to the utterance. Then let us call the utterance true (or false) just in case ϕ is true (or false) w.r.t. F in the sense of (C).

Note that the criterion is incomplete, since it fails to say anything about cases where F (the file w.r.t. which ϕ is evaluated, i.e., the file prior to the utterance that ϕ represents) is false. In such cases, F′ will always be false as well. (This is so because, as we have conceived of file change, it is by its nature subject to the law: Once false, always false.) But (C) does not assign either truth or falsity to ϕ if both F and F′ are false. This strikes one as a defect, and we will have more to say about it in section 3.3 below. For now, let us look at the positive aspects of having (C) in our theory.

To the extent that (C) is applicable, it yields intuitively adequate predictions. This is especially noteworthy in regard to utterances with free indefinites in them. Consider the monosentential text "A woman was bitten by a dog," for which (3) is a permissible logical form under our current assumptions.

---

[14] This is similar, though not equivalent, to a proposal by von Stechow (1981) that the "assertion" made by the utterance of a sentence be identified with the material implication whose antecedent is the common ground prior to the utterance, and whose consequent is the content of the sentence. The difference between that and my criterion (C) is this: (C) does not imply that an utterance automatically asserts something true if it is performed when a false common ground obtains. Stechow's proposal implies that, and in this respect his defined notion of truth diverges from its pretheoretical counterpart.

(3)

```
              T
              |
              S
            /    \
         NP₁      S′
          |      /    \
      a woman  NP₂     S″
               /\      /\
            a dog   e₁ was bitten by e₂
```

(3)

T

S

NP_1 — a woman

S'

NP_2 — a dog

S'' — e_1 was bitten by e_2

Suppose the sentence is uttered under the reading represented by (3) in a situation where some file F obtains. It does not matter what exactly F is like, but we must assume two things about it: that it be true, and that (3) be felicitous w.r.t. it. We make the latter assumption because for infelicitous utterances (or readings of utterances), the question of truth never arises in the first place. (If you like, you can write this explicitly into (C), i.e., by amending it to: "A formula ϕ *that is felicitous w.r.t. a file F* is true w.r.t. F … (etc., as above)".) In this particular case, felicity means that neither 1 nor 2 are members of $\text{Dom}(F)$, since they are referential indices of indefinites in (3). Now since F is true, (C) applies and yields, in conjunction with rules (i″) and (ii″), the result:

> The formula (3) is true w.r.t. F if $\text{Sat}(F')$ is nonempty, and false w.r.t. F if $\text{Sat}(F')$ is empty, where:
>
> $\text{Sat}(F') = \{a_N \in \text{Sat}(F): a_1$ is a woman and a_2 is a dog, and a_2 bit $a_1\}$.

It can be shown that $\text{Sat}(F')$ is non-empty if and only if there are a woman and a dog such that the latter bit the former. The proof runs as follows: Trivially, if there isn't any pair of a woman and a dog such that the latter bit the former, then $\text{Sat}(F')$ will be empty. Assume now that there is some woman x and some dog y such that y bit x. Now pick an arbitrary sequence $a_N \in \text{Sat}(F)$. (There must be at least one, since F is by assumption true.) Change a_N into a_N' by replacing a_1 by x. By the assumption of felicity, we know that $1 \in \text{Dom}(F)$. Therefore, by principle (B) of section 2.1.1, it follows that $a_N' \in \text{Sat}(F)$. (Recall the intuition behind that principle: A file cannot impose conditions on the n-th number of its satisfying sequences unless it contains an n-th card.) Now change a_N' further into a_N'' by replacing a_2' by y. Since $2 \in \text{Dom}(F)$, again by the assumption of felicity, it follows by principle (B) that $a_N'' \in \text{Sat}(F)$, too. Given that $a_1'' = x$ and $a_2'' = y$, and given the properties

we have assumed x and y to have, it follows that $a_N'' \in \text{Sat}(F')$, hence that $\text{Sat}(F')$ is non-empty.

We have just seen an illustration of an important point: Indefinites need not be bound in logical form for them to receive what strikes us as existential readings. How is this possible? The answer lies basically in the way the notion of truth is defined for files: a file is true iff *there is* a sequence that satisfies it. So, in a manner of speaking, existential quantification is built into the truth definition and therefore need not be explicitly expressed in logical form.[15] Still — and this is just as important a point — we do not predict an existential reading for just *any* free variable, but only for those that are novel w.r.t. their file of evaluation, which means — thanks to the Novelty-Familiarity-Condition — only to those which are indefinite. Let us look at a text with a free *definite* to appreciate this point.

Consider the trisentential text "She is a woman. He is a dog. She was bitten by him." under the reading represented by (4).[16]

(4)

```
                        T
          ┌─────────────┼───────────────┐
          S             S               S
         ╱ ╲           ╱ ╲             ╱   ╲
   she₁ is a woman  he₂ is a dog  she₁ was bitten by him₂
```

(4)
$$
\begin{array}{c}
\text{T} \\
\diagup \quad \vert \quad \diagdown \\
\text{S} \qquad \text{S} \qquad\qquad \text{S} \\
\end{array}
$$

she$_1$ is a woman $\quad$ he$_2$ is a dog $\quad$ she$_1$ was bitten by him$_2$

I choose this example because it forms in some sense a minimal pair with the one discussed above: (3) and (4) have identical satisfaction conditions. Now imagine the text represented by (4) to be uttered in a situation where initially a file F obtains. Assume, once again, that F is true, and that the utterance is felicitous w.r.t. F. The felicity assumption of course amounts to a different (and even contrary) requirement in this case than it did with the preceding example: This time, it means that both 1 and 2 must be members of $\text{Dom}(F)$. The truth criterion (C) applies with the following result:

> Formula (4) is true w.r.t. F if $\text{Sat}(F')$ is non-empty, and false w.r.t. F if $\text{Sat}(F')$ is empty, where:
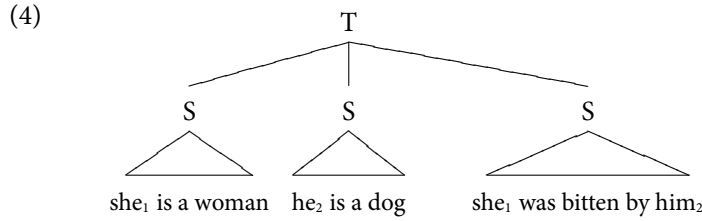
---

[15] Kamp (1981) is to be credited with having been the first to explicitly propose and defend a definition of truth that has "existential quantification built into it" in this sense.

[16] In my treatment of this example, I pretend that indefinites in predicate nominal position do not have referential indices. This may be false, but irrelevantly so. I could just as well have used an example without predicate nominals to illustrate the truth predictions for sentences with definites.

$$\textsc{Sat}(F') = \{a_N \in \textsc{Sat}(F): a_1 \text{ is a woman, } a_2 \text{ is a dog, and } a_2 \text{ bit } a_1 \}$$

This sounds just like the truth condition that was determined for formula (3) above, and yet amounts to something quite different. We can show that (4) will not necessarily be true w.r.t. F just because some woman was bitten by some dog. For suppose the latter to be the case. Then the set $M = \{a_N: a_1 \text{ is a woman, } a_2 \text{ is a dog, and } a_2 \text{ bit } a_1\}$ is non-empty. We also assume $\textsc{Sat}(F)$ to be non-empty, as always when we apply the truth criterion. Still, $\textsc{Sat}(F')$, which is the intersection $M \cap \textsc{Sat}(F)$, may well be empty. Suppose, e.g., that $\{x: \text{there exists } a_N \in \textsc{Sat}(F) \text{ such that } a_1 = x\}$ includes only one element, namely the cat Charlotte, which is not a woman. Given that $1 \in \textsc{Dom}(F)$ (by the assumption of felicity), this is quite possible. In that case, clearly $M \cap \textsc{Sat}(F) = \varnothing$. In other words, not any old pair of a woman and a dog that bit her will do for F' to be true, but it must be a pair whose members also fit cards number 1 and 2 of F.

The difference in truth conditions between an utterance with indefinites (e.g., (3)) and a comparable utterance with definites (e.g., (4)) is thus the following: To verify the former, the individuals to satisfy the uttered sentence may be found anywhere in the domain of individuals. To verify the latter, the individuals to satisfy the uttered sentence have to be found amongst the ones which fit certain already established file cards.

In discussing (4), I have not specified whether the definites are meant anaphorically or deictically, and I did not need to. The truth condition that was derived applies in either case. It does not matter how the cards 1 and 2, which the woman and dog are supposed to fit, ever got into F, by previously uttered NPs, or by contextual salience of certain individuals.

So we have seen that the truth criterion (C) for utterances, which embodies the natural idea that to be a true utterance means, more or less, to lead to a true file, makes adequate predictions where it applies. In contrast with the truth definition we had in Chapter II, it permits us to dispense with at least one half of the construal rule of Existential Closure. (Why this is an advantage, I have yet to say; see below section 4.3.) Let me draw attention to another respect in which it is superior to the previous truth definition:

I have so far considered only utterances of text-size in my illustration of (C). For example, (3) and (4) were logical forms of entire texts, rooted in a T-node. But (C) applies with equally intuitive results to smaller units of an utterance, in particular individual (matrix) sentences that are part of a multisentential text. Take for instance the two-sentence text of (3) followed by (5) and focus on its second half.

(3)   A woman$_1$ was bitten by a dog$_2$.
(5)   She$_1$ hit him$_2$ with a paddle$_3$.

218 | *Definiteness in File Change Semantics*

If we make the usual assumptions, i.e., that we have a true file F at the point where it comes to evaluating (5), and that (5) is felicitous w.r.t. that file F, then criterion (C) predicts:

(5) is true w.r.t. F if $\text{Sat}(F')$ is non-empty, where $\text{Sat}(F') = \{a_N \in \text{Sat}(F): a_3 \text{ is a paddle and } a_1 \text{ hit } a_2 \text{ with } a_3\}$.

And this amounts to (by a reasoning of the sort I have illustrated with respect to (3) and (4)):

(5) is true w.r.t. F if:

there is a sequence $a_N \in \text{Sat}(F)$ and an individual x, such that x is a paddle and a_1 hit a_2 with x.

Now in this case, we know something more specific about F, because we assume (5) to have been uttered as a continuation of (3), and therefore assume F to have been shaped (in part) by the utterance of (3). This means that:

For every $a_N \in \text{Sat}(F)$: a_1 is a woman and a_2 is a dog and a_2 bit a_1.

Therefore, (5) will not be true w.r.t. F unless there is a paddle with which a woman hit a dog that had bitten her.

Likewise, we predict (5) to be false w.r.t. F if, although there is a woman and a dog that bit her, there is no paddle with which any woman hit any dog that bit her.

Recall that the truth definition of Chapter II, section 3, could not have been applied to (5) in this case, for the simple reason that we were then lacking a felicity criterion for anything but complete texts, and felicity was then, as it is now, a precondition for either truth or falsity. So it is ultimately to the credit of our new conception of felicity (as introduced in section 2 above) that the current truth criterion (C) does a better job than its predecessor when it comes to individual sentences that are part of a text.

3.3 Truth of an utterance with respect to a false file

As a definition for the truth and falsity of utterances, (C) falls short of capturing our pretheoretical use of the terms, because it fails to apply in the full range of cases where we are in the habit of making truth value judgments. (C) never applies when an utterance is performed against the background of a false file, a situation which is probably the rule, rather than the exception, in every-day conversation.

Files can be false for either of two reasons: because something false has previously been uttered (and has not been rejected by the audience), or because some never-verbalized falsehoods are among the shared presuppositions of the discourse-participants. Yet, our ability to assess the truth of an utterance is seldom impeded by this. If, for example, we are presented with an utterance of "Alpacas are sheep" that occurs in a situation where it is presupposed that alpacas are native to Australia, we have no difficulty (presuming that we know the facts) to make up our minds that this utterance was false. Likewise, if an utterance of "Alpacas are not sheep" had occurred in the same situation, we would have judged it true. Yet, (C) fails to assign either truth or falsity to either utterance, and indeed, no criterion that solely looks at the truth-values of the files before and after the utterance possibly could predict those judgments: In both cases, we have a false file turning into another false file, because alpacas are not native to Australia.

I said that our naive truth value judgments were *seldom* impeded by the falsity of the file on which the utterance is to take its effect. *Sometimes*, though, they are. Suppose an utterance has just occurred, and been believed, of the sentence:

(6) There will be a concert$_1$ tonight.

But in fact, there will not be. This utterance of (6) was false, and so is the file F that it has produced. Now suppose the next utterance is (7):

(7) It$_1$ will start at eight.

Is the utterance of (7) true or false, and what facts matter for the decision? There is no straightforward answer to this question. About the utterance of (6), and about the utterance of the whole text (6) + (7), we would say they are false. But for (7), the question is somehow inappropriate. And the reason for its inappropriateness is just the falsity of the preceding utterance, i.e., the falsity of the file F at which (7) is supposed to be evaluated.

This example suggests that it is not altogether a defect in (C) that it determines no truth values for utterances evaluated w.r.t. false files. Sometimes such inapplicability coincides with insecure intuitions. But not always, and (C) definitely predicts more insecurity than there is.

What makes the last example different from the one about the alpacas, and makes it different in such away that falsity of the file interferes with our intuitions about truth in one case, but not in the other? Vaguely speaking, in comprehending the truth conditions of "Alpacas are sheep," we can abstract away from the particulars of the current file, including the falsehood it contains, to a degree to which we cannot abstract away from the particulars of the actual file of evaluation when confronted with "It will start at eight." This

is not too surprising, considering that the latter utterance depends for its very felicity on specific features of the actual file, while the former is felicitous with respect to more or less any old file. I suggest that a readiness to abstract away from as much as possible is typical of the way we form intuitive judgments about truth and falsity, and that the readiness for such abstraction is what is missing when we rigidly apply (C).

What does "abstracting away from the particular properties of a file" amount to? Apparently, it amounts to disregarding cards that are in some sense irrelevant, and moreover, disregarding entries on some of the cards that remain. So if we are given an utterance of a logical form ϕ that occurs in a situation where a false file F obtains, then we do not stubbornly insist on considering F itself, but we try to find some true file F^- that is as similar as possible to F, except that (a) certain cards of F whose presence is not required for the felicity of ϕ may be missing from F^-, and (b) certain entries on file cards of F (that are also irrelevant to the felicity of ϕ) may be missing from the corresponding cards of F^-. How many and which entries may be disregarded to get from F to F^- in this way? That depends presumably on considerations of relevance and other vague factors.

The truth criterion that we seem to be applying in real life is thus more like (C′) below than like (C):

(C′) ϕ is true w.r.t. F if and only if, for some true file F^- which relates to F as described above, $F^- + \phi$ is true.

This truth criterion is of course as vague as our characterization of the relation between F and F^- was. The vagueness shows up when we try to decide which entries in the file F we are supposed to consider removed along with a card that we are disregarding. Suppose, for example, you are to disregard card 2 of a file F_1 with the satisfaction set $\textsc{Sat}(F_1) = \{a_N: a_1$ is a woman, a_2 is a dog, and a_2 bit $a_1\}$, and the domain $\textsc{Dom}(F) = \{1, 2\}$. Certainly, the resulting file $F_1{}^-$ will have $\textsc{Dom}(F_1{}^-) = \{1\}$, and $\textsc{Sat}(F_1{}^-)$ will be a superset of $\textsc{Sat}(F_1)$. But the rest is unclear. Will $F_1{}^-$ look as specified in (a), or as specified in (b)?

(a) $\textsc{Sat}(F_1{}^-) = \{a_N: a_1$ is a woman and a_1 was bitten$\}$

(b) $\textsc{Sat}(F_1{}^-) = \{a_N: a_1$ is a woman$\}$

These two and other choices of $F_1{}^-$ are compatible with what I have said about the relation that $F_1{}^-$ is supposed to bear to F_1. They differ in how many of the "entries" in the file are ignored along with the card that is disregarded.

Imprecise though (C′) is, I will leave it at that. It hopefully could be made as precise as it need be to capture our intuitive notion of truth, but the effort required would exceed the importance of the matter for the present purposes.

4 Quantification

For the first half of this chapter, we more or less ignored quantification. This was especially true of sections 1.5 and 2.4, where we were concerned with breaking up the file change that a complex formula brings about into successive smaller file changes corresponding to each of the formula's constituents. The rules by which this was accomplished (cf. (i″) and (ii″), section 2.1.1) covered only cumulative, not quantified, formulas. Moreover, we did not consider in our elimination of Existential Closure that subrule which applies in the nuclear scopes of quantifiers. I will now proceed as follows:

First, I will give a reason for why the file changes induced by quantified utterances should be viewed as sequences of smaller steps. Then I will examine how they could be. Thereafter, I will ask whether Operator Indexing can be eliminated, and also whether Existential Closure is as dispensable inside quantified structures as it turned out to be with respect to texts. The answer will be positive in both respects, a desirable result insofar as it permits imposing narrower limits than we have been assuming on the variety of possible construal rules.

4.1 Quantified structures and file change

Consider an utterance of a simple quantified sentence:

(1)   Everyone laughed,

which, according to our current assumptions, has a logical form like this:

(1′)
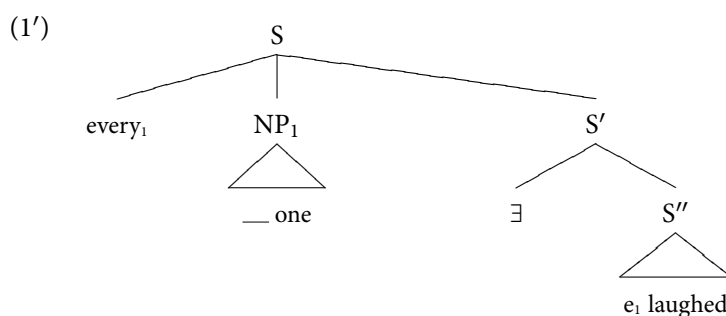
Suppose that (1), under reading (1′), is uttered in a situation where a file F_1 obtains. What is the file $F_1 + (1')$ going to be like? Given the systematic relation between file change and satisfaction conditions that we stated in the form of principle (A) in section 1.2, and given the satisfaction conditions that

were assigned to $(1')$ in Chapter II, section 3, we expect that F1 + $(1')$ will have the following satisfaction set:

$\textsc{Sat}(F_1 + (1')) = \{a_N \in \textsc{Sat}(F_1)$: for every b: if b is a person, then b laughed$\}$.

This is either the whole set $\textsc{Sat}(F_1)$, or the empty set, depending on whether everyone laughed, or not. (Recall that closed formulas like $(1')$, i.e., formulas without free variables in them, are satisfied either by every sequence or by none at all.)

We know that $F_1 + (1')$ has to be the final result of the file change that $(1')$ causes F_1 to undergo. But how do we get there, what are the intermediate steps, if any? Example (1) is too simple to provide any direct evidence of intermediate steps. But there are other examples, whose felicity conditions can only be predicted if we assume certain intermediate steps. Look at (2) or (3):

(2) $Everyone_1$ bought a $pretzel_2$ and ate it_2.
(3) $Everyone_1$ who bought a $pretzel_2$ ate it_2.

Each of these contains the definite pronoun "it_2," which is subject to a requirement of familiarity under the Novelty-Familiarity-Condition. If we evaluated the whole utterance at once, we would predict that this requirement applies to the file which obtains prior to (2) or (3) and that this file must already contain a card number 2 if the utterance is to be felicitous. But intuitively, either utterance is perfectly acceptable when an empty file obtains. Of course, a card number 2 must have been introduced by the time the "it_2" gets evaluated. But by that time, we apparently have a file which has already been affected by the evaluation of "a $pretzel_2$" and therefore cannot fail to meet the requirement. The situation is simple enough, but if we want to deduce the correct predictions from the Novelty-Familiarity-Condition, we have to be able to account for the file change induced by utterances such as these as composed of successive smaller changes. Quantified formulas are no different from cumulative ones in this respect. (Cf. section 2.4. for discussion of analogous facts in cumulative formulas.)

What succession of smaller steps would effect the overall change that we are taking to result from, e.g., $(1')$? Imagine proceeding as follows: Initially, the file is F_1. Now you first take the restrictive term, in this case the NP '__ one_1," and tentatively let it take its effect on F_1. By rule (i''), this brings you from F_1 to F_2, where $\textsc{Dom}(F_2) = \textsc{Dom}(F_1) \cup \{1\}$, and $\textsc{Sat}(F_2) = \{a_N \in \textsc{Sat}(F_1)$: a_1 is a person$\}$. But as I said, the change from F_1 to F_2 is merely tentative, and to keep open your option of undoing it later, you retain in your memory a copy of F_1 as it was before the change. Second, you take the nuclear scope,

in this case $[_{S'} \exists [_{S''} e_1 \text{ laughed}]]$, and let it, again tentatively, take its effect on F_2. Supposing that you ignore the existential quantifier (which here is vacuous anyway), this brings you from F_2 to F_3, where $\text{Dom}(F_3) = \text{Dom}(F_2)$, and $\text{Sat}(F_3) = \{a_N \in \text{Sat}(F_1): a_1 \text{ is a person and laughed}\}$. Your memory now contains copies of two previous files, F_1 and F_2. The third and final step is of a different nature and not as simple as the first two. On the basis of F_2 and F_3, you change F_1, this time definitively, into a file F_4 with the following satisfaction set and domain:

$\text{Sat}(F_4) = \{a_N \in \text{Sat}(F_1): \text{for every } b_N \text{ which agrees with } a_N \text{ on all } i \in \{1\}: \text{if } b_N \in \text{Sat}(F_2), \text{ then } b_N \in \text{Sat}(F_3)\}.$

$\text{Dom}(F_4) = \text{Dom}(F_1)$

Note that in this third step, you take into account the universality of the quantifier "every" and the set of selection indices that it carries (here: $\{1\}$). This completes the evaluation of $(1')$. You can convince yourself that the file F_4 we have arrived at is indeed $F_1 + (1')$, as predicted above on the basis of principle (A) and the satisfaction condition of $(1')$.

The procedure just described is certainly not as simple as the successive evaluation of a cumulative formula. But then, quantification is indeed felt to be more intellectually demanding than mere conjunction. And the basic idea behind this 3-step procedure is not an unnatural one. The first step consists of making a supposition: Suppose something is a person. The second step makes a further supposition: Suppose that person is moreover laughing. The third step draws the generalization that you can always proceed from the first supposition to the second, i.e., that the latter will be true in any instance where the former is.

The general rule which dictates the 3-step procedure just illustrated may be given the following preliminary formulation:

(iii') Let ϕ be a quantified formula, consisting of a universal quantifier with the selectional indices $i_1,...,i_n$, and of the two formulas ϕ^1 and ϕ^2.

Then, for any file F:
$\text{Sat}(F + \phi) = \{a_N \in \text{Sat}(F): \text{for every } b_N \text{ that agrees with } a_N \text{ on all } i \notin \{i_1,...,i_n\}: \text{if } b_N \in \text{Sat}(F + \phi^1), \text{ then } b_N \in \text{Sat}((F + \phi^1) + \phi^2);$

$\text{Dom}(F + \phi) = \text{Dom}(F).$

To give some further illustration of the functioning of this rule, let us look at a slightly more complex sentence, with a logical form as follows:

(4′)

```
                              S
          ┌───────────────────┼───────────────────┐
       every₁,₂            NP₁                    S‴
                    ┌────────┴────────┐       ┌────┴────┐
                  NP₁′              S̄          ∃        S‴‴
                   △           ┌────┴────┐              △
                 ___ one     who₁       S′        e₁ got it₂ back
                          ┌────────┴────────┐
                        NP₂               S″
                         △                 △
                       a cat        e₁ gave e₂ to her₃
```
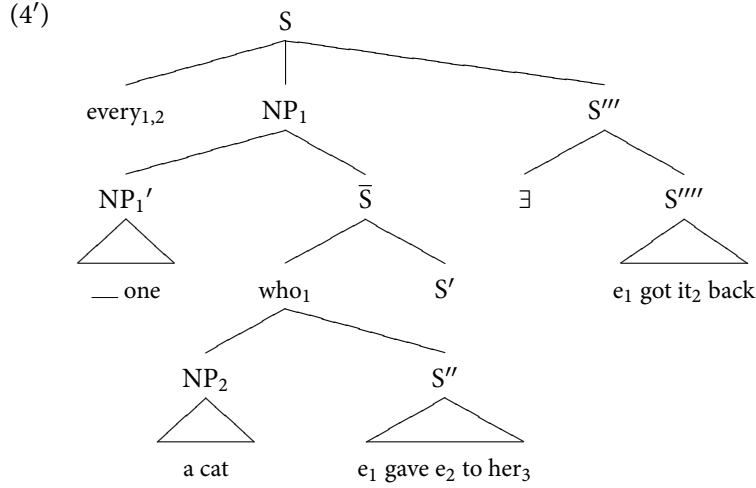
The pronoun "her₃" might either be used deictically or have an antecedent somewhere in the discourse that has preceded (4). The file change that (4) induces can again be viewed in three steps, analogous to the three steps in the case of (1′). For ease of comparison, I will again refer to the initial file as "F_1," and to the files after the first, second, and final steps as "F_2," "F_3," and "F_4," respectively (although they are of course different files from the ones in the other example).

The first step, in which F_1 is tentatively changed to F_2, is this time itself a succession of smaller steps, because it involves the complex cumulative formula NP_1. Without spelling these smaller steps out in detail, let me mention two things: (a) Like every other file change, be it tentative or definitive, the steps from F_1 to F_2 are not carried out without first checking whether the Novelty-Familiarity-Condition is met each time. In this instance, this means that F_1 must be such that $1 \notin \text{Dom}(F_1)$, $2 \notin \text{Dom}(F_1)$ and $3 \in \text{Dom}(F_1)$, since 1 and 2 are indices of indefinites in NP_1 and 3 is the index of a definite in NP_1. (b) After all of NP_1 is evaluated w.r.t. F_1, which is a matter of multiple applications of rules (i″) and (ii″), the resulting file F_2 has the following domain and satisfaction set:

$$\text{Dom}(F_2) = \text{Dom}(F_1) \cup \{1, 2\}$$

$$\text{Sat}(F_2) = \{a_N \in \text{Sat}(F_1): a_1 \text{ is a person, and } a_2 \text{ is a cat, and } a_1 \text{ gave } a_2 \text{ to } a_3\}.$$

Now comes the second step, in which S‴ (or rather, S‴‴, as we are again ignoring the vacuous existential quantifier) is evaluated w.r.t. F_2. This is another occasion for the Novelty-Familiarity-Condition to apply, and this

time it requires that $2 \in \text{Dom}(F_2)$, because of the presence of the definite "it$_2$." The requirement is met, of course. (Perhaps, if traces are definite, there is the additional requirement that $1 \in \text{Dom}(F_2)$, which is likewise met.) At the end of the second step, we have the following file F_3:

> $\text{Sat}(F_3) = \{a_N \in \text{Sat}(F_1):$ a_1 is a person, and a_2 is a cat, and a_1 gave a_2 to a_3, and a_1 got a_2 back$\}$.

We are ready for the third step, which consists of the operation that was specified under (iii′) above. If we supply the actual values of $\text{Sat}(F_2)$ and $\text{Sat}(F_3)$ for $\text{Sat}(F + \phi^1)$ and $\text{Sat}((F + \phi^1) + \phi^2)$, respectively, (iii′) yields the following:

> $\text{Sat}(F_4) = \{a_N \in \text{Sat}(F_1):$ for every b_N that agrees with a_N on all $i \notin \{1, 2\}$: if b_1 is a person, b_2 is a cat, and b_1 gave b_2 to b_3, then b_1 is a person, b_2 is a cat, b_1 gave b_2 to b_3, and b_1 got b_2 back$\}$.

Because of the agreement requirement, b_3 must always equal a_3. So the above reduces to:

> $\text{Sat}(F_4) = \{a_N \in: \text{Sat}(F_1):$ for every b_1 and b_2:
> if b_1 is a person, b_2 is a cat, and b_1 gave b_2 to a_3, then b_1 got b_2 back$\}$.

This is as it should be. The present formulation of (iii′) thus works for this example as well as the previous one.

It should be noted that (iii′) also takes care of the problem that I pointed out at the beginning of this section, i.e.: how do we predict the felicity conditions of complex quantified sentences correctly? By evaluating quantified structures in three steps (the first two of which may themselves be successions of smaller steps), and by applying the Novelty-Familiarity-Condition with each minimal step, we can account for the felicity facts. Just look at our discussion of (4) again and focus on the definite pronoun "it$_2$." If we had evaluated (4) in one single step, the Novelty-Familiarity-Condition would have inadequately required for 2 to be in the domain of the initial file F_1. But as we proceeded in small steps, the requirement that the Novelty-Familiarity-Condition imposed with respect to "it$_2$" was that 2 be in the domain of a certain intermediate file, F_2. And this requirement was met because F_2 had in part been shaped by the evaluation of the indefinite "a cat$_2$" earlier in the same sentence. Examples (2) and (3) can be dealt with for fully analogous reasons.

Once we have characterized the file change potential of universal sentences, it is not hard to give analogous characterizations for other quantifiers and operators:

(iv′) Let ϕ be a quantified molecular formula, consisting of an existential quantifier with the selection indices $i_1,...,i_n$ and of the formula ψ.

Then, for any file F:
SAT(F + ϕ) = {$a_N \in$ SAT(F): there is some b_N such that b_N agrees with a_N on all i $\notin$ {$i_1,...,i_n$} and $b_N \in$ SAT(F+ψ)};
DOM(F + ϕ) = DOM(F).

(v′) Let ϕ be an operator-headed molecular formula, consisting of a negation operator and the formula ψ.

Then, for any file F:
SAT(F + ϕ) = {$a_N \in$ SAT(F): $a_N \notin$ SAT(F + ψ)};
DOM(F + ϕ) = DOM(F).

These rules describe two-step procedures rather than three-step procedures, because existential and negated formulas do not have restrictive terms. Otherwise, the idea is the same as with universal formulas: You first perform a tentative file change, which amounts to adding ψ as a supposition. Then you calculate the definitive file change on the basis of the initial file F and the auxiliary file F + ψ, where the particular nature of this calculation depends on the force of the operator involved (and on its selection indices, if any). Note that (iv′) and (v′), just like (iii′), interact with the Novelty-Familiarity-Condition to yield certain predictions concerning the projection of felicity conditions in existential and negated sentences. For instance, (v′) implies that negation is a "hole" in the sense of Karttunen (1973). But I will not elaborate here on this aspect of the rules any further. The reader is referred to Heim (in preparation) [Heim (1983b)].

Given rule (iv′), we are now able to determine in particular the file change induced by a universal sentence with non-vacuous existential closure over the nuclear scope, such as (5).

(5) Every man saw a cat.

However, we will reconsider our analysis of such sentences shortly, and will come to abandon the rule of Existential Closure altogether, thus rendering rule (iv′) inapplicable in the evaluation of examples like (5).

4.2 Selection indices eliminated

In this and the next section we will strive to eliminate from our system of construal rules both Quantifier Indexing and (the remaining subrule of) Existential Closure. It is easily seen that this is, in principle at least,

possible. But the proof of possibility is trivial and in itself not interesting. The interesting question does not concern possibility, but desirability: What do we gain by eliminating those rules, and do the gains make up for the losses (if any)? I will argue that they do.

Here is why elimination of Existential Closure and Quantifier Indexing must be possible in principle: Both rules apply obligatorily and in a completely deterministic fashion. In other words, where the existential quantifiers go, and where which selectional indices belong, is always predictable from other properties of the logical form under consideration; the two rules add only redundant material.

Given that selectional indices are redundant, we should be able to rewrite, e.g., rule (iii′) in such a way as to make no reference to them. Let us try that. Suppose we do not assign any selectional indices during construal, neither in the course of applying Quantifier Construal, nor by Quantifier Indexing. This means that the logical form of sentence (4), for instance, no longer looks like (4′) above, but rather like (4″):

(4″)
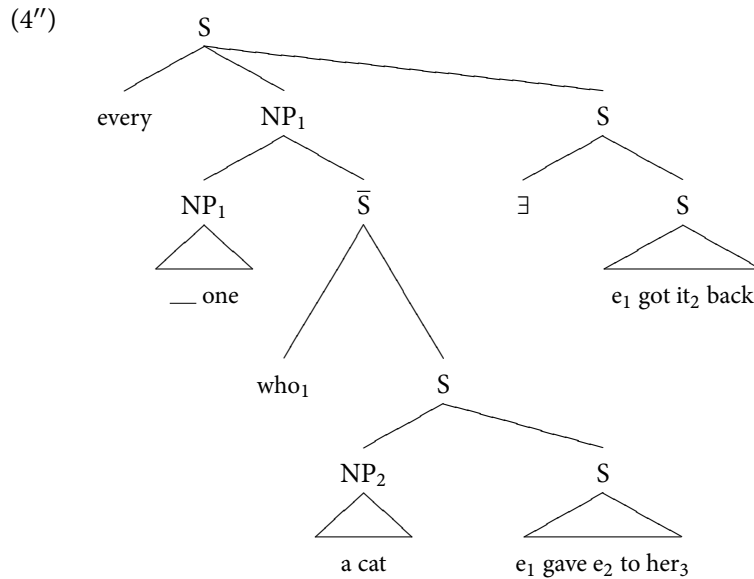
We could still assign (4″) the same file change potential that we used to assign to (4′), if we had the following rule (iii″) instead of (iii′):

(iii″) Let ϕ consist of a universal quantifier and the two formulas ϕ^1 and ϕ^2.

Then, for any file F:
$\text{Sat}(F + \phi) = \{a_N \in \text{Sat}(F): \text{for every } b_N \text{ which agrees with } a_N \text{ on all } i \notin$

$\{i: \phi^1$ contains some [-definite] $NP_i\}$:
if $b_N \in \text{Sat}(F + \phi^1)$, then $b_N \in \text{Sat}((F + \phi^1) + \phi^2)\}$.

Of course, the reason why (iii″) would apply to (4″) with the intended result is simply that (iii″) has a version of Quantifier Indexing built into it, so to speak: The agreement requirement imposed on b_N in (iii″) makes reference to the indices of indefinites contained in ϕ^1 just like Quantifier Indexing used to do. (Note that for (iii″) to apply as intended we have to adopt a stipulation we have toyed with before, i.e., that quantifying NPs, and hence their determinerless remnants in logical form, are [-definite].)

To abolish Quantifier Indexing while replacing rules like (iii′) by ones like (iii″) would seem to be a rather pointless exercise. One can even make an argument that it leads to a worse theory than we had before. The argument is based on the observation that (iii″) lacks a property that we have so far tacitly assumed to be a general property of the rules of interpretation, namely, the property of "compositionality." A "compositional" rule, in the sense intended here, makes reference only to the interpretations of the immediate constituent formulas of the formula being interpreted, but is insensitive to the internal composition of those constituents. The rules assigning satisfaction conditions to logical forms that we assumed in Chapter II were all compositional in this sense, and so were the rules assigning file change potentials that were given in previous sections of this chapter.

But (iii″) is not compositional, because the file change potential it assigns to ϕ depends on properties of ϕ^1 other than its file change potential, viz., on the presence of indefinites inside ϕ^1 and on their indices. The hypothesis that only compositional rules are possible in the interpretation component is, ceteris paribus, a worthwhile hypothesis to entertain, and this, if nothing else, should stop us from eliminating Quantifier Indexing at the expense of admitting interpretation rules like (iii″).

However, there is a different and more interesting way of bypassing the need for Quantifier Indexing as a construal rule and of interpreting structures like (4″) directly. Suppose we assume the following rule (iii‴) instead of either (iii′) or (iii″):

(iii‴) Let ϕ consist of a universal quantifier and the two formulas ϕ^1 and ϕ^2.

Then, for any file F:
$\text{Sat}(F + \phi) = \{a_N \in \text{Sat}(F):$ for every b_N which agrees with a_N on $\text{Dom}(F):$
if $b_N \in \text{Sat}(F + \phi^1)$, then $b_N \in \text{Sat}((F + \phi^1) + \phi^2)\}$.

(iii‴) neither violates compositionality, nor is it otherwise more complicated than (iii′). Nevertheless, it can be shown to correctly predict the file change

potentials of universal sentences while operating on impoverished logical forms like $(4'')$ with no selection indices marked.

Applied to $(4'')$, (iii''') yields the following:

(6) $\text{SAT}(F + (4'')) = \{a_N \in \text{SAT}(F):$ for every b_N which agrees with a_N on $\text{DOM}(F):$
if b_1 is a person, b_2 is a cat, and b_1 gave b_2 to b_3, then b_1 is a person, b_2 is a cat, b_1 gave b_2 to b_3, and b_1 got b_2 back$\}$.

(In calculating this, I relied on our earlier calculations, in which we already determined the satisfaction sets of the auxiliary files $F + \phi^1$ and $(F + \phi^1) + \phi^2$. Note that there is no difference between the old logical form $(4')$ and the new logical form $(4'')$ as far as those two files are concerned, at least not if we continue to disregard the vacuous existential quantifier.)

Is (6) the right result? Well, that depends on what $\text{DOM}(F)$ is. If $\text{DOM}(F)$ happens to contain 3, but neither of 1 or 2, then (6) amounts to precisely what we want to predict, namely:

$\text{SAT}(F + (4'')) = \{a_N \in \text{SAT}(F):$ for every $b_1, b_2:$
if b_1 is a person, b_2 is a cat, and b_1 gave b_2 to a_3, then b_1 got b_2 back$\}$.

However, if $\text{DOM}(F)$ contains 1 or 2, or fails to contain 3, we seem to get counterintuitive predictions of various sorts. For instance, if $\text{DOM}(F)$ contains all three numbers, (6) turns out to describe the sort of file that we would expect to result from an utterance like: "Everyone who gave a cat to *anyone* got it back," but which cannot result from any intuitively available reading of (4):

$\text{SAT}(F + (4'')) = \{a_N \in \text{SAT}(F):$ for every $b_1, b_2, b_3:$
if b_1 is a person, b_2 is a cat, and b_1 gave b_2 to b_3, then b_1 got b_2 back$\}$.

So it does not look like (6) is generally the right prediction; it appears to be right in some cases and wrong in others, depending on the membership of $\text{DOM}(F)$.

However, the cases where (6) seems to come out wrong need not really concern us at all, because *in no such case will $(4'')$ be felicitous w.r.t. F.* For remember that felicity of $(4'')$ w.r.t. F requires that each minimal step in the evaluation of $(4'')$ must meet the Novelty-Familiarity-Condition. Given that both "__ one$_1$" (i.e., the remnant of "everyone") and "a cat$_2$" are indefinites, and "her$_3$" is a definite, the restrictive term "__ one$_1$ who$_1$ gave a cat$_2$ to her$_3$" can only be evaluated w.r.t. F if $1 \notin \text{DOM}(F)$, $2 \notin \text{DOM}(F)$, and $3 \in \text{DOM}(F)$. If $\text{DOM}(F)$ meets these conditions, (6) is in fact the right prediction, as we have seen. If $\text{DOM}(F)$ does not meet these conditions, then the evaluation of

(4″) w.r.t. F is blocked at the first step already, and we never get a chance to construct F + (4″) at all. In that case, the question whether (6) describes the satisfaction set of F + (4″) correctly or incorrectly simply never arises.

What we have just seen is important: An interpretation rule like (iii‴), which does not by itself entail adequate predictions about the file change potentials (and hence the truth conditions) of universal sentences, nevertheless yields adequate predictions in the context of a theory in which it interacts with conditions on felicity, especially the Novelty-Familiarity-Condition.

Of course, we have seen only one example to substantiate this: But it should not be hard to see that (iii‴) will work quite generally. The crucial point is for (iii‴) to draw just the right line between those indices on which b_N is to agree with a_N and those on which it is to vary. It used to be the selection indices that determined where this line was drawn: b_N was to vary on the selection indices and remain fixed everywhere else. The selection indices, in turn, correlated with indefinites occurring in the restrictive term, because that correlation was imposed by Quantifier Indexing. So there was in effect a correlation between indices on which b_N varied and indefinites which occurred in the restrictive term. As the rule stands now, the line is drawn according to the domain of F: b_N gets to vary on the indices not in Dom(F) and remains fixed elsewhere (elsewhere being Dom(F)). Dom(F), in turn, stands in a systematic relation to the set of occurrences of indefinites in the restrictive term, a relation that is imposed by the Novelty-Familiarity-Condition. As a result, there is again a correlation between indices on which b_N varies and indefinites in the restrictive term.

It ultimately comes to the same thing, whether you rely on selection indices or on the domain of F to draw the line. (The only difference is an irrelevant one: If you go by Dom(F), b_N will be permitted to vary on any indices which do not occur in either the previous discourse or the universal formula, whereas if you go by selection indices, b_N is kept fixed on those.)

So selection indices have turned out to be superfluous: The purpose they were designed to fulfill can equally well be served by making reference to the domain of the file and the properties that the Novelty-Familiarity-Condition imposes on that domain. The logical step to take is to get rid of selection indices altogether, and to abolish Operator Indexing, the rule that introduces them, along with them. At the same time, let us adopt (iii‴) as the rule that interprets universally quantified formulas, as well as analogously revised rules for other quantifiers and operators. We must now also adopt definitively the stipulation that quantifying NPs are [-definite], since this is a crucial precondition for the adequate functioning of rules like (iii‴). (Example (4″) above shows this clearly: Unless "__ one₁" is definite, 1 could be an element of Dom(F) without violating the Novelty-Familiarity-

Condition, in which case we would wrongly predict that (4″) quantifies over cats that *some fixed person* gave to some other fixed person.)

The revision of the theory that has just been accomplished brings us one step closer to a "pure" familiarity theory of definiteness, by which I mean a theory which considers all systematic distinctions between definites and indefinites to be derivative upon the association of definiteness with familiarity and indefiniteness with novelty. As long as Quantifier Indexing was part of the theory, there were three independent differences between definites and indefinites: the Novelty-Familiarity-Condition (formerly: the Novelty Condition) discriminated between them, only definites presupposed their descriptive content, and Quantifier Indexing applied exclusively to indefinites. With the elimination of Quantifier Indexing, these three differences have been reduced to two. We still predict the empirical generalization that Quantifier Indexing was designed to capture, i.e., that indefinites get bound by quantifiers in their environment, while definites do not — but we predict it as an indirect consequence of the Novelty-Familiarity-Condition.

As far as we can tell at this point, the empirical predictions of the theory have remained unaffected by the revision: It was not motivated by a desire to describe the facts more accurately, but was just an application of Occam's razor. However, we will see in section 5 that the simplified theory is also better equipped to predict correctly the behavior of definite descriptions, which, as we noticed earlier (see Chapter II, section 6), sometimes seem to get bound by quantifiers despite their definiteness.
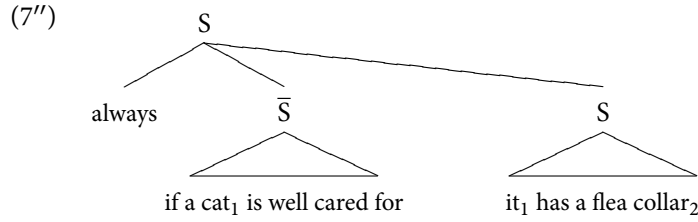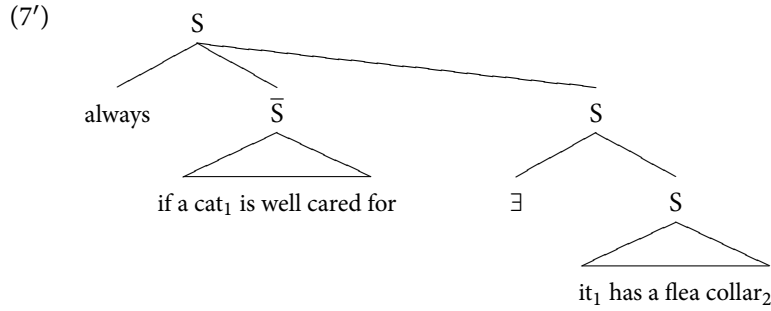
Note, incidentally, that our success in eliminating Operator Indexing depended on the assumption that file change potentials, not satisfaction conditions, are assigned by the recursive system of interpretation rules. A rule like (iii‴) characterizes $\text{SAT}(F + \phi)$ by making crucial reference to $\text{DOM}(F)$, which could not be referred to in a rule that assigns satisfaction conditions to ϕ independently of which file ϕ is evaluated against. This provides another reason to give file change potentials priority over satisfaction conditions in the formulation of the interpretation rules.

4.3 Why are the nuclear scopes of operators existentially closed?

We have reconsidered the Novelty Condition, the second subrule of Existential Closure, and Operator Indexing, and in each case we have come to the conclusion that the relevant generalizations are less successfully accounted for within the construal component than they are in terms of principles of file construction (in particular, the Novelty-Familiarity-Condition) and principles of file verification (i.e., the definition for truth of a file). It is only natural at this point to conjecture that the first subrule of Existential Closure,

the one that applies to the nuclear scopes of operators, does not really belong in the construal component either. Should we not try to assume logical forms like (7″) instead of the currently generated (7′) for sentences like (7), and leave it to semantic interpretation to predict the perceived narrow-scope existential reading of "a flea collar?"

(7)   If a cat is well cared for, it always has a flea collar.

(7′)

```
                    S
          ／       ｜         ＼
                             ＼
     always        S̄            S
              ／  ＼        ／  ＼
             ／    ＼      ／    ＼
      if a cat₁ is well cared for   ∃       S
                                        ／  ＼
                                       ／    ＼
                                it₁ has a flea collar₂
```

(7′) tree with nodes:
- S dominating: always, S̄, S
- S̄ dominating: if a cat$_1$ is well cared for
- lower S dominating: ∃, S
- lowest S dominating: it$_1$ has a flea collar$_2$

(7″)

```
                    S
          ／       ｜         ＼
                             ＼
     always        S̄            S
              ／  ＼         ／  ＼
             ／    ＼       ／    ＼
      if a cat₁ is well cared for   it₁ has a flea collar₂
```

(7″) tree with nodes:
- S dominating: always, S̄, S
- S̄ dominating: if a cat$_1$ is well cared for
- S dominating: it$_1$ has a flea collar$_2$

Suppose we rewrite the interpretation rule for universal quantification once more, changing (iii‴) to (iii⁗):

(iii⁗) Let ϕ consist of a universal quantifier and two formulas ϕ^1 and ϕ^2.

Then, for any file F:
$\textsc{Sat}(F + \phi) = \{a_N \in \textsc{Sat}(F)$: for every b_N which agrees with a_N on all $i \in \textsc{Dom}(F)$: if $b_N \in \textsc{Sat}(F + \phi^1)$, then there is some c_N which agrees with b_N on all $i \in \textsc{Dom}(F + \phi^1)$ such that $c_N \in \textsc{Sat}((F + \phi^1) + \phi^2)\}$.

Obviously, this rule assigns the intended interpretation to (7″), because it has existential quantification over the nuclear scope built into it.[17] If we give analogous reformulations for negation and whatever other operators there are, then Existential Closure as a construal rule becomes superfluous and can be eliminated.

---

[17] Kamp (1981) also interprets quantified sentences in such a way that existential quantification over the nuclear scope is built into the interpretation, rather than being explicitly represented in the structures to be interpreted.

It may not seem that we gain much insight by eliminating Existential Closure in this way, but there is at least one thing to be said for it: It clears the way for a more restrictive theory of what kinds of construal rules are possible: We can now hypothesize that there cannot be obligatory construal rules. As far as our system is concerned, we have eliminated all those rules that would have been counterexamples to this hypothesis: Existential Closure (both subrules) and Quantifier Indexing. The remaining ones (NP-Indexing, NP-Prefixing, and Quantifier Construal) are supposedly optional rules. (Note that that does not preclude their being in effect obligatory, say, because non-application would mean that the resulting structure is not a well-formed, or not an interpretable, logical form. By calling them nevertheless optional, I mean that obligatory application need not be stipulated for them.) This hypothesis makes a worthwhile contribution to constraining the class of possible grammars.

Let us hence continue with the assumption that logical forms look like (7″) and that the existential force of indefinites in the nuclear scope is somehow a matter of how such structures get interpreted. It remains an open question why the existential force should arise in just this environment. If we build existentialization into each individual rule that interprets an operator-headed structure, as I have suggested, we are surely missing a generalization. It would be better to distinguish the idiosyncratic lexical meaning of each operator from the general rule that interprets operator-headed formulas. That way, we would not have to build existentialization of the nuclear scope into more than one rule. There are presumably various ways of implementing this suggestion, but I will not be more concrete here.

A further question that I want to raise, but will not answer, is this: Is there anything that the evaluation of the nuclear scope in a quantified formula has in common with the assessment of a file in terms of its truth or falsity? And if so, could that common property explain why both involve implicit existential quantification?

With these issues remaining unresolved, the elimination of Existential Closure from the construal component and the concomitant revision of the interpretation rules for operators rest on rather weak motivation. Future research will hopefully lead to an understanding of operators that supports our decision to regard existential closure as a genuinely semantic, rather than a construal, phenomenon.

4.4 Summary of the revised interpretation rules

We have now concluded the process of successive modification whose point of departure were the five rules of interpretation of the extensional satis-

faction semantics of Chapter II, section 3.2. What we have arrived at may be called an extensional file change semantics. It consists of the following four rules, which jointly amount to a simultaneous recursive characterization of $\textsc{Sat}(F + \phi)$ and $\textsc{Dom}(F + \phi)$. Note that the fifth rule for existentially quantified formulas is no longer needed, due to the elimination of existential quantifiers from logical forms.

Let a model $\langle A, \text{Ext} \rangle$ for English be given.

(I) Let ϕ be an atomic formula, consisting of an n-place predicate ζ and an n-tuple of variables $\langle \alpha^1,...,\alpha^n \rangle$ whose indices are $i_1,...,i_n$, respectively. Then:
$\textsc{Sat}(F + \phi) = \{a_N \in \textsc{Sat}(F): \langle a_{i_1},...,a_{i_n} \rangle \in \text{Ext}(\zeta);$
$\textsc{Dom}(F+ \phi) = \textsc{Dom}(F) \cup \{i_1,...,i_n\}.$

(II) Let ϕ be a cumulative molecular formula with the immediate constituent formulas $\phi^1,...,\phi^n$ (in that order). Then:
$\textsc{Sat}(F + \phi) = \textsc{Sat}(... (F + \phi^1) + \phi^n);$
$\textsc{Dom}(F + \phi) = \textsc{Dom}(... (F + \phi^1) + \phi^n).$

(III) Let ϕ be a quantified molecular formula, consisting of a universal quantifier and the two formulas ϕ^1 and ϕ^2 (in that order). Then:
$\textsc{Sat}(F + \phi) = \{a_N \in \textsc{Sat}(F):$ for every $b_N \mathrel{\widetilde{\textsc{Dom}(F)}} a_N$ such that $b_N \in \textsc{Sat}(F+ \phi^1)$, there is some $c_N \mathrel{\widetilde{\textsc{Dom}(F + \phi^1)}} b_N$ such that $c_N \in \textsc{Sat}((F + \phi^1) + \phi^2)\};$

$\textsc{Dom}(F+ \phi) = \textsc{Dom}(F).$

(IV) Let ϕ be an operator-headed molecular formula, consisting of a negator and the formula ψ. Then:
$\textsc{Sat}(F + \phi) = \{a_N \in \textsc{Sat}(F):$ there is no $b_N \mathrel{\widetilde{\textsc{Dom}(F)}} a_N$ such that $b_N \in \textsc{Sat}(F + \psi)\};$
$\textsc{Dom}(F + \phi) = \textsc{Dom}(F).$

The notation "$a_N \mathrel{\widetilde{\textsc{m}}} b_N$" that I have employed here abbreviates: "a_N agrees with b_N on all $i \in M$."

5   Definite Descriptions and Related Issues

This section is designed to tie up various loose ends, especially from section 6 of Chapter II. We will first restate the stipulation that definites presuppose their descriptive content in the context of the current theory. Then we will address the phenomenon of novel definites, which appears to falsify the Novelty-Familiarity-Condition. We will account for it by appealing to a
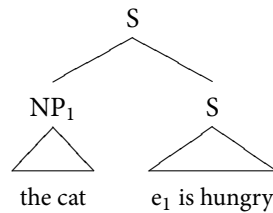
strategy of "acconmodation" in the sense of Lewis (1979). This account will naturally generalize to the examples of "narrow-scope" definite descriptions that were seen to create problems in the previous chapter. Finally, there will be some remarks concerning the distribution of pronouns as opposed to full definites, including some brief discussion of "paycheck"-pronouns and other instances of what is sometimes called "pronouns of laziness." In dealing with these topics, we will benefit from the file change conception of semantics in general, and from the reduction of Quantifier Indexing to the Novelty-Familiarity-Condition, as accomplished in the previous section, in particular.

5.1   *The Extended Novelty-Familiarity-Condition*

It was argued in Chapter II, section 6.2, that descriptive definites presuppose their descriptive content. For instance, if (1) is given the reading (1′), then it carries the presupposition (1a).

(1)   The cat is hungry

(1′)

```
                    S
              _____|_____
             |             |
           NP_1            S
            /\             /\
           /__\           /__\
         the cat      e_1 is hungry
```

(1a) $cat(x_1)$

The function of the presupposition, as we saw, was to narrow down the choice of referents that could be associated with "the cat$_1$" when used deictically. We also saw how the choice of antecedents that could be associated with an anaphoric use of "the cat$_1$" (e.g., when (1) occurs as part of a larger sentence which contains other NPs that could potentially be coindexed with "the cat$_1$") is restricted because of presuppositions that arise from (1a) by principles of presupposition projection (cf. Chapter II, section 6.3).

What does it mean in the framework of our current theory that (1′) presupposes (1a)? We have assumed, and will continue to assume, that presuppositions are felicity conditions, i.e., that a sentence under a certain reading can only be uttered felicitously in a context in which the discourse participants already presuppose the presuppositions that the sentence has under that reading. What the discourse participants presuppose in a context is represented by the file which obtains in that context. (See 1.4 above, where files were viewed as "common grounds.") What it means for (1′) to

presuppose (1a) is therefore the following: (1′) is felicitous w.r.t. a file F only if F already contains the information expressed by (1a).

For a file F to "contain," in the intended sense, a piece of information can, but need not, mean that that information is explicitly written down in F. It will generally suffice for it to be merely entailed by what is written in F.[18] For instance, F will contain the information expressed by (1a) if it says "is a cat" on card 1 of F, but it will also contain that information if it says "is a calico kitten" on card 1, or if it says somewhere in F (presumably not on any particular card at all) that all individuals are cats. The relevant notion of entailment is the following:

A file F entails a formula ϕ iff for every world w:
$$\text{SAT}_w(F) \subseteq \text{SAT}_w(F + \phi).$$

For this definition to be applicable, files must be characterized not just in terms of their *actual* satisfaction sets, but rather in terms of their satisfaction sets in each possible world. We have been confining ourselves to extensional semantics throughout this chapter, but it would be easy to give intensional variants of the interpretation rules (I) to (IV) so that they specify $\text{SAT}_w(F + \phi)$ for an arbitrary world w.

The stipulation that (1′) has (1a) as a presupposition may now be formulated as follows:

Suppose (1′) is to be evaluated w.r.t. a file F.
Then it is required that F entails "cat (x_1)."

This falls under a general rule , call it the "Descriptive-Content-Condition":

Whenever a [+definite] formula NP_i is to be evaluated w.r.t. a file F, it is required that F entails NP_i.

This condition is taken to function in the same way as the Novelty-Familiarity-Condition: It must be met each time before a minimal step of file change is taken at which it is applicable.[19] If it is not met, file change cannot proceed any further.

---

[18] Wasow (1972) formulated an analogous observation in his "Novelty Constraint," which is a prohibition against anaphoric NPs whose descriptive content carries "novel," i.e., non-entailed, information about the referent. This is not the same sense in which we have used the term "novel" so far, but I will suggest below that it is a related sense.

[19] Note that the minimal steps at which the Descriptive-Content-Condition may become applicable are not always "minimal" in the sense of involving atomic formulas. If NP_i is complex, it may well be that every one of its atomic subformulas is felicitous w.r.t. the relevant intermediate file, yet $F + NP_i$ cannot be constructed without violating the Descriptive-Content-Condition.

Like the Novelty-Familiarity-Condition, the Descriptive-Content-Condition applies with equally adequate results to deictic and anaphoric definites, and it does so without there being a special mechanism of presupposition projection. Sentence (1), for instance, may be used to talk about a contextually salient cat. We assume (cf. 2.2 above) that when a cat is making itself salient it causes a file card to crop up on which are written the properties under which it is making itself salient. Normally (especially when we are dealing with *perceptual* salience) being-a-cat will be among those properties.[20]. Therefore, the file will entail that x_i is a cat (i being the number of the card that has thus been caused to crop up, whatever number that may be), and it will therefore be felicitous, both in view of the Novelty-Familiarity-Condition and of the Descriptive-Content-Condition, to utter (1) under a reading where the index of "the cat" is i, i.e., under a reading in which "the cat" in effect refers to the cat that has been making itself salient.

On the other hand, sentence (1) might appear as part of a text in which a cat has been previously mentioned, e.g., text (2):

(2)   There is a cat_i behind you. The cat_i is hungry.

Here it will be a result of the evaluation of the first sentence that the file comes to contain a card with the entry "is a cat" on it and thus comes to entail that x_i is a cat (i being again the number of the relevant card, whatever it may be). So the definite "the cat_i" is again licensed by the pertinent felicity conditions (Novelty-Familiarity and Descriptive-Content), and we get what is in effect an anaphoric reading of the definite. This case is completely analogous to the case of the deictic reading.

Neither the Novelty-Familiarity-Condition nor the Descriptive-Content-Condition can be deduced from the other, but there is nevertheless a close relation between the two. On an impressionistic level, one might say that both conditions have to do with "familiarity" in some sense: The Novelty-Familiarity-Condition imposes the requirement that the definite must be associated with an already established, i.e., "familiar," card. The Descriptive-Content-Condition adds to this the requirement that the *descriptive content* of the definite be already established, or as we might say, already "familiar" too. (As we have seen, to be "familiar" in this second sense means to be entailed.) On a more technical level, we observe that the two conditions overlap to a certain extent: Whenever NP_i is a noun phrase with a non-trivial

---

[20] It need not always be: a cat may sometimes make itself salient under disguise. But we need not worry about this possibility here, since this is precisely the sort of situation in which "the cat" is not a suitable expression to refer to a salient cat. The Descriptive-Content-Condition automatically predicts this.

descriptive content, then if it meets the Descriptive-Content-Condition, it automatically meets the Novelty-Familiarity-Condition as well. By a "trivial" descriptive content, I here mean a predicate that is already presupposed to be true of every individual on the domain whatsoever. For instance, if F entails [the cat$_i$], and there is at least one world in which F is true and which has a non-cat in it, then principle (B) of section 2.1.2 ensures that i ∈ Dom(F).[21]

Let us therefore think of the two conditions as two aspects of one more inclusive condition, which we will refer to as the *Extended-Novelty-Familiarity-Condition*:

For ϕ to be felicitous w.r.t. F it is required for every NP$_i$ in ϕ that:

(i) if NP$_i$ is [-definite], then i ∉ Dom(F);

(ii) if NP$_i$ is [+definite], then

    (a) i ∈ Dom(F), and

    (b) if NP$_i$ is a formula, F entails NP$_i$.

It might be possible to formulate this condition in simpler and less redundant ways, but let us adopt the present version for the time being.

5.2   Novel definites and accommodation

According to the theory I have presented so far, every use of a definite NP requires that there already be an appropriate card in the file. Violations of this requirement result in infelicity and the process of file change comes to a halt. A definite can never introduce a new discourse referent. The Novelty-Familiarity-Condition strictly excludes this option, there are no two ways about that.

But in fact, there are many uses of definites, in particular definite descriptions, which do not fit this theory. If we look at the taxonomies of usages of the definite article that are found in the literature, it is apparent that we cannot account for the majority of them. Hawkins (1978), for instance, lists eight usage types, only two of which seem to obey the Novelty-Familiarity-Condition under the assumptions we have been making: the anaphoric use,

---

[21] Proof: By hypothesis, there are w, a_N and x such that $a_N \in SAT_w(F)$ and x is not a cat in w. Consider the sequence b_N which is defined as follows: $b_i = x$, and for all $j \neq i$, $b_j = a_j$. Suppose now that i ∉ Dom(F). Then, by (B), $b_N \in SAT_w(F)$. Since F entails [the cat$_i$], $b_N \in SAT_w(F + [the cat_i])$ as well. This means that b_i is a cat in w, contrary to our assumption. This disproves i ∉ Dom(F). (It should somehow be possible to prove the corresponding claim for non-trivial NP$_i$ in general.)

and what he calls the "visible situation use," i.e., a use that requires perceptual salience of the referent. Some of the other six uses are exemplified as follows (cf. Hawkins, op. cit., Ch. 3):

As I am walking up a driveway, someone says to me:

(3)   Watch out, the dog_i will bite you.

There was no previous discourse, there is no dog in sight, and I had no reason to believe that any dog lived here before I heard the utterance. ("Immediate situation use")

Someone tells me on the phone:

(4)   The sun_i is shining.

There was no previous mention of a sun. ("Larger situation use")

(5) strikes one as a self-contained text, which does not presuppose contextual salience or previous mention of an author:

(5)   John read [a book about Schubert]$_i$ and wrote to the $author_j$.

("Associative anaphoric use")

To accommodate such examples, we will not abandon the Extended Novelty-Familiarity-Condition, but we will assume that there is a mechanism by which utterances that fail this condition can be rendered felicitous after all. This is a mechanism of the sort that Lewis (1979) discusses under the name of "accommodation," and it has been called "bridging" in the psychological literature.[22] In our terms, accommodation is an adjustment of the file that is triggered by a violation of a felicity condition and consists of adding to the file enough information to remedy the infelicity. For instance, suppose a file F with $i \notin \text{Dom}(F)$ obtains when (3) gets uttered. (3) is infelicitous w.r.t. F according to the Extended Novelty-Familiarity-Condition. If no accommodation occurs, file change stops there. If the option of accommodation is taken, however, F is adjusted to F′, which contains an additional card i with an entry more or less like: "is a dog somewhere close by." (3) is now felicitous w.r.t. F′, and file change can proceed from there, turning F′ into F′ + [the dog_i], then turning that into (F′ + [the dog_i]) + [e_1 will bite you], and so forth.

Under which conditions is accommodation an available option, and what exactly is added to the file when the option is taken? These questions

---

[22] See, e.g., Clark and Haviland (1977), Clark (1977).

are by no means easy to answer, as can be seen from the attention that they (or rather the analogous questions that correspond to them in other theoretical frameworks) have received in the literature, ranging from the work of traditional grammarians to much recent work in psychology and artificial intelligence. From the point of view of much of that work, they are perhaps the only non-trivial questions that a theory of definiteness faces. Be that as it may, I can say only very little about the rules that govern accommodation, none of which is new.[23]

Contrary to what one might have thought, accommodation in response to definites is not normally a matter of adding just the minimal amount of information that would restore felicity. Consider example (5). By the time "the author" comes along, the file already contains a card i describing a book about Schubert that John read, but it does not contain any card describing an author. The simplest way to fix up the file in this case would seem to be by just adding a card j with the entry "is an author" on it. This would suffice for the Extended Novelty-Familiarity-Condition to be met and file change to proceed, but it is not what happens. Instead, one adds a card j with a somewhat more elaborate entry, viz., "is author of i." How do I know? Because (5) is clearly felt to convey more information than (6):

(6)   John read a book about Schubert and wrote to an author.

If accommodation in the case of (5) involved only the minimal entry "is an author" on the added card j, then the overall effect of (5) on the file should be the same as that of (6). But not any old pair of a book about Schubert and an author that is good enough to make (6) true will do for (5); it has to be a pair whose second member authored the first.

This example illustrates a general fact: When a new file card is introduced under accommodation, it has to be linked by crossreferences to some already-present file card(s). (Hence the term "bridging": the crossreferences form a "bridge" that connects the new discourse referent to the network of discourse referents that is already established.) This is particularly evident with the so-called "associative anaphoric" use which (5) exemplifies. It carries over to cases which are analogous to associative anaphoric uses, except that the role of the "indirect antecedent" (in (5), "a book about Schubert") is played by a contextually salient referent. For instance, I might say: "the price is not marked" in a situation where the two of us have been (silently) looking at a table that is for sale. I will then be understood as talking about the price of that table. Presumably, there has already been a file card i

---

[23] I am drawing on Hawkins (1978) and especially on Clark and Marshall (1981).

for the salient table, and my utterance now triggers accommodation, in the course of which a new card j is introduced that carries a crossreference to i.

In most of the "immediate" and "larger situation" uses, the newly added card is maybe not so much linked to another card as it is linked directly to the utterance situation. For instance , the card that gets introduced in response to "the dog" in example (3) above seems to get an entry like "is a dog close to here" (or "close to the addressee"), where "here" is the place of this utterance, and "the addressee" is the addressee of this utterance. Maybe every file should be taken to contain as one of its cards a card describing the utterance situation, so this sort of entry could be subsumed as a special case under entries with crossreferences to other cards. Be that as it may, a reference to the utterance situation counts as the kind of link that is required between an accommodated card and the previous file. There seems to be no exception to this requirement. Mere addition of a card without crossreferences, as it would happen with an indefinite instead of the novel definite, is never acceptable in accommodation.

Notice that some definites are built in such a way that the requirement for a crossreference will be automatically fulfilled, even if only the minimal entry to satisfy the Extended Novelty-Familiarity-Condition is put on the accommodated card. This is the case with the definites in (7):

(7)   John read [a book about Schubert]$_i$ and
      wrote to $\left\{ \begin{array}{l} \text{[its}_i \text{ author]}_j \\ \text{[the woman who had written it}_i\text{]}_j \end{array} \right\}$

Here, too, we must rely on accommodation (presuming that there was no previous introduction of a card j). But because the descriptive content of NP$_j$ explicitly contains a crossreference to card i, it follows from the Extended Novelty-Familiarity-Condition alone that the accommodated card must include such a crossreference. We need not appeal to an additional requirement for crossreferences here, but the examples are of course consistent without assumption that such a requirement is generally in force.

What explains the requirement that accommodated cards must be connected by bridges of this sort to the previous file? I do not know. An explanation would seem to be called for, but I cannot do more here than state the observation.

Accommodation triggered by novel definites can occur at any point in the evaluation of a complex utterance, and in particular, it can occur while one of the auxiliary steps in the evaluation of an operator headed formula is under way. This assumption turns out to provide the key to a treatment of "narrow-scope definites," which created problems for us in Chapter II (6.4). Consider (8) or (9):

(8)   [Every man who hates [his$_1$ father]$_2$]$_1$ is unhappy.
(9)   [Every man]$_1$ hates [his$_1$ father]$_2$.

Suppose we want to evaluate (9), starting from an initial file F. The first step is the calculation of F + [__ man]$_1$ which goes through only if 1 ∉ Dom(F). Suppose it does go through, and we get to move on to the second step, which is to be the calculation of ((F + [__ man]$_1$) + [his$_1$ father]$_2$) + [e$_1$ hates e$_2$]. Here we run into a problem: "his$_1$ father$_2$" is definite and thus should be entailed by F + [__ man]$_1$ to meet the Extended Novelty-Familiarity-Condition. It can be shown that this entailment could hold only if F happened to conform to (10).[24]

(10)   F entails that x$_2$ is every man's father.

In other words, F would have to contain the information that there is some one person that is every man's father. If F is indeed of that sort, we end up deriving a truth-condition for (9) according to which it says that every man hates that universal father. However, we ordinarily assume that different men have potentially different fathers, and so we ordinarily do not have a file F that meets (10). In the ordinary case, F will either contain no card number 2 at all, or else its card number 2 will describe something other than a universal father. We may ignore the second case, because it is unlikely that we would have disambiguated "his$_1$ father" in favor of index 2 in the first place if card 2 did not describe anything that fits the descriptive content of that NP. So let us assume that we have 2 ∉ Dom(F). This means we have to perform accommodation before we can incorporate "[his$_1$ father$_2$]" into the file.

The accommodation that is required consists of adding to F + [__ man]$_1$ a card number 2 on which it says: "is father of 1." (Note that this addition is a permissible accommodative measure in that it links the new card by a crossreference to an already present card.) In other words, we accommodate F + [__ man]$_1$ to (F + [__ man]$_1$) + [his$_1$ father]$_2$. Then we proceed as usual: we add — vacuously, as it were — [his$_1$ father]$_2$, and then add to that [e$_1$ hates

---

[24] Let us prove that the negation of (10) implies that F + [__ man]$_1$ does not entail [his$_1$ father]$_2$.

Assume that F does not entail that x$_2$ is every man's father. This means that there are a$_N$, w, x such that: a$_N$ ∈ SAT_w(F), and x is a man in w, and a$_2$ is not x's father in w. We have been assuming that 1 ∉ Dom(F). Therefore, by principle (B), b$_N$ ∈ SAT_w(F), where b$_N$ is the sequence that agrees with a$_N$ except that b$_1$ = x. Given that b$_1$ is a man in w, we may conclude that b$_N$ ∈ Sat(F + [__ man]$_1$). Given that b$_2$ (who equals a$_2$) is not b$_1$'s father in w, we have b$_N$ ∉ SAT((F + [__ man]$_1$) + [his$_1$ father]$_2$). Therefore, F + [__ man]$_1$ does not entail [his$_1$ father]$_2$.

e_2]. So the end result of the second step is $((F+ [\underline{\quad} man]_1) + [his_1 father]_2) +$ [e_1 hates e_2]. Now we are ready for the third step, which builds on the results of the first two steps and gives us:

$\text{Sat}(F + (9)) = \{a_N \in \text{Sat}(F) : \text{for every } b_N \widetilde{\text{Dom}(F)} a_N: \text{if } b_N \in \text{Sat}(F + [\underline{\quad}$ $man]_1), \text{then there exists } c_N \widetilde{\text{Dom}(F) \cup \{1\}} b_N \text{ such that } c_N \in \text{Sat}(((F + [\underline{\quad}$ $man]_1) + [his_1 father]_2) + [e_1 hates e_2]))\}.$

Given our assumption that neither 1 nor 2 are in $\text{Dom}(F)$, this turns out to amount to:

$\text{Sat}(F + (9)) = \{a_N \in \text{Sat}(F): \text{for every } b_1 \text{ such that } b_1 \text{ is a man, there is}$ a c_2 such that c_2 is b_1's father and b_1 hates $c_2\}$.

In other words: every man has a father that he hates. This is a reading where "his father" has in effect narrower scope than the universal NP.

In fact, it is just the reading that we would have arrived at in the theory of Chapter II by applying to the NP "[his_1 father]$_2$" the construal rule of Quantifier Indexing. That theory forced us to make a decision: does Quantifier Indexing apply to the definite in (9), or does it not? If not, we had no way of predicting the "narrow-scope" reading. If yes, we had to stipulate appropriate qualifications for Quantifier Indexing. In the present theory, we are better off. There is no construal rule of Quantifier Indexing, so we need not worry about, its conditions of application. We also do not need to stipulate any qualifications to the interpretation rule for the universal quantifier. That rule works as before, keeping fixed the assignments to the old indices, and varying the assignments to the new indices. Indices that correspond to accommodated cards are new, that is in the nature of accommodation. Hence, quantification will treat them like it treats all new indices: it will, in effect, bind them.

When we first abolished selection indices and had the interpretation rules for operators exploit the domain of the file instead, it seemed to make no empirical difference: There was a perfect correlation between indefiniteness and novelty-w.r.t.-the-file, and between definiteness and familiarity-w.r.t.-the-file, so we could not tell whether it was indefiniteness or novelty-w.r.t.-the-file that made an index get bound under quantification. Now that we have acknowledged accommodation, the correlation is no longer perfect: accommodated definites are novel-w.r.t.-the-file. So the question whether quantification is sensitive to indefiniteness or to novelty can be decided by looking at the behavior of definites that involve accommodation. As it turns out, they decide in favor of the choice we have in fact made, i.e., that quantifiers bind whatever indices are novel, indefiniteness plays only an indirect role, insofar as it is a sufficient condition for novelty.

I need not go into the evaluation of example (8), as it is essentially analogous to (9). Since this time the definite triggering accommodation is in the restrictive term of the universal quantifier, rather than in the nuclear scope, accommodation occurs in the course of the first step. As a result, we get a reading in which the universal quantifier ranges in effect over pairs of a man and his father.

In our discussion of "narrow-scope" definites in Chapter II, we also had an example with a definite description under negation:

(10)   Mary didn't have lunch with the king of France (because France doesn't have a king).

This, too, can be given a consistent reading by assuming accommodation. Suppose we start from a file F with $i \notin \text{Dom}(F)$, where i is the index of "the king of France," In order to calculate the file change that the negated sentence brings about, we must first calculate the auxiliary file F + [Mary had lunch with [the king of France]$_i$], which is only possible if we add a card i (describing a king of France) by way of accommodation. We then arrive at:

Sat(F + [Mary had lunch with [the king of France]$_i$]) = $\{a_N \in \text{Sat}(F):$ a_i is king of France and Mary had lunch with $a_i\}$.

The final calculation according to the interpretation rule for negation will then give us:

Sat(F + [not[Mary had lunch with [the king of France]$_i$]]) = $\{a_N \in$ $\text{Sat}(F)$: there is no b_i such that b_i is king of France and Mary had lunch with $b_i\}$.

This is the "narrow-scope" reading we were after.

As we have just seen, our theory enables us to predict narrow-scope readings for definites under negation, and this is a good thing because such readings do occur, e.g., with example (10). One might object, however, that the theory generates such readings all too easily. It is a fact, after all, that such readings arise only when they are forced, as in (10) by the explicit denial in the "because"-clause. If the "because" clause were missing, then we would ceteris paribus understand "Mary didn't have lunch with the king of France" as implying that France has a king, even if this is news to us. In other words, even if the file F that obtains prior to the utterance fails to contain a card i describing a king of France, we prefer a reading under which the utterance is about the king of France and denies that Mary had lunch with him. Can our theory predict such a reading at all, and if yes, can it predict that it should be ceteris paribus preferred?

If we think about these questions and try to meet the challenge they present, we are forced to be more specific than we have been about the way in which accommodation works, especially when it occurs in the midst of evaluating an operator-headed formula. Operator-headed formulas typically require *tentative* file changes as part of their evaluation, i.e., they require that we keep earlier stages of the file in our memory until we have reached the final step. This raises a question about accommodation: if we have to accommodate at a point where we are holding several files in memory, will it affect just the one we are currently operating on, or the other ones, too? We may ask the same question in a different way: When file change blocks because of a prima facie violation of felicity conditions and we are forced to back up, accommodate, and start over, *how far back* do we go?

For example, suppose we are to evaluate a formula of the form "not ϕ" w.r.t. an initial file F. To do this, we have to calculate F + ϕ while keeping F in memory. Suppose now that ϕ is prima facie infelicitous w.r.t. F, so we have to accommodate, i.e., we have to turn F into a file F′ from which we can proceed to F′ + ϕ without violating felicity conditions. There are in principle two ways of doing that: (a) we calculate F′ + ϕ instead of F + ϕ, but leave the copy of F that we are holding in memory untouched; (b) we substitute F′ for F throughout, i.e., we assume F′ instead of F both as a basis for the evaluation of ϕ and as a basis for any further steps that were supposed to depend on F. If we proceed according to (a), then the final product of the evaluation of "not ϕ" will be (11).

(11) $\text{Sat}(F + [\text{not } \phi]) = \{a_N \in \text{Sat}(F): \text{there is no } b_N \; \widetilde{\text{Dom}(F)} \; a_N \text{ such that } b_N \in \text{Sat}(F' + \phi)\}$;

$\text{Dom}(F + [\text{not } \phi]) = \text{Dom}(F)$.

If, on the other hand, we follow (b), then the final product is (12).

(12) $\text{Sat}(F + [\text{not } \phi]) = \{a_N \in \text{Sat}(F'): \text{there is no } b_N \; \widetilde{\text{Dom}(F')} \; a_N \text{ such that } b_N \in \text{Sat}(F' + \phi)\}$;

$\text{Dom}(F + [\text{not } \phi]) = \text{Dom}(F')$.

(11) and (12) are by no means equivalent.

Suppose ϕ is "Mary had lunch with [the king of France]$_i$" and F is such that $i \notin \text{Dom}(F)$. This is the situation we had with example (10). When we went through the evaluation of this example above, we accommodated according to option (a) and thus ended up with a result that was an instance of (11). F′, the fixed-up version of F, was in this case such that $\text{Dom}(F') = \text{Dom}(F) \cup \{i\}$ and $\text{Sat}(F') = \{a_N \in \text{Sat}(F): a_i \text{ is king of France}\}$. We did not realize that we had another option, namely (b). Suppose we now had chosen

(b). F′ would still have been as just described. But the final result would have been an instance of (12), namely the following:

(13) $\text{Sat}(F + [\text{not}[\text{Mary had lunch with } [\text{the king of France}]_i]]) = \{a_N: a_N \in \text{Sat}(F), \text{ and } a_i \text{ is king of France, and it is not the case that Mary had lunch with } a_i\}.$

$\text{Dom}(F+ [\text{not}[\text{Mary had lunch with } [\text{the king of France}]_i]]) + \text{Dom}(F) \cup \{i\}.$

This corresponds to a reading where the existence of a king of France is implied.

As we noted, the latter sort of reading is ceteris paribus preferred. This suggests the hypothesis that accommodations of type (b), which amount to an across-the-board adjustment of F to F′, are generally preferred over accommodations of type (a), with more localized effects. Whether this hypothesis is borne out for a wider range of examples, and how one might explain the existence of such a preference, these remain open questions.[25] Supposing that the preference exists, it is clear why example (10), when it includes the "because"-clause as originally assumed, must still be evaluated according to the non-preferred option (a). For imagine we had followed (b) and had thus arrived at the result file described in (13), which entails that France has a king. This entailment is explicitly contradicted in the "because"-clause, and so the only way of making sense of the utterance after all would have been to start over and follow (a) instead of (b).

Obviously, there is a lot more to be said about this kind of example in particular and about accommodation in general. Still, I hope to have said enough to justify the conclusion that a theory along the lines of this dissertation need not simply capitulate when confronted with the sort of "narrow-scope" readings that have often been considered compelling evidence in favor of a quantificational analysis of definite descriptions.

5.3   On the distribution of pronominal versus non-pronominal definites

Throughout this dissertation, I have been concerned with giving necessary and sufficient conditions for the use of a definite NP as opposed to an indefinite NP. Yet, it is well known that not all kinds of definite NPs are alike in their felicity conditions. For instance, we often find speech contexts in which a definite description is acceptable, but a personal pronoun is not:

---

[25] See Heim (in preparation) [Heim 1983b].

(11) John has a cat and a dog. $\left\{ \begin{array}{c} ? \text{ Its} \\ \text{The cat's} \end{array} \right\}$ name is Felix.

(12) John is married. $\left\{ \begin{array}{c} ? \text{ She} \\ \text{His wife} \end{array} \right\}$ is nice.

The inacceptability of "it" in (11) has an obvious reason: There are two equally possible antecedents, giving rise to an unresolvable ambiguity. The pronoun does not have enough descriptive content to limit the choice of antecedents to one, whereas the definite description "the cat" does. Not any definite description would have, though: "the pet" would have been no better than "it" in (11). So the contrast in acceptability that (11) illustrates is not really a matter of pronouns versus non-pronouns. It is rather a matter of more versus less descriptive content. That the amount of descriptive content should become relevant in this way, and that it should lead to unacceptability of pronouns in certain cases, follows naturally from conversational principles and does not call for any new provisions in our theory.

(12) seems to be a different case: The problem is not that there are two competing file cards with which "she" might be associated, but that there are prima facie none at all. While the definite description can here be saved by accommodation, this option is unavailable for the pronoun. Why? One might again seek an explanation in the lack of descriptive content. Accommodation, as you recall, always requires that a bridge of crossreferences be built between the new card and the preexisting file. In the case of "his wife," the descriptive content fully determines the bridge that is to be built, whereas in the case of "she," some guessing is required, However, I am not convinced that this suffices to explain the contrast in acceptability. For notice that the example becomes noticeably better (though not as impeccable as the variant with "his wife") when "she" is replaced by "the woman," which has hardly more descriptive content and calls for just as much guessing as to which bridge is intended. The impression that one gathers from the difference between "she" and "the woman" in the context of (12) is that pronouns obey a constraint of their own that restricts their use even when there is no danger of ambiguity: For a pronominal definite NP_i to be felicitous w.r.t. a file F, i must be a *prominent* element of Dom(F). Call this the "Prominence Condition."

What does "prominent" mean? Let us assume that a file is not just an amorphous bunch of cards, but is organized in such a way that a small number of cards enjoy a privileged place, "on the top of the file," so to speak. These are always the cards that the file clerk had to handle most recently, i.e., that were most recently introduced or updated. The number of those cards, a small and constantly shifting subset of the domain of the file, are the

prominent elements of the domain, and only they can appear as the indices of pronouns. So anaphoric pronouns will have to have antecedents in the recent previous discourse, and deictic pronouns will have to refer to objects that still are, or very recently have been, making themselves salient.

How does the Prominence Condition apply to occurrences of a pronoun NP_i where there is no card number i in the file at all unless accommodation creates one? Certainly, such cases will be prima facie violations of the Prominence Condition, since a number i that is not in Dom(F) is a fortiori not a prominent member of Dom(F). But can this violation be remedied by accommodation? Can accommodation apply in such a way that the accommodated card not only satisfies the Extended Novelty-Familiarity-Condition, but moreover counts as prominent? What we have seen so far suggests that it cannot. The badness of the pronoun in example (12) indicates that a card that owes its very existence to accommodation is not prominent at the point at which the NP that triggers its accommodation is evaluated. (It would of course become prominent as a result of the successful evaluation of that NP, if indeed it were successful.) Prominence, it would seem, is not accommodable.

But there are exceptions: we do find pronouns used in examples which we can only analyze in terms of accommodation. The exceptions fall into several types, some of them perhaps marginal[26], but others not at all. Among the latter are the famous "paycheck"-sentences of Karttunen (1969) and other so-called "pronouns or laziness."[27] Consider the underlined pronoun in (13).

(13) [Every wise man]$_1$ gives [his$_1$ paycheck]$_2$ to his$_1$ wife. John$_3$ gives $\underline{it}_4$ to his$_3$ mistress.

The first sentence of (13) is basically like our example (9), "Every man hates his father." Its evaluation includes a step at which card number 2, with the entry "is a paycheck of 1," is introduced by accommodation.

When we are done with the evaluation of the first sentence, however, the file includes neither 1 nor 2, since both were introduced only tentatively, as part of the auxiliary files that were used to interpret the universal quantifier. As we proceed to the second sentence, we first find (or maybe accommodate) the card number 3 for "John$_3$." Then we come to "it$_4$" and fail to find a card number 4. (Notice that things would be no different if we had chosen the indexing "it$_2$." Card 2 is no more in the file than card 4 at the relevant

---

[26] Among the marginal ones, I would count an example that I have heard from Barbara Partee: "I glued two pieces of paper together, and *it* flew."

[27] Partee (1970) discusses a wealth of examples that are of relevance here.

moment, as was just pointed out. In other words, "it" could at best be coindexed vacuously with "his paycheck".) So what do we do? Apparently, we simply create a card number 4, with the entry "is a paycheck of 3," and the Prominence Condition does not seem to reject this accommodated card as unsuitable for the pronoun "it_4." This is how we get the intended reading, which could not come about in another way.

Why can the pronoun in (13) be saved by accommodation when the one in (12) could not? Apparently, it plays a role that the "it" in (13) occurs so shortly after "his paycheck." This means that the evaluation of the latter, which included accommodative introduction of a card for x_1's paycheck, is still "fresh in the file clerk's memory," so to speak, and he just needs to do the same thing again when he faces the "it." Presumably, the analogy to the recent accommodation triggered by "his paycheck" accounts for the ease with which "it" is interpreted: both times the accommodated card receives an entry of the form "is a paycheck of i," where i is different in the two cases, but is the index of the respective subject in both. It is not surprising, from a psychological point of view, that accommodation should go through most smoothly when it is sanctioned by a recent parallel precedent in this way.

But where does this leave the Prominence Condition? Even if we are correct in assuming a bias in favor of parallelism, we should still expect the pronoun in (13) to be ruled out because it violates the requirement for prominence. I am not entirely sure how to reconcile the condition with the facts here. Rather than permitting for exceptions, I prefer to think that the Prominence Condition can, so to speak, be satisfied "by proxy": card 4, while not itself prominent in the required sense, is introduced in analogy to the prominent card 2. The prominence of the latter seems to suffice to license the former as a suitable discourse referent for a pronoun. One might want to be more precise about the conditions under which file cards can be prominent by proxy. For the time being, this vague suggestion is all I have to offer.

The "paycheck"-pronoun in (13) has a definite "antecedent." (I use "antecedent" here in an extended sense that does not imply non-vacuous coindexing. In this extended sense, a pronoun's antecedent may be the NP whose evaluation sets the precedent for the analogous evaluation of the pronoun and serves as its proxy for the purpose of the Prominence Condition.) It appears as though *in*definite antecedents do not license "paycheck"-pronouns as easily. Consider the contrast in (14):

(14) a. Almost every man sold the he owned. John rented *it* out.
      b. Almost every man sold a house he owned. John rented *it* out.

In (14a), "it" can be the house John owned, just like "it" could be John's paycheck in (13). But in (14b), "it" cannot be interpreted at all. Why

should this be so? I can think of two possible reasons, which probably both play a role. One reason could be that the analogy between a regular (non-accommodative) card introduction and an introduction by accommodation is not as strong as the analogy between two introductions by accommodation. This could mean that the evaluation of "a house he owned" in the first sentence of (14b) just does not count as the sort of precedent that would license accommodation in response to the pronoun, even though the entry that the card to be accommodated would have to carry (viz.,"is a house that i owned") is fully parallel to the entry of the card that was introduced for the antecedent "a house he owned."

A second reason for the unacceptability of the pronoun in (14b) might have to do with the message that the speaker conveys by choosing the indefinite "a house he owned," rather than using the definite "the house he owned" and thereby relying on accommodation. By refraining from the definite, the speaker signals that it is not a matter of course that an arbitrary man can be associated with a house he owns. But then it will not be treated as a matter of course that John can be associated with a house he owns either, and this will interfere with the intended disambiguation of the pronoun "it." Again, these are merely vague speculations and there is more to be said before the contrast in (14) can be considered explained.

While (14b) is no good, some cases of pronouns where accommodation appears to be licensed by an indefinite antecedent are not all that bad. Examples like (15) are rejected by some speakers, but for others they are much better than (14b).

(15)   Almost every man sold a house. One of them got as much as 300,000 dollars for *it*.

And perfectly good is (16):

(16)   Every motel room has a copy of the Bible in it. In this room, *it* was hidden under a pile of TV Guides.

(15) and (16) differ crucially from (14b) in the following respect: In (14b), it is understood that John is not one of the majority of men who are claimed to have sold houses they own. In (15), "one of them" means: one of the men that sold a house, and in (16), "this room" is understood to be one of the motel rooms that were just said to generally have Bibles in them.

We may speculate that accommodation is ceteris paribus more acceptable when it adds no new information to the file, i.e., when the original file and the file with the accommodated card added to it have identical satisfaction sets (in all worlds). The accommodations required to interpret the pronouns in (15) and (16) are of this information-preserving sort: Given that the file

already entails that x_i (where i is the index of "one of them") is a man who sold a house, its satisfaction conditions will stay the same when a card j with the entry "is a house that i sold" is added. Similarly for (16). The precise role of this facilitating factor in accommodation (if indeed it is one) remains mysterious, as does the fact that (16) is better than (15).

Let us finally look at some examples that I noted as problematic in Chapter II, section 7, when I compared Karttunen's predictions about the "lifespan of discourse referents" to the corresponding predictions of my theory.

(17)   John wants to catch *a fish*. Can you see *it* from here?
(18)   John knew that Mary had *a car*. But he had never seen *it*.
(19)   Mary wants to marry *a rich man*. *He* must be a banker.

In each case, we limit our attention to readings where the operator ("wants," "knew," and "wants," respectively) takes scope over the indefinite in the first sentence. This implies that none of the three examples will permit non-vacuous coindexing between the indefinite and the underlined pronoun. The relevant facts are that in both (18) and (19), but not in (17), we perceive a reading in which the pronoun has the indefinite as its "antecedent," in some intuitive sense. Karttunen observed, as I reported above, that the relevant difference between (17) and (18) is that "know," unlike "want," is factive, and the relevant difference between (17) and (19) is that the second sentence in (19), but not in (17), contains another modal operator (here: "must").

Since non-vacuous coindexing is not an option in these examples, the only possible relation between the indefinite and the pronoun is one where the pronoun triggers an accommodation for which the evaluation of the indefinite is somehow a licensing precedent. Our task, then, is to explain why this sort of accommodation is possible in (18) and (19), but not in (17).

Both "want" and "know" involve universal quantification over a certain set of possible worlds, in one case the set of worlds that conform to the subject's desires, in the other case the set of worlds which are compatible with the subject's knowledge. The worlds that conform to someone's desires often do not include the actual world; but the worlds that conform to someone's knowledge always do, that is what makes "know" factive. The difference between (17) and (18) can now be seen as analogous to the difference between (14b) and (15). In (17), the first sentence says that every world w that fits John's desires has in it a fish that John catches in w. The second sentence is about the actual world w_0. In order to get the intended interpretation for "it," we would have to accommodate a card describing a fish that John catches in w_0. But we are well aware that the existence of such a fish in w_0 is not at all a matter of course, and this awareness apparently makes

us resist the required accommodation. In (18), on the other hand, the first sentence says that every world w that is compatible with what John knows is such that Mary has a car in w. The second sentence is about w_0. To interpret the "it," we have to accommodate a card describing a car that Mary has in w_0. This accommodation adds no new information, because w_0 is known to be among John's knowledge-worlds and hence known to contain a car that Mary owns in it. Presumably, the facilitating influence of information-preservingness is here strong enough to license accommodation.

In (19), both the first and the second sentence involve universal quantification over worlds, The operator "must" in the second sentence is understood to express necessity in view of what Mary wants, and so it ranges over the same set of worlds as the "want"-operator in the first sentence. In evaluating the second sentence, we introduce a card i with an entry like "is a world that conforms to Mary's desires." That is the first step of the familiar three-step sequence that applies with operators just like we have seen it apply with quantifiers. (Notice that we are now using cards that describe worlds, along with cards for individuals.) In the second step, we have to deal with the pronoun "he," which we do by accommodating a card j with the entry "is a rich man that Mary marries in world i." Like in the case of example (18), this accommodation does not bring any new information into the file. We already know from the first sentence that worlds conforming to Mary's desires have rich men that Mary marries in them, and so this must hold for any world that satisfies card i. If information-preserving accommodation was licensed in (18), we expect it to be licensed in (19) as well, as indeed it appears to be.

In this section, we have explored, albeit rather inconclusively, the conditions under which pronouns can be interpreted by accommodation. The examples of pronouns for which accommodation must be appealed to are famous examples, because they are all counterexamples to the hypothesis that anaphoric relations can always be analyzed as either variable-binding or coreference. In their capacity as counterexamples to that hypothesis, they have received much attention in the literature, and the reader may wonder how the present discussion relates to the conclusions about them that have been reached elsewhere.

A recurrent idea, which was eventually made fully precise by Cooper (1979), has been the idea that pronouns must sometimes be analyzed as definite descriptions. I agree with this in a certain sense, although the relation between pronouns and definite descriptions appears in a somewhat different light in my theory. I did not take the standard Russellian analysis of definite descriptions and extend it to pronouns. Rather, I started with a move in the opposite direction: I tried to give definite descriptions an

analysis of the sort that is standardly applied to pronouns, viz., a variable-analysis. So I analyzed definite descriptions as pronouns, rather than vice versa. This attempt met with certain prima facie limitations, which could eventually be overcome by providing for a mechanism of accommodation. Once accommodation had been made available, there were in principle two possibilities:

(a) Accommodation could have been limited strictly to definite descriptions, to the exclusion of pronouns. That would have been a way of regaining most of the empirical predictions that result from the contrasting standard analyses for definite descriptions and for pronouns (with the exception of predictions concerning donkey-anaphora).

(b) There was the alternative of permitting accommodation for definites of any kind. By choosing this alternative, we have come to predict that some pronouns will show just the definite-description-like behavior that Cooper's proposal is designed to capture.

Notice that, to the extent that Cooper's and my views are in agreement, we are also facing some of the same problems: If pronouns are not intrinsically different from definite descriptions, then how come their distribution is in effect so much more limited? Some of the limitations follow in an obvious way from the lack of descriptive content, but we have seen in this section that there are subtle contrasts without obvious explanations. As far as I can see, we would have faced analogous problems if our starting point had been an analysis of pronouns as (context-dependent) Russellian definite descriptions.

6   Conclusion: Motivation for the file change model of semantics

In this dissertation, one analysis has been presented two times. Both Chapter II and the current chapter are elaborations of a variable analysis for indefinite and definite noun phrases, but they employ different theoretical frameworks. Supposing that the common analysis is basically on the right track, are we in a position to choose one of the frameworks over the other? I have indicated at various points in the course of this chapter that the approach in terms of files and file change offers advantages over the more conventional approach of the previous chapter. Drawing on some of these earlier remarks, and adding a few more, I will try to make a final plea for the semantic theory that has started to emerge in this chapter and that I have dubbed "file change semantics."

First, let me briefly remind you of some facts about the theory. Files were introduced as an additional level of analysis to intervene between language

and the world — more precisely: between the linguistic structures that the grammar generates and their referents and truth values. Unlike levels of representation that are strictly grammar-internal (e.g., logical form), files encode information from nonlinguistic sources (perception, permanently stored knowledge) along with information contributed by the linguistic structures. The rules and principles that are involved in the generation of files are accordingly diverse. We have tried to isolate some of them and have grouped them into three components: (a) interpretation rules (such as our rules (I) to (IV) in section 4.4), (b) felicity conditions, and (c) accommodation rules. The so-called felicity conditions are concerned with the way in which the logical form to be interpreted and the file to be updated mutually constrain each other. The theory of definiteness, as embodied in the Extended Novelty-Familiarity-Condition, belongs in this component.

I will now present four more or less tentative arguments in favor of file change semantics. The last one relates directly to the theory of definiteness and to considerations that this chapter has focused on. The other three concern points that I think deserve further attention in future research.

(1) The first argument is based on the privileged status that file change semantics accords to conjunction, as opposed to other logical operations, like negation, disjunction,[28] and quantification. Conjunction is the simplest: a formula that expresses the conjunction of p and q is evaluated by first evaluating p and then evaluating q. Negation or disjunction are on a par with quantification in requiring the construction of auxiliary files while previous files must be held in memory, and are hence less simple. From the point of view of a conventional semantics, by contrast, all of them are truth-functional connectives, hence alike in complexity. Empirical evidence favors the asymmetry that file change semantics suggests: not only does it appear, from a psychological point of view, that conjoined structures are the easiest to understand, but there is also significance in the banal fact that conjunction is the operation which is automatically understood when formulas are juxtaposed without any explicit connective.

---

[28] An interpretation rule for disjunction would have to look more or less like this:

(V) Let ϕ be a molecular formula consisting of "or" and the two formulas ψ^1 and ψ^2. Then: $\text{SAT}(F + \phi) = \{a_N \in \text{SAT}(F): \text{there is some } b_N \overset{\frown}{\text{DOM}(F)} a_N \text{ such that either } b_N \in \text{SAT}(F + \psi^1) \text{ or } b_N \in \text{SAT}(F + \psi^2)\}$;
$\text{DOM}(F + \phi) = \text{DOM}(F)$.

This argument will require some further elaboration to become clear and convincing, especially with respect to what it presumes about the relation of the present competence-model of semantics to a processing model. At this point, it is not so much an argument, but a suggestion of where to look for one. Note that, if tenable, it will be an argument that indiscriminately favors theories which interpret expressions in terms of "context change" potentials of some sort over theories that primarily assign truth conditions. It is independent of how in particular the changing aspect of the "context" is construed, whether in terms of the file notion or in terms of a similar notion of common ground.

(2) A second argument, and perhaps one of the strongest to be found, is that file change semantics offers a solution to the projection problem for presuppositions. This would take another whole dissertation to establish convincingly, but we have seen enough to warrant optimism. l am almost tempted to say that the essentials of file change semantics can be deduced from the projection problem as soon as that problem is formulated in the right terms.

As a starting point, take the familiar idea that presuppositions are felicity conditions, which means that the presupposition of an expression constrains the range of contexts in which that expression can be evaluated. Now suppose we observe, e.g., that p calls for a context that entails q, but "if r, p" merely calls for a context that entails s (a weaker proposition than q). Looking for an explanation of our observations, we ask ourselves: What rule will yield the value s when applied to the values of p, q, r, and the meaning of "if ... then"? That is a standard way of formulating the projection problem, but it is not the only possible formulation. The observations just mentioned could equally well have led us to ask as follows: How come a context that entails merely s becomes good enough for an utterance of p if you say "if r" first? From this question, it is not far to the answer: Apparently, saying "if r" changes a context that used to entail s into one that entails q.

Once the projection problem is rephrased in this way, it turns into the problem of specifying, in a systematic fashion, the context-changing effects of all the meaningful expressions of the language, and of specifying them in such a way that the observations about presupposition projection are accounted for. It turns out that not any old way of construing the notion of "context" will lead to success in this task. If presupposition projection in quantified sentences is to be covered, then it must be acknowledged that the smallest context-changing units of utterance are not always closed formulas, but may contain free

variables. Since such open formulas do not express propositions, context-change cannot generally consist of simply adding propositions, and contexts can hence not be characterized in terms of mere sets of propositions. Characterizing them in terms of files, however, enables us to assign context-change potentials to units as small as open formulas. The projection problem thus provides both a general argument for context-change semantics and a specific argument for characterizing contexts as files. Even if file change semantics were just a theory of presupposition projection that had to be added on top of a component that assigns truth-conditions in the conventional way, it might stand a chance against competing theories of presupposition projection. If it can moreover replace conventional truth-conditional interpretation, as this dissertation has argued, then so much the better.

(3) A third advantage of file change semantics emerges when we consider the interaction between rules of accommodation and rules of interpretation. Both accommodation and interpretation are updating operations on the file that are in some sense dictated by the meaning of the expression under evaluation. The difference is that interpretation rules evaluate the pure semantic content, in the narrowest sense of "semantic," whereas accommodation rules, whose application depends upon the demands of felicity conditions, deal with a more "pragmatic" dimension of meaning (with the information that an expression "implicates," as some would say). This division of labor is not particular to file change semantics, but mirrors the two-dimensional organization of semantics that has been adopted elsewhere.[29]

What is particular to the file change model, however, is that accommodation can apply at any point during the evaluation of a complex logical form and can thereby affect the outcome of the next interpretation rule that applies. For instance, we saw in section 5.2 how a definite description in the scope of a quantifier may trigger accommodation during the construction of an auxiliary file, and how the interpretation of the quantified structure will then be based on the accommodated auxiliary file, leading in effect to substantially different truth-conditions than would have resulted without accommodation. This would not have been possible in a theory where truth-conditions, rather than context-change potentials, were assigned directly by the recursive interpretation rules. We have derived considerable benefit from permitting accommodation to be "interspersed" with interpretation in this way:

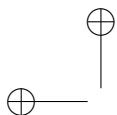[29] See in particular Karttunen and Peters (1975; 1979).

We have been able to maintain a uniform logical analysis of all definites as variable-NPs and a very simple account of how quantifiers select the indices they bind, while still predicting the "narrow-scope" readings of complex definites.

This seems to me to add plausibility to the idea of a level of analysis like our files, which are, so to speak, generated simultaneously by rules of diverse components. The precise implications of this idea, and what it means for the relation between semantics and pragmatics, remain to be explored.

(4) Last, but not least, file change semantics has provided us with a framework in which we could state the theory of definiteness in a single principle, the Extended Novelty-Familiarity-Condition. Recall the conclusion to Chapter II, where we identified three seemingly independent contrasts that correlated with the definite-indefinite distinction: (i) only indefinites were subject to the Novelty Condition, (ii) only indefinites underwent Quantifier Indexing, and (iii) only definites presupposed their descriptive content. From our present point of view, (i) and (iii) were incomplete precursors of the Extended Novelty-Familiarity-Condition. (i), the Novelty Condition, was a prohibition against anaphoric indefinites, which now follows from a more general prohibition against linking indefinites to familiar file cards. (iii) used to be the only stipulation about definiteness that was concerned with felicity conditions. Now the theory of definiteness is entirely a matter of felicity conditions, and (iii), rephrased as the Descriptive-Content-Condition, corresponds to one of two clauses in the Extended Novelty-Familiarity-Condition that define two related aspects of the notion of familiarity.

Assumption (ii) has disappeared from the theory of definiteness altogether. Its purpose had been to secure, through the mediation of selection indices, that quantifiers bind just the indices of indefinites. This is now secured by the Extended Novelty-Familiarity-Condition through the mediation of the domain of the file, given a revised treatment of quantification under which interpretation rules for quantifiers refer to file-domains instead of selection indices. The elimination of (ii) has both simplified the construal component and led to the welcome prediction that quantifiers may bind definites in certain cases involving accommodation.

Of course, the Extended Novelty-Familiarity-Condition is a more complex stipulation than anyone of the three conditions it replaces. Nevertheless, it amounts to a simpler theory of the definite-indefinite
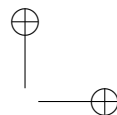
contrast and makes it appear less mysterious that definites and indefinites should differ in just the ways they do.
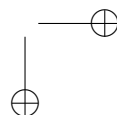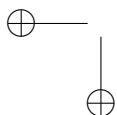
References

Annear, S. 1968. 'Relative Clauses and Conjunctions'. In *Working Papers in Linguistics No. 1*. Ohio State University, pp. 80–109.

Baker, C. L. 1966. *Definiteness and Indefiniteness in English*. Master's thesis, University of Illinois.

Ballmer, T. 1979. 'Context change, truth, and competence'. In *Semantics from Different Points of View*. Berlin: Springer.

Barwise, J. and Cooper, R. 1981. 'Generalized Quantifiers and Natural Language'. *Linguistics and Philosophy*, 4: 159–219.

Carlson, G. 1977. *Reference to Kinds in English*. Ph.D. thesis, University of Massachusetts, Amherst. (also, New York: Garlan 1979.).

Chastain, C. 1975. 'Reference and Context'. In *Minnesota Studies in the Philosophy of Science, vol. VII - Language, Mind, and Knowledge*. Minneapolis: University of Minnesota Press, pp. 194–269.

Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Christophersen, P. 1939. *The Articles – A study of their Theory and Use in English*. Copenhagen: Munksgaard.

Clark, H. 1977. 'Bridging'. In Johnson-Laird, P. and Wason, P. (eds.) *Thinking*. Cambridge: Cambridge University Press, pp. 411–420.

Clark, H. and Haviland, S. 1977. 'Comprehension and the Given-New contrast'. In Freedle, R. (ed.) *Discourse Production and Comprehension*. Norwood, N.J.: Ablex, pp. 1–40.

Clark, H. and Marshall, C. 1981. 'Definite reference and mutual knowledge'. In Joshi, A., Webber, B. and Sag, I. (eds.) *Elements of Discourse Understanding*. Cambridge University Press, pp. 10–63.

Cooper, R. 1979. 'The interpretation of pronouns'. In Heny, F. and Schnelle, H. (eds.) *Syntax and Semantics 10 – Selections from the Thirtd Groningen Round Table*. Academic Press, pp. 61–92.

Egli, U. 1979. 'The Stoic concept of anaphora'. In v. Stechow, A. (ed.) *Semantics from Different Points of View*. Springer, pp. 266–283.

Evans, G. 1977. 'Pronouns, Quantifiers, and Relative Clauses'. *Canadian Journal of Philosophy*, 7: 467–536.

———1980. 'Pronouns'. *Linguistic Theory*, 11: 337–362.

Fodor, J. D. and Sag, I. 1982. 'Referential and Quantificational Indefinites'. *Linguistics and Philosophy*, 5, 355-398.

Gazdar, G. 1979. *Pragmatics: Implicature, Presupposition and Logical form*. New York: Academic Press.

Geach, P. 1962. *Reference and Generality*. Ithaca, NY: Cornell University Press.

Grice, P. 1967. 'William James Lectures on Logic and Conversation'. In Davidson, D. and Harman, G. (eds.) *The Logic of Grammar*. Encino, CA: Dickenson, pp. 64–75.

Grosz, B. 1977. 'The Representation and Use of Focus in a System for Understanding Dialogs'. SRI Report.

Hamblin, C. 1971. 'Mathematical models of dialogue'. *Theoria*, 37: 130–155.

Hawkins, J. 1978. *Definiteness and Indefiniteness*. London: Croom Helm.

Heim, I. 1983a. 'File Change Semantics and the Familiarity Theory of Definiteness'. In Bäuerle, Rainer, Schwarze, Christoph and von Stechow, Arnim (eds.) *Meaning, Use, and the Interpretation of Language*. Berlin: Walter de Gruyter, pp. 164–189.

Heim, Irene 1983b. 'On the Projection Problem for Presuppositions'. In Barlow, M., Flickinger, D. and Wescoat, M. (eds.) *WCCFL 2: Second Annual West Coast Conference on Formal Linguistics*. Reprinted in *Partee and Portner (2002)*.

Higginbotham, J. 1980. 'Pronouns and Bound Variables'. *Lingusitic Inquiry*, 11: 679–708.

Hintikka, J. and Carlson, L. 1979. 'Conditionals, generic quantifiers, and other applications to subgames'. In *Formal Semantics and Pragmatics for Natural Languages*. Dordrecht: Reidel.

Jackson, F. 1979. 'On assertion and indicative conditionals'. *Philosophical Review*, 88.

Jespersen, O. 1949. *A Modern English Grammar on Historical Principles Vol. VII (completed by N. Haislund)*. Copenhagen: Munksgaard.

Kamp, H. 1981. 'A Theory of Truth and Semantic Representation'. In Groenendijk, Jeroen, Janssen, Theo M.V. and Stokhof, Martin (eds.) *Formal Methods in the Study of Language, Part 1*. Mathematical Centre Tracts, pp. 277–322.

Kaplan, D. 1970. 'What is Russell's theory of definite descriptions?' In Yourgrau, W. and Beck, A. (eds.) *Physics, Logic, and History*. New York: Plenum, pp. 277–288.

Karttunen, L. 1968a. 'What do Referential Indices Refer To?' RAND Corporation Report No. P3854.

———1968b. 'What Makes Definite Noun Phrases Definite?' RAND Corporation Report No. P3871.

———1969. 'Pronouns and Variables'. *CLS*, 5.

———1973. 'Presuppositions of Compound Sentences'. *Linguistic Inquiry*, 4: 169–193.

———1976. 'Discourse Referents'. In McCawley, J. (ed.) *Syntax and Semantics 7*. Academic Press, pp. 363–385.

Karttunen, L. and Peters, S. 1975. 'Conventional Implicature in Montague Grammar'. In *Proceedings of the Berkeley Linguistics Society I*. pp. 226–278.

———1979. 'Conventional Implicature'. *Syntax and Semantics*, 11: 1–56.

Katz, J. 1972. *Semantic Theory*. New York: Harper & Row.

Kratzer, A. 1978. 'Semantik der Rede - Kontexttheorie, Modalwörter, Konditionalsätze.' Könighstein Ts.: Scriptor.

———1981a. 'Blurred Conditionals'. In Klein, W. and Levelt, W. (eds.) *Crossing the Boundaries in Linguistics*. Reidel, pp. 201–209.

———1981b. 'The Notional Category of Modality'. In Eikmeyer, H.-J. and Rieser, H. (eds.) *Words, Worlds, and Contexts, New Approaches to Word Semantics*. Berlin: Walter de Gruyter, pp. 38–74.

Kripke, S. 1977. 'Speaker's Reference and Semantic Reference'. *Midwest Studies in Philosophy*, 2, 1: 255–276.

Langendoen, D. T. and Savin, H. B. 1971. 'The Projection Problem for Presuppositions'. In Fillmore, Charles and Langendoen, D. Terence (eds.) *Studies in Linguistic Semantics*. New York: Holt, Reinhardt and Winston.

Lasnik, H. 1976. 'Remarks on coreference'. *Linguistic Analysis*, 2: 1–22.

Lawler, J. 1973. *Studies in English generics*. University of Michigan Papers in Linguistics 1, Ann Arbor, MI: University of Michigan Press.

Lerner, J. and Zimmermann, T. 1987. 'Mehrdimensiona1e Semantik: Die Präsupposition lind die Kontextabhängigkeit von 'nur''. In *Working paper no. 50 of the SFB 99*. Germany: University of Konstanz: University of Konstanz.

Lewis, D. 1973. *Counterfactuals*. Oxford: Blackwell.

———1975. 'Adverbs of Quantification'. In Keenan, E. (ed.) *Formal Semantics of Natural Language*. Cambridge: Cambridge University Press, pp. 3–15.

———1979. 'Scorekeeping in a Language Game'. *Journal of Philosophical Logic*, 8: 339–359.

May, R. 1977. *The Grammar of Quantification*. Ph.D. thesis, Massachussetts Institute of Technology, Cambridge, Massachusetts.

Montague, R. 1974. 'The proper treatment of quantification in ordinary English'. In Thomason, R. (ed.) *Formal Philosophy*. New Haven: Yale University Press, pp. 247–270.

Nunberg, G. and Pan, C. 1975. 'Inferring quantification in generic sentences'. *CLS*, 11: 412–422.

Parsons, T. 1978. 'Pronouns as paraphrases'. Unpublished ms.

Partee, B. 1970. 'Opacity, Coreference, and Pronouns'. *Synthese*, 21: 359–385.

Partee, B. H. and Portner, P. (eds.) 2002. *Formal Semantics - The Essential Readings*. Oxford: Blackwell Publishing.

Postal, P. 1969. 'Anaphoric Islands'. *CLS*, 5: 205–239.

———1970. 'On coreferential complement subject deletion'. *Linguistic Inquiry*, 1: 439–500.

Prince, E. 1981. 'On the inferencing of indefinite-<u>this</u> NPs'. In A. Joshi, I. Sag and Webber, B. (eds.) *Linguistic Structure and Discourse Setting*. Cambridge: Cambridge University Press.

Quine, W.V.O. 1960. *Word and Object*. Cambridge, Massachusetts: MIT Press.

Reinhart, T. 1976. *The Syntactic Domain of Anaphora*. Ph.D. thesis, Massachussetts Institute of Technology, Cambridge, MA.

Rodman, R. 1976. 'Scope phenomena, movement transformations, and relative clauses'. In Partee, B. (ed.) *Montague Grammar*. New York: Academic Press, pp. 165–176.

Russell, B. 1905. 'On Denoting'. *Mind*, 14: 479–493.

———1919. *Introduction to Mathematical Philosophy*. London: Allen and Unwin.

Smaby, R. 1979. 'Ambiguous coreference with quantifiers'. In Guenther, F. and Schmidt, S. (eds.) *Formal Semantics and Pragmatics for Natural Languages*. Dordrecht: Reidel, pp. 37–75.

Stalnaker, R. C. 1968. 'A Theory of Conditionals'. In Rescher, N. (ed.) *Studies in Logical Theory*. Oxford: Blackwell, pp. 98–112.

———1973. 'Presuppositions'. *Journal of Philosophical Logic*, 2, 4: 447–457.

———1974. 'Pragmatic Presuppositions'. In Munitz, Milton and Unger, Peter (eds.) *Semantics and Philosophy: Essays*. New York University Press.

———1979. 'Assertion'. In Cole, P. (ed.) *Syntax and Semantics 9*. New York: New York Academic Press, pp. 315–32.

Stockwell, R., Schachter, P. and Partee, B. 1973. *The Major Syntactic Structures of English*. New York: Holt, Rinehart and Winston.

Strawson, P. F. 1952. *Introduction to Logical Theory*. London: Methuen.

Tarski, A. 1936. 'Der Wahrheitsbegriffin den formalisierten Sprachen'. *Studia Philosophica*, 1.

von Stechow, Arnim 1981. 'Topic, Focus and Local Relevance'. In Klein, W. and Levelt, W. (eds.) *Crossing the Boundaries in Linguistics*. Dordrecht: Reidel, pp. 95–130.

Wasow, T. 1972. *Anaphoric Relations in English*. Ph.D. thesis, Massachussetts Institute of Technology, Gent. Published as *Anaphora in Generative Grammar*. E. Story-Scientia (1979).

Webber, B. 1978. 'A Formal Approach to Discourse Anaphora'. Ph.D. Dissertation, BBN Report No. 3761.