Resolution of quantifier scope ambiguities

Howard S. Kurtzman
Division of Neuroscience and Behavioral Science, National Institute of Mental Health, 5600 Fishers Lane, Rockville, MD 20857, USA

Maryellen C. MacDonald
Hedco Neuroscience Building, University of Southern California, Los Angeles, CA 90089-2520, USA

Received March 31, 1992, final version accepted May 25, 1993

Abstract

Various processing principles have been suggested to be governing the resolution of quantifier scope ambiguities in sentences such as Every kid climbed a tree. This paper investigates structural principles, that is, those which refer to the syntactic or semantic positions of the quantified phrases. To test these principles, the preferred interpretations for three grammatical constructions were determined in a task in which participants made speeded judgments of whether a sentence following a doubly quantified sentence was a reasonable discourse continuation of the quantified sentence. The observed preferences cannot be explained by any single structural principle, but point instead to the interaction of several principles. Contrary to many proposals, there is little or no effect of a principle that assigns scope according to the linear order of the phrases. The interaction of principles suggests that alternative interpretations of the ambiguity may be initially considered in parallel, followed by selection of the single interpretation that best satisfies the principles. These results are discussed in relation to theories of ambiguity resolution at other levels of linguistic representation.

The order of authors was determined by the alphabet; both contributed equally to this work. We thank the following for helpful discussion and suggestions: John Bowers, Gennaro Chierchia, Robin Clark, Janet Dean Fodor, Irene Heim, Philip Johnson-Laird, Mario Montalbetti, Massimo Poesio, Barry Schein, and audiences at the 10th Cognitive Science Society Conference (1988), the 3rd CUNY Conference on Sentence Processing (1990), and the 1st SALT (Semantics and Linguistic Theory) Conference (1991). We thank Jeffrey Kishner, Paula O’Day, and James Wagner for assistance in running subjects. Kurtzman was supported by NICHD grant R01 HD20824 at Cornell University, and MacDonald was supported by BRSG2 S07 RR07047-23 at MIT, and by NSF grant DBS-9120415 and a grant from the Zumberge Research Innovation Fund at USC. No official endorsement of this paper by the National Institute of Mental Health is intended or should be inferred.

Correspondence to: H.S. Kurtzman, e-mail: kwh@cu.nih.gov or to M.C. MacDonald, e-mail: mcm@gizmo.usc.edu.
Introduction

Ambiguity resolution has been a central topic within the study of sentence comprehension since the rise of contemporary psycholinguistics in the 1960s (Carlson & Tanenhaus, 1989; Fodor, Bever, & Garrett, 1974; Small, Cottrell & Tanenhaus, 1988). Humans' ability to resolve ambiguities is remarkable; ambiguity is pervasive in everyday language, but we are usually able to adopt a single appropriate interpretation quickly and unconsciously. Ambiguity resolution is also interesting because the manner in which the human comprehension system responds to ambiguities can be viewed as reflecting basic operating principles of the system. Most research to date has examined ambiguities that arise at the lexical level (e.g., the various lexical items corresponding to bat) and at the syntactic level (e.g., the two phrase structures underlying Mary said that the tree fell yesterday).

The present paper extends this research by examining a common type of ambiguity that arises at a logico-semantic level and which so far has received little attention in psycholinguistics. This ambiguity concerns quantifier scope, and it can arise when two or more noun phrases (NPs) in a sentence contain a quantifier term such as every, some, a, many, or a few in the determiner position. For example, the meaning of the sentence in (1) may correspond to either of the logical structures represented in (2) and (3):

(1) Every kid climbed a tree.
(2) (\forall x)(\exists y) (x \text{ is a kid} \& y \text{ is a tree} \& x \text{ climbed } y)
   \[\text{read as: "For every kid } x \text{, there is a tree } y \text{, such that } x \text{ climbed } y\”]\n(3) (\exists y)(\forall x) (x \text{ is a kid} \& y \text{ is a tree} \& x \text{ climbed } y)
   \[\text{read as: "There is a tree } y \text{, such that for every kid } x \text{, } x \text{ climbed } y\”]\n
The two logical structures differ only in the relative scope orderings of their quantified terms. According to the interpretation in (3), in which a tree has "wide scope" and every kid has "narrow scope," there is one particular tree that all of the kids climbed. According to (2), in which every kid has wide scope and a tree has narrow scope, each kid did not necessarily climb the same tree: rather, there may be more than one tree with different kids climbing each tree (and perhaps, though not necessarily, just one kid per tree).

Another ambiguous sentence is produced when the positions of the quantifier terms in (1) are reversed, as shown in (4). The two possible logical structures are shown in (5) and (6).

(4) A kid climbed every tree.
(5) (\exists x)(\forall y) (x \text{ is a kid} \& y \text{ is a tree} \& x \text{ climbed } y)
   \[\text{read as: "There is a kid } x \text{, such that for every tree } y \text{, } x \text{ climbed } y\”]\n
(6) $(\forall y)(\exists x) (x \text{ is a kid} \& \ y \text{ is a tree} \& \ x \text{ climbed } y)$

[read as: “For every tree y, there is a kid x, such that x climbed y’’]

According to (5), in which a kid has wide scope, there is one particular kid who climbed all the trees. According to (6), in which the term expressing every tree has wide scope, every tree got climbed by some kid or another, though not necessarily the same kid for each tree.

Though they have received little attention in psycholinguistics, quantifier terms have been of substantial interest to linguists and philosophers of language, largely because, as Chierchia and McConnell-Ginet (1990) point out, quantifier terms contribute to the expressive capacity of natural language by making possible the expression of generalizations about variously sized sets of individuals. Most linguists agree that quantifier scope is represented at a semantic level that is distinct from lexical and syntactic levels of linguistic analysis (though Katz, 1980, has argued that information associated with scope is encoded at a pragmatic level). Various linguistic and computational approaches have been proposed to describe the semantic level at which quantifier scope is represented: standard logical formulae (Hobbs & Sheiber, 1987; Kempson & Cormack, 1981; Woods, 1977), a quasi-syntactic level of logical form (May, 1977, 1985), indices on syntactic structure (Williams, 1986), model theory (Cooper, 1983), mental models (Fodor, 1982), discourse-representation models (Kamp, 1981), and propositional networks (Anderson, 1983).

For psycholinguists, the interest in quantifier scope ambiguities lies in the operations that comprehenders employ to resolve the ambiguities. As with other ambiguities, comprehenders often report informally that one interpretation of a quantifier scope ambiguity is preferred. By “preferred” interpretation, we mean the one that is consciously arrived at first or is otherwise initially considered to be more likely or appropriate. For (1), English speakers almost always report that the interpretation in (2) is preferred, while for (4), they often report that the interpretation in (5) is preferred. Analyses of similar preferences for lexical and syntactic ambiguities have resulted in the development of specific models of comprehension processes (Carlson & Tanenhaus, 1989; Small et al., 1988) and lead to experiments investigating the time course of ambiguity resolution. It would appear that the same approach could be pursued for quantifier scope preferences. Because ambiguities of quantifier scope are not the result of any individual words being ambiguous or of the word string corresponding to more than one (surface or deep) phrase structure, the process of resolving these ambiguities may be different from lexical or syntactic ambiguity resolution processes. Discovering the similarities and differences in ambiguity resolution across the various levels of linguistic representation can lead to a broader theory of the language processor.

Previous research suggests that there are several types of information that
might constrain which interpretation a comprehender adopts for a scope ambiguous sentence. These types include (at least): (1) structural positions of the quantified phrases in the syntactic or semantic representation of the sentence; (2) lexical biases of particular quantifier terms to take wide or narrow scope; and (3) real-world knowledge such that one interpretation might be more plausible in context than the other(s). Our view is that all of these types of information play a role in quantifier scope disambiguation, though not necessarily at the same time. A complete theory of scope disambiguation will have to specify how and when these types of information are utilized, just as theories of lexical and syntactic disambiguation must do for the roles of similar constraints and for the time course in which they are applied.

Our goal in this paper is more modest, however. Rather than develop a complete theory, we focus upon the role of structural position. We show that this type of information by itself has a substantial impact on interpretation preferences and provide an initial account of precisely how it is used. We also peripherally address the role of lexical biases, as it is impossible to study quantifier scope without incurring the effects of the particular quantifier terms appearing in the stimulus sentences. In this work we do not investigate the role of real-world knowledge, but rather control for it by maintaining high plausibility for all of our stimuli.

Processing principles

In investigating ambiguities and the types of information that affect their resolution, psycholinguists have found it fruitful to assume that interpretation preferences are due largely to the comprehension system operating according to principles that lead to one interpretation being constructed earlier or being assigned a higher ranking. For cases in which both interpretations of an ambiguity are fully well formed, the most likely source for the broad range of preferences would indeed seem to be comprehension-processing principles. This approach has in fact been adopted, at least implicitly, by various researchers concerned with quantifier scope preferences. They have suggested principles (reviewed below) which they claim govern the processing responsible for the initial assignment of a scope ordering to an ambiguity. In investigating processing in terms of “principles”, we are adopting a terminology that is common in this literature and in some work on ambiguity resolution of other levels of linguistic representation (e.g., Frazier & Fodor, 1978), but it should be noted that our and others’ view is that these principles arise from the particular architecture of the processing system, so that identification of principles is an intermediate, rather than final goal of a processing theory.

The proposed principles for quantifier scope processing have generally been
expressed in broad, functional terms that are compatible with any of the specific formats for semantic representation referred to above. The most commonly discussed principle is that of linear order, according to which the preferred scope ordering of quantified phrases corresponds to the left-to-right ordering of the phrases in the surface structure of the sentence (Bunt, 1985; Fodor, 1982; Johnson-Laird, 1969; Kroch, 1975; Lakoff, 1971; VanLehn, 1978). This principle does not refer directly to hierarchical syntactic relations among quantified phrases or to any aspects of semantic structure. For active sentences of the form of (1) and (4), the principle predicts that the preferred interpretation will be that in which the first-encountered (or leftward) phrase – Every kid in (1) and A kid in (4) – has wide scope. For the passive counterparts, as in (7) and (8), the principle again predicts a preference for wide scope for the leftward phrase – A tree in (7) and Every tree in (8).

(7) A tree was climbed by every kid.
(8) Every tree was climbed by a kid.

Other theorists have proposed that the positions of the quantified phrases in the syntactic structure determine preferred scope ordering. For example, Iooup (1975) argued for a hierarchy of grammatical relations as a primary determinant. According to this hierarchy, a phrase occupying a surface- or deep-structure subject position preferably has wide scope over a phrase in a non-subject position. This preference is stronger if the phrase occupies the subject at both surface and deep structure, as does the leftward phrase in active sentences like (1) and (4). If one quantified phrase occupies the surface subject and the other occupies the deep subject, as in passives like (7) and (8), there may be no sharp preference. In more recent linguistic theory, the deep-structure subject is often described as the external argument (e.g., Levin & Rappaport, 1986). Therefore, we will reformulate Iooup’s claims for preferences for deep and surface subjects to have wide scope as two principles: a surface subject principle and an external argument principle. For actives, the two principles agree on which phrase to assign wide scope, while for passives the principles conflict.

Reinhart (1983, especially pp. 197–198) also seeks to account for scope preferences in terms of syntactic positions. Rather than referring to specific positions, as Iooup does, she refers to the c-command relations among quantified phrases. Reinhart claims that one phrase will preferably have wide scope over another if the former c-commands the latter at surface structure (and not vice versa).

1 A simple definition of c-command is given by Chomsky (1986) (slightly paraphrasing): Node A c-commands Node B iff A does not dominate B and every Node C that dominates A dominates B. Variations on this definition have been suggested (e.g., Aoun & Sportiche, 1983; Reinhart, 1983), but none would result in alterations of any of the predictions derived from the c-command principle which are discussed here. VanLehn (1978) also suggests a version of the c-command principle.
Reinhart’s grammatical theory that seeks to account for the cases in which scope ambiguity does and does not arise in sentences with multiple quantified phrases. For all of the actives and passives exemplified in (1), (4), (7), and (8) above, this principle predicts that the leftward phrase will preferably have wide scope, since as surface subject, it c-commands the other constituents of the sentence.

The fact that Loup’s and Reinhart’s principles are syntactically based does not contradict the view that scope ambiguities arise at a semantic level. Since semantic representations are constructed on the basis of syntactic representations, these syntactically based principles can be understood as guiding that construction during sentence processing. Other theorists, however, have argued that preferences should be explained by principles that refer only to the semantic level. Both Kempson and Cormack (1981) and May (1985) claim, within their different approaches, that the phrase expressing the topic, as opposed to comment, of the sentence preferably has wide scope, and Katz (1980) makes a related claim—we will call this view the **topic principle**. The precise definition of topic is unclear (see Kempson & Cormack, 1981, and Reinhart, 1982, for discussion), but for our purposes it is sufficient to characterize it broadly as the constituent expressing “what the sentence is about”. The topic is canonically expressed by the surface subject or the left-most surface constituent in English—and so it predicts that for both the actives and the passives the leftward phrase will preferably have wide scope. Crucially on this view, however, it is the phrase’s semantic status as topic, rather than its syntactic position, which determines that it preferably has wide scope.

Other semantic approaches could be developed. One possibility, which we explore below, is that preferred scope ordering is determined by the thematic hierarchy (Grimshaw, 1990; Jackendoff, 1972). That is, a phrase that expresses a thematic role that is higher on the hierarchy than that expressed by another phrase will have wide scope over that phrase. According to this **thematic hierarchy principle**, and given the hierarchy relations Agent > Experiencer > Theme, a phrase that expresses an Agent or Experiencer will preferably have wide scope over a phrase expressing a Theme, and this preference might be stronger for a phrase expressing an Agent than one expressing an Experiencer. It is not unreasonable to suggest that quantifier scope is related to thematic roles, since on many linguistic theories the principles that govern them apply at the same or related levels of structure (e.g., Halvorsen, 1983; May, 1985). This principle predicts that for actives such as (1) and (4), in which the leftward phrase expresses an Agent and the rightward phrase a Theme, the leftward phrase will preferably have wide scope. For the passive counterparts in (7) and (8), it predicts that the rightward phrase will preferably have wide scope.

In sum, all of the principles predict that for actives the leftward phrase will preferably have wide scope. As for passives, the principles vary in their predictions: the linear order, surface subject, c-command, and topic principles
predict that the leftward phrase will preferably have wide scope, while the external argument and thematic hierarchy principles predict that the rightward phrase will preferably have wide scope.

Evidence

The interpretation preference intuitions previously reported in the literature do not unequivocally support the predictions of any of these processing principles. Although it is usually reported that the leftward phrase preferably has wide scope in active sentences, for sentences of the form of (4), in which the a-phrase precedes the every-phrase, it is sometimes reported that the rightward phrase preferably has wide scope (e.g., Moran, 1988). As for passives, there is no general consensus on which interpretation is preferred. For example, Johnson-Laird (1969) claims that the leftward phrase preferably has wide scope, while Ioup (1975) and Kempson and Cormack (1981, citing Keenan, 1976), claim that there is not a strong preference for either phrase to have wide scope.

It is possible that these disagreements over interpretation preferences stem from each researcher relying on judgments for just one or a few sentences that are different from those used by other researchers, without sufficient control for lexical, syntactic and semantic properties of the sentences. VanLehn (1978) did examine a larger set of sentences (121 sentences), but various properties of his sentences were also not adequately controlled: the sentences' alternative scope interpretations were unequal in plausibility, they contained other lexical or syntactic ambiguities, and they contained non-obligatory material extraneous to the scope ambiguity. Any of these characteristics could have influenced judgments and masked the effects of scope processing principles.

A second concern with this work is the nature of the measure: the activity of providing conscious judgments calls attention to the sentences as potentially ambiguous and is not strictly time-limited. Informants are thus able to reflect
upon a sentence and upon their judgment before reporting; they may also rely on special strategies or be biased by their judgments on other sentences. Informal judgments may thus be influenced by factors that do not normally contribute to sentence comprehension and may not yield appropriate data for theories of processing.

In lexical and syntactic ambiguity work, researchers have developed alternative measures that produce data that reflect initial comprehension processing more directly. These measures include cross-modal priming (Swinney, 1979; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982); eye-movement tracking (Frazier & Rayner, 1982); self-paced reading (Just, Carpenter, & Woolley, 1982); and speeded grammaticality judgment (Clifton, Frazier, & Connine, 1984; Kurtzman, Crawford, & Nychis-Florence, 1991). By contrast, only one such alternative to informal judgments has been utilized in earlier published research on scope processing. Johnson-Laird (1969) and Micham, Catlin, Van Derveer, and Love-land (1980) presented participants with a scope-ambiguous sentence and had them judge the truth values, relative to the sentence, of variously filled matrices in which the rows and columns stood for the quantified phrases. From the pattern of responses, they inferred which interpretation had been assigned to the sentence. However, this task still suffers from the same limitations as the elicitation of conscious judgments: it directs attention to the ambiguities and is not time-restricted. For example, Johnson-Laird (1969) reports that nearly half the participants mentioned some aspect of the ambiguity of the sentences during testing, and the mean time to judge each series of 10 matrices per sentence exceeded 40 s. Further, the task was quite difficult, as indicated by the fact that Micham et al. had to drop about 25% of their participants from the data analysis because they were incorrect on control items; this difficulty increases the likelihood that participants often adopted special strategies.

The experiments reported here use a method that taps initial comprehension in a more direct fashion. Participants read an ambiguous quantifier sentence followed by a continuation sentence that is a reasonable discourse continuation of the quantifier sentence under just one of its interpretations. Participants judge whether the continuation sentence is indeed a reasonable continuation. For example, for (1), the continuation sentence would be either the plural The trees were full of apples (reasonable only for the interpretation in (2)) or the singular The tree was full of apples (most reasonable for the interpretation in (3)). Participants are instructed to read the sentences and to make their judgments as rapidly as possible. The pattern of responses to the continuation sentences indicates what interpretation(s) had been preferably assigned to the ambiguous quantifier sentences. Control stimuli in which the quantifier sentence is unambiguous, as well as varied unambiguous filler stimuli, are also tested. The task thus does not call attention to the ambiguity of any of the sentences, and participants are encouraged to answer based on their first impressions.
Performance on the task reveals which interpretations of the ambiguity are initially available for integration with a succeeding sentence in a discourse. The task is not strictly an “on-line” one, since it does not tap into the interpretations that might be only briefly available on the way to adopting particular interpretations, which has been possible for lexical and (perhaps to a lesser extent) syntactic ambiguities. Therefore, in testing the various preference principles described above, we construe them as predicting which interpretations are initially adopted rather than merely initially computed. This is consistent with how the researchers who have proposed the principles have construed them. Nonetheless, the task is significantly more immediate than the tasks previously used to investigate scope ambiguity and hence should provide more valid and reliable data. (Time to read a continuation sentence and respond to it was generally between 2.0 and 3.5 s in the experiments below.)

Experiment 1 investigates active sentences, Experiment 2 investigates passive sentences, and Experiments 3 and 4 investigate scope ambiguities within complex NPs.

EXPERIMENT 1: ACTIVE SENTENCES

This experiment investigated ambiguous active sentences containing the quantifier terms every and a. Sentences were tested with these terms in both orders, as in examples (1) and (4). These particular terms were used because they are high in frequency and are natural-sounding in the first sentence of a discourse. Further, their meanings are simpler than those of some other quantifier terms such as many and a few.3

We varied the order of quantifier terms in order to control for their lexical biases. Vendler (1967), Ioup (1975), Kroch (1975), VanLehn (1978), and Fodor (1982) all report informal judgments suggesting that quantifier terms differ in their biases to preferably take wide scope. They primarily discuss each, every, and all, whose biases toward wide scope decline in that order. Ioup tentatively suggests that a falls between every and all.4 Although this study is primarily

3Hornstein (1984) refers to every and a as type II quantifiers (see also Aoun & Hornstein, 1985). In particular, their scope is generally bounded within a tensed clause and they show effects of the Empty Category Principle (assuming a framework in which these quantified phrases undergo movement, as originally proposed in May, 1977).

Some is commonly used in other researchers’ examples. It was not used in these experiments because it has more wide-ranging uses than every and a: some co-occurs with singular and plural count nouns and with mass nouns, while every and a co-occur only with singular count nouns.

4More recent work by Fodor and Sag (1982) leads to the suggestion that the tendency of a toward wide scope is even lower than all. Fodor and Sag point to a non-quantifier, referential use of a, in which the a-phrase does not enter into scope relations but whose meaning is often readily confusable with that of the wide-scope quantifier use. If this non-quantifier use is excluded, the tendency for quantifier a to take wide scope is likely lower than Ioup suggested.
concerned with the effects of the quantified phrases’ positions within the sentence (as addressed by the processing principles), rather than effects of the quantifier terms themselves, the inclusion of the quantifier order variable allows us to make preliminary observations of any interaction between structural and lexical factors.

Method

Participants

College students in the Pittsburgh, Pennsylvania, and Ithaca, New York, areas volunteered to serve as paid participants. A total of 48 participants were tested, half in Pittsburgh and half in Ithaca. All participants in this and subsequent experiments were native speakers of American English; none participated in more than one experiment.

Materials

Thirty-two sets of two-sentence experimental passages were created, each with eight different versions. A complete example is shown in Table 1, and all stimuli are contained in Appendix A. Each set consisted of a 5- or 6-word quantifier sentence followed by a 6-word continuation sentence. Three independent variables—ambiguity, quantifier order, and verb—were manipulated in the quantifier sentence, and a fourth variable—interpretation—was manipulated in the continuation sentence.

Table 1. Example stimuli, Experiment 1

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Ambiguous, WS1</th>
<th>Ambiguous, WS2</th>
<th>Unambiguous, WS1</th>
<th>Unambiguous, WS2</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Every . . . a&quot; quantifier order</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantifier</td>
<td>Every kid climbed a tree.</td>
<td>Every kid climbed a tree.</td>
<td>Every kid climbed a different tree.</td>
<td>Every kid climbed the same tree.</td>
</tr>
<tr>
<td>Continuation</td>
<td>The trees were full of apples.</td>
<td>The tree was full of apples.</td>
<td>The trees were full of apples.</td>
<td>The tree was full of apples.</td>
</tr>
</tbody>
</table>

| "A . . . every" quantifier order |
| Quantifier | A kid climbed every tree. | A kid climbed every tree. | The same kid climbed every tree. | A different kid climbed every tree. |
| Continuation | The kid was full of energy. | The kids were full of energy. | The kid was full of energy. | The kids were full of energy. |

Note: WS1 indicates an interpretation with wide scope on the first quantified noun phrase; WS2 indicates wide scope on the second quantified noun phrase. The fourth variable, verb, was manipulated between-items. Only an action verb example is shown here; an example of a perception verb sentence is Every tourist saw a statue.
The ambiguity variable had two levels: ambiguous and unambiguous. In the ambiguous sentences, if the $a$-phrase is assigned wide scope over the every-phrase, then the $a$-phrase refers to a single entity. However, if the every-phrase is assigned wide scope over the $a$-phrase, then the $a$-phrase typically refers to multiple entities. For all ambiguous quantifier sentences, both interpretations were equally highly plausible, even with no specific context provided, as judged by the authors and colleagues.

The unambiguous quantifier sentences provide a baseline comparison for the ambiguous condition. The unambiguous sentences contained every and, instead of $a$, either the same or a different. A phrase containing the same refers to a single entity, while a phrase containing a different typically refers to multiple entities, but these unambiguous sentences are not, precisely, disambiguated versions of the ambiguous sentences. Such disambiguated versions would be much more lengthy and complex than the ambiguous sentences. Thus, Every kid climbed a different tree means that there was a one-to-one mapping of kids to trees, but Every kid climbed a tree, with wide scope on Every kid, does not specify this or any particular mapping. Both the same and a different also have deictic uses, as when the same tree or a different tree refers to a tree that is the same as or different from one already indicated in the context.5

Quantifier order varied according to the order of the quantified terms in the quantifier sentence. We refer to the orders as “Every . . . a” and “A . . . every” for both the ambiguous quantifier sentences and their unambiguous counterparts.

The verb variable was manipulated between passages. In 16 passages, the quantifier sentence’s verb was an action verb, so that the leftward subject NP expressed an agent and the rightward object NP a theme, and in the other 16 it was a perception verb so that the leftward subject NP expressed an experiencer and the rightward object NP a theme. The content words of the quantifier sentences with action and perception verbs were matched for length and frequency. Due to the paucity of high-frequency perception verbs, some of them were repeated across passages: three of the verbs appeared in two passages and one verb appeared in three passages.

In the continuation sentence, the noun in the subject NP was the same as the noun within the quantifier sentence’s NP containing $a$, the same, or a different. This subject NP and its agreeing verb varied between singular and plural. The singular continuation sentence implied that the $a$-phrase in the ambiguous

5When used deictically, both the same tree and a different tree refer to a single entity. We assumed that it was unlikely that participants would adopt the deictic readings, since no preceding context is provided for the quantifier sentences. The results below support this assumption; in particular, participants interpreted phrases containing a different as referring to multiple entities. Another possible difference between the ambiguous and unambiguous quantifier sentences is that in the latter there might not even be a relation of scope between the quantified phrases (see alternative analyses in Clark & Keenan, 1986, and Moltmann, 1992).
quantifier sentence (or the phrase containing the same in its unambiguous counterpart) referred to a single entity, whereas the plural continuation sentence implied multiple reference of the NP containing a or a different in the quantifier sentence. Each ambiguous quantifier sentence could be paired with either continuation sentence. When the continuation sentence is compatible with the quantifier sentence under the reading in which wide scope is assigned to the leftward (linearly first) phrase, the interpretation variable is labeled as WSl (for wide scope on the 1st phrase). When the continuation sentence is compatible with wide scope assignment to the rightward (second) phrase, the interpretation is labeled WS2. Note that the WSl/WS2 distinction is orthogonal to the singular/plural distinction.

Each unambiguous quantifier sentence was paired only with the continuation sentence that was compatible with it. The passage’s interpretation is given the same WSl or WS2 label as the passage comprised of an ambiguous quantifier sentence and the same continuation sentence.

In addition, 88 filler passages were created, with varied lengths and syntactic and semantic structures (including both actives and passives), and some contained quantifier terms. For 40 of these, the second sentence was a reasonable continuation of the first, while for the other 48 the second sentence was not. Thus, of the total of 120 passages that each participant judged (32 experimental plus 88 filler passages), 60% contained a second sentence that was compatible with the first (all versions of the experimental passages were compatible).

Twenty of the incompatible fillers contained a first sentence that was identical in structure, but not content words, to the unambiguous quantifier sentences in the experimental passages. The second sentence was structurally identical to the continuation sentences in those passages, except that its singular/plural value was mismatched to the first sentence, so that the correct response to the continuation sentence was “No”; for example, Every player admired a different coach. The coach was an experienced veteran. These fillers, which we term similar incompatibles, were included in order to detect a response bias to judge as compatible any passages with the general form of the experimental passages. Unlike the experimental passages, there was only one version of each of these fillers (with five each corresponding to the four types of unambiguous quantifier sentences in the experimental passages); all participants were presented with the same versions.

Eight stimulus files were created including all fillers and one member of each of the 32 sets of experimental passages (four members exemplifying each of the eight types).

Procedure

Each participant was tested individually, sitting in front of a CRT, with thumbs resting on the space bar of the terminal and index fingers resting on Yes and No
keys. Pressing the space bar on the terminal brought the first sentence of the passage onto the middle row of the screen. Participants read the sentence and then pressed the space bar again to remove the first sentence and present the continuation sentence, beginning in the same position as the first. Subjects used the Yes and No keys to respond whether the second sentence “made sense” and was a “natural continuation” of the first. It was stressed to participants that they should read both sentences quickly but make sure to understand them, and that they should respond as quickly and accurately as possible. Progression through the trials was self-paced.

All participants were first presented 20 practice passages (similar to the fillers). Then they were presented with one of the stimulus files (six participants per file). The passages in a stimulus file were presented in a different random order to each participant. The session lasted approximately 30 min. In post-session interviews, few participants reported noticing any ambiguities, and none reported noticing more than five. All reported that the task was at least moderately easy.

Results

The extent to which continuation sentences were judged compatible with the quantifier sentences is shown in Fig. 1. For the ambiguous quantifier sentences, participants judged the WS1 continuation sentences to be compatible more frequently than the WS2 continuation sentences, consistent with most intuitions reported in the literature. For the unambiguous quantifier sentences, participants usually accurately judged all continuation sentences to be compatible (recall that only compatible continuation sentences were presented for unambiguous trials). This pattern produced a robust Ambiguity × Interpretation interaction, \( \text{min } F'(1, 68) = 30.15, \ p < .001 \) (here and elsewhere, \( \text{min } F' \) values are given if
significant, otherwise $F_1$ and $F_2$ are given). The lack of an interpretation preference for the unambiguous quantifier sentences suggests that there was not any general plausibility difference between the WS1 and WS2 interpretations.

Further examination of ambiguous conditions revealed several interesting effects. Though the preference for the WS1 interpretation was highly preferred for these items ($\text{min } F'(1, 65) = 41.01, p < .001$), the WS2 continuation sentences were not always rejected. In particular, they were judged to be compatible more often following “Every...a” sentences with perception verbs than following other quantifier sentence types, producing a significant Quantifier order $\times$ Interpretation $\times$ Verb interaction, $\text{min } F'(1, 80) = 4.24, p < .05$.

To explore this interaction further, separate analyses were conducted on the responses to each of the continuation sentence types in the ambiguous condition. For just the WS1 continuations, there were more Yes responses following “A...every” than “Every...a” quantifier sentences, producing a marginal effect of quantifier order, $F_1(1, 47) = 3.43, p < .10, F_2(1, 30) = 6.27, p < .05$. For just the WS2 continuation sentences, there were more Yes responses following “Every...a” than “A...every” quantifier sentences, producing a significant effect of quantifier order, $\text{min } F'(1, 75) = 12.07, p < .01$. Also, for these WS2 continuation sentences, there was a significant Quantifier order $\times$ Verb interaction such that Yes responses to this continuation were very low, except in one case: the combination of the “Every...a” order and perception verb, $\text{min } F'(1, 76) = 5.43, p < .05$. Because the verb factor was manipulated between items, we considered the possibility that other changes across the verb factor (e.g., different nouns in the sentences) might have produced this effect. No obvious differences in the items were apparent to us, although we noted that while the first NP was animate in all items, the second NP was sometimes inanimate (for both some action verb and some perception verb items). However, a post hoc analysis of this animacy factor revealed no reliable effects or interactions with any other factors.

The strong interpretation preferences demonstrated in the ambiguous condition were not present in the unambiguous condition, so that there was no reliable effect of interpretation. There was a small effect of quantifier order, with slightly more Yes responses to continuation sentences following “Every...a” than “A...every” quantifier sentences ($87.5\%$ vs. $76.6\%$, respectively), $\text{min } F'(1, 54) = 4.15, p < .05$. There was also a significant interaction of Quantifier order $\times$ Interpretation, $F_1(1, 47) = 6.21, p < .05, F_2(1, 30) = 6.26, p < .05$. This interaction was the result of a slight preference for the WS1 interpretation in the “A...every” order but no clear preference for either interpretation in the “Every...a” order.

Given the overwhelming preference for WS1 interpretations in the ambiguous condition, we might ask whether the ambiguous sentences were being treated exactly like the unambiguous sentences with WS1 interpretation. The answer is that the strong preference in the ambiguous condition ($77.9\%$ Yes responses to WS1 interpretations) was not quite comparable to the unambiguous WS1
condition (84.1% Yes responses), producing a marginally significant effect of ambiguity, $F_1(1, 47) = 6.13, p < .05$, $F_2(1, 30) = 3.27, p < .10$. There was also a marginally significant interaction of Ambiguity $\times$ Quantifier order, $F_1(1, 47) = 4.02, p < .10$, $F_2(1, 30) = 6.80, p < .05$, which reflects that the main effect of ambiguity was due primarily to the relatively weaker WS1 preference for the ambiguous “Every... a” quantifier sentences.

For the similar incompatible fillers, participants correctly responded No on 85.0% of the trials, which is comparable to the overall accuracy level of 82.3% on the unambiguous conditions in the experimental passages, for which the correct answer was Yes. These results indicate that subjects could make these judgments, and there was no particular bias to respond Yes or No.

Discussion

The results of Experiment 1 indicate that for active scope ambiguous sentences there is a general preference for the interpretation in which the leftward quantified phrase has wide scope. This preference corresponds to most informal judgments and is predicted by all the processing principles described above. In addition, however, the present data show that this preference is modulated by two other effects. One effect is the order of quantifiers: the preference is stronger for the “A... every” order than for the “Every... a” order. This result was manifested by both greater availability of the preferred interpretation and lesser availability of the non-preferred interpretation for the “A... every” order than for the “Every... a” order. The other effect is the type of verb: the non-preferred interpretation is more available for “Every... a” sentences with perception verbs.

The effect of the order of quantifiers is not likely to be due to any lexical biases of the quantifiers to take wide scope. As mentioned, if there is a difference between every and a, then every probably has the greater bias toward wide scope, so that the strongest WS1 preferences should have been observed in the “Every... a” condition. The opposite result was observed, however.

We suggest that the reason for the sharper preference for the “A... every” order is that when an a-phrase in the surface subject or topic of a sentence is received it is immediately interpreted as referring to just a single entity; we call this the single reference principle (Fodor, 1982, makes a related claim). This principle is a plausible one, both because single reference is simpler to represent than multiple reference (cf. Murphy, 1984; Crain & Steedman, 1985) and because single reference is always possible and often obligatory (as when a is the only quantifier; e.g., A kid climbed trees). Only upon subsequent reception of the every-phrase, which can take wide scope over the a-phrase, would it be reasonable to treat the a-phrase as potentially having multiple reference. But to switch from single to potential multiple reference involves a complex alteration of
the semantic representation (cf. Crain & Steedman, 1985; Gillon, 1984). Consequently, the general preference for the a-phrase to have wide scope remains strong, because that interpretation allows its single reference to be maintained. By contrast, in the “Every...a” order, the a-phrase in the direct object is received after the every-phrase. Immediately upon reception of the a-phrase, it is apparent that it might have single or multiple reference, based upon whether it has wide or narrow scope. Thus, the a-phrase can be interpreted as having either single or multiple reference without having to switch from a previously assigned reference. (Every always has multiple reference, and so its reference never needs to be switched.) As a result, it is more common in the “Every...a” order than the “A...every” order for the general preference for wide scope on the leftward phrase to be overridden. We would therefore expect that other factors would be more likely to affect preferences in “Every...a” sentences than “A...every” sentences, because the single reference principle so strongly determines scope preferences in the latter case. Consistent with this expectation, an effect of verb type was observed in the “Every...a” condition but not in the “A...every” condition.

The interaction of the single reference principle with other principles suggests why verb effects should be more evident in “Every...a” sentences than in “A...every” sentences, but it does not predict the direction of the effect that was observed, that the WS1 preference was stronger for action verbs than perception verbs. This effect can be explained by adopting the thematic hierarchy principle as an account of the general preference for wide scope on the leftward phrase (without excluding the possible influence of other principles). The stronger preference observed for action-verb sentences than perception-verb sentences with the “Every...a” order would be due to agents having a stronger preference than experiencers for wide scope over themes, in accord with the hierarchy.

This first experiment indicates that at least two principles – thematic hierarchy and single reference – contribute to scope preferences. Note that the single reference principle is not a scope-processing principle per se – it does not refer directly to scope relations – but it does influence scope assignment because single reference of an a-phrase is strongly associated with wide scope in the absence of any particular context. This experiment has not ruled out the additional influence of any of the other principles, because all the principles predict that wide scope is preferably assigned to the leftward phrase in active sentences. To begin to distinguish among these principles, Experiment 2 examines passive sentences, for which the principles do not all make the same predictions.

**EXPERIMENT 2: PASSIVE SENTENCES**

In this experiment, the active quantifier sentences of Experiment 1 were replaced...
by their passive counterparts; for example, *A tree was climbed by every kid, Every statue was seen by a tourist.* The linear order, surface subject, topic, and c-command principles all predict that the leftward quantified phrase will preferably have wide scope over the rightward quantified phrase within the *by*-phrase, so that the WS1 interpretation is preferred. In passives, the leftward phrase obviously linearly precedes the rightward one, and it is also both surface subject and topic (Hornby, 1972; Olson & Filby, 1972). The exact structural position of the *by*-phrase in passives remains under debate, but on virtually all current views it is c-commanded by the surface subject NP (Baker, 1988; Grimshaw, 1990). By contrast, both the external argument principle and the thematic hierarchy principle predict that the rightward phrase will preferably have wide scope—a WS2 preference. This rightward phrase is the external argument. It also expresses the Agent or Experiencer role, which is higher on the thematic hierarchy than the Theme expressed by the leftward phrase.

**Method**

Materials were identical to those used in Experiment 1, with the following changes: the quantifier sentences were replaced by their passive counterparts, such that sentences of the form “NP1 VERBed NP2” were replaced by those of the form “NP2 was VERBed by NP1.” A complete example is shown in Table 2. In the same manner, the first sentence in each of the 20 similar incompatible fillers was replaced by its passive counterpart. Also, some of the other fillers that had

<table>
<thead>
<tr>
<th>Table 2.</th>
<th>Example stimuli, Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sentence</strong></td>
<td><strong>Ambiguous, WS1</strong></td>
</tr>
<tr>
<td>&quot;<em>Every . . . a</em>&quot; quantifier order</td>
<td></td>
</tr>
<tr>
<td>Quantifier</td>
<td>Every tree was climbed by a kid.</td>
</tr>
<tr>
<td>Continuation</td>
<td>The kids were full of energy.</td>
</tr>
<tr>
<td>&quot;<em>A . . . every</em>&quot; quantifier order</td>
<td></td>
</tr>
<tr>
<td>Quantifier</td>
<td>A tree was climbed by every kid.</td>
</tr>
<tr>
<td>Continuation</td>
<td>The tree was full of apples.</td>
</tr>
</tbody>
</table>

*Note: WS1 indicates an interpretation with wide scope on the first quantified noun phrase; WS2 indicates wide scope on the second quantified noun phrase. The fourth variable, *verb,* was manipulated between-items. Only an action verb example is shown here; an example of a perception verb sentence is *Every statue was seen by a tourist.*
been in the active voice were replaced by passive versions. The same procedure was followed as in Experiment 1. Forty-eight college students in the Ithaca area were paid for their participation.

Results

Figure 2 presents the compatibility judgment data. The strong WS1 preference observed for the active ambiguous sentences in Experiment 1 was not observed in Experiment 2’s passive ambiguous sentences. There was a slight WS1 preference for the ambiguous passages (62.2% for WS1 vs. 56.5% for WS2), but this effect was not robust, \( F_1(1, 47) = 4.56, p < .05, F_2(1, 30) = 1.29, \) n.s. Moreover, there were no marked effects of quantifier order or other factors for the ambiguous items in this experiment; the only other reliable result for these items was a Quantifier order \( \times \) Interpretation \( \times \) Verb interaction, \( \min F'(1, 70) = 4.16, p < .05 \), which, as can be seen in Fig. 2, seems to be due to small fluctuations in responses to WS1 interpretation sentences.

It is possible that the moderate frequency of compatible judgments (in the 50–70% range) in all ambiguous conditions is actually due to an averaging of the responses of two groups of participants – one group that strongly prefers the WS1 interpretation and another that strongly prefers the WS2 interpretation. To assess this possibility, we counted, for each participant, the number of passages judged compatible within each of the four types of ambiguous passages, as defined by crossing quantifier order and interpretation. The mean percentages (averaged across types) of participants who judged as compatible 0, 1, 2, 3, and 4 items within a type (out of a maximum of 4 in each type) were 5%, 19%, 27%, 29%, and 20%, respectively. This unimodal distribution is not consistent with the idea that there are two groups of participants. Inspection of each participant’s
responses to the types did not reveal any participant who clearly preferred WS1 or WS2 interpretations.6

Although the ambiguous items in Experiment 2 produced very different patterns than in Experiment 1, the patterns for the unambiguous items were similar across experiments. First, the continuation sentences following unambiguous sentences in Experiment 2 were often judged to be compatible (84.5% of the time), and significantly more so than the ambiguous items (59.4% compatible judgments), \( \min F'(1,76) = 55.85, \ p < .001 \). Second, there was a small but reliable preference for the WS1 over the WS2 interpretation for the unambiguous passages (89.1% for WS1 vs. 80.0% for WS2), \( \min F'(1,79) = 5.43, \ p < .05 \). Third, performance on the passive similar incompatible filler items (with correct response No) was 85% correct, comparable to Experiment 1.

Discussion

The important result in this experiment is that for scope ambiguous passive sentences, there is no consistent preference for one interpretation over the other, in sharp contrast to the ambiguous active sentences tested in Experiment 1. Since each of the processing principles under consideration predicted either a strong WS1 or WS2 preference, none is supported by this result. The combined results from Experiments 1 and 2 suggest that more than one principle is operative during quantifier scope ambiguity resolution, and that these principles lead to conflicting preferences for the passive. So, if two principles are operative (the simplest case), one principle favors the WS1 interpretation in passives (either

6It might also be suggested that the lack of clear preferences was due to participants simply guessing, because they found the ambiguous passive quantifier sentences to be very difficult to comprehend. Such difficulty could arise from the combination of the scope ambiguity and the added processing complexity of the passive structure (Forster & Olbrei, 1973). To test this possibility, we had a new group of 32 participants rate the comprehension difficulty of the ambiguous quantifier sentences used in Experiments 1 and 2. The sentences were presented in an item-counterbalanced design and were embedded in a large list of fillers of known varying difficulty. On a 5-point scale with 5 indicating most difficult, the Experiment 2 passive sentences were rated as slightly more difficult than the Experiment 1 active sentences (1.64 vs. 1.33, respectively), \( \min F'(1,51) = 17.75, \ p < .001 \). The only other reliable effect was that “A . . . every” sentences were rated as slightly more difficult than “Every . . . a” sentences (1.56 vs. 1.42, respectively), \( \min F'(1,59) = 4.40, \ p < .05 \). There was no reliable effect of verb or any significant interactions. Participants did use the full range of the scale in rating the fillers.

These results do not support the view that the ambiguous passive sentences were so difficult to comprehend that they would lead participants to guess. Though the passives were rated a bit more difficult than the actives, the difference was small and both were rated as quite easy. Also, the other significant effect, that “A . . . every” sentences were rated slightly more difficult than “Every . . . a” sentences, does not reflect the results of Experiments 1 and 2, which indicate clearer preferences for the former than the latter. It appears then that small differences in rated difficulty do not directly determine responses in the compatibility judgment task.
linear order, surface subject, c-command, or topic), and the other favors the WS2 interpretation (either thematic hierarchy or external argument). One or the other principle will “win” this conflict on different occasions, so that no single overall preference is observed for the passive items. For the actives, the two principles converge on the same interpretation rather than conflict, thereby leading to a consistent preference for the WS1 interpretation.

The Quantifier order × Interpretation × Verb interaction in the judgment data for the passive ambiguities is a sign of the complexity of the process by which one interpretation is chosen over the other. The fact that verb is a factor in this interaction suggests that one of the operative principles is the thematic hierarchy principle, which also explains the effect of the verb in Experiment 1. It was also found in Experiment 1 that preferences were sharper for the “A . . . every” order than the “Every . . . a” order. This was attributed to the single reference principle, that comprehenders treat an a-phrase in the surface subject or topic as referring to a single entity. No clear effect of this principle was observed in Experiment 2, but the fact that quantifier order was one of the factors in the interaction suggests that single reference might enter into the conflict among scope preference principles and thereby have some indirect influence.

Obviously we cannot yet explain the particular pattern of judgments revealed by the Quantifier order × Interpretation × Verb interaction in Experiment 2, but the fact that the interaction exists at all is explicable within a framework that posits conflicting principles operating upon the passive ambiguities. We delay any more detailed discussion of the framework until after investigation of another construction in Experiments 3 and 4, which will allow the contributions of specific principles to be distinguished further.

EXPERIMENT 3: COMPLEX NOUN PHRASES

This experiment investigated a complex NP construction. An example of a stimulus set is given in Table 3. In the quantifier sentence, the two quantifier terms occur in a complex NP that is the direct object of the main verb. The complex NP contains a prepositional Phrase (PP) complement. Following the current standard view (Radford, 1988), we assume a structure like that in Fig. 3 for the complex NP. This structure differs from the actives and passives examined.

This structure has two important properties which distinguish it from earlier proposed structures: the PP is sister of the noun, and the higher determiner c-commands all the other constituents of the complex NP. See May (1977) and Radford (1988) for evidence for this structure. An alternative structure in which NPs occur within determiner phrases (e.g., Fukui, 1986) maintains these two properties of the structure adopted here, and essentially the same predictions would be derived from the various processing principles.
Table 3. Example stimuli, Experiment 3

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Ambiguous, WS1</th>
<th>Ambiguous, WS2</th>
<th>Lexical bias, WS1</th>
<th>Lexical bias, WS2</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>&quot;Every ... a&quot; order</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantifier</td>
<td>George has every photograph of an admiral.</td>
<td>George has every photograph of an admiral.</td>
<td>George has each photograph of an admiral.</td>
<td>George has every photograph of the admiral.</td>
</tr>
<tr>
<td>Continuation</td>
<td>The admirals were quite famous.</td>
<td>The admiral was quite famous.</td>
<td>The admirals were quite famous.</td>
<td>The admiral was quite famous.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>&quot;A ... every&quot; order</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantifier</td>
<td>George has a photograph of every admiral.</td>
<td>George has a photograph of every admiral.</td>
<td>George has the photograph of every admiral.</td>
<td>George has a photograph of each admiral.</td>
</tr>
<tr>
<td>Continuation</td>
<td>The photograph was quite famous.</td>
<td>The photographs were quite famous.</td>
<td>The photograph was quite famous.</td>
<td>The photographs were quite famous.</td>
</tr>
</tbody>
</table>

Figure 3. Complex NP structure of stimuli in Experiments 3 and 4.

previously in that one of the quantified phrases is entirely embedded within the other.

The thematic hierarchy principle by itself predicts that there will not be any consistent preference for either the higher or lower NP to have wide scope. This is because in all the stimuli tested here, the complex NP is assigned the Theme role by the verb, and in most stimuli the prepositional object NP (the lower NP) is also assigned the Theme role.\(^8\) Because the ambiguity occurs entirely within the

\(^8\)Grimshaw (1990) has recently argued that in complex NPs in which the higher noun does not express an event, as are tested in this experiment, the lower NP is actually not assigned a thematic role at all. It seems natural to assume that such a phrase which does not bear a thematic role would be outranked on the hierarchy by a phrase which does bear a thematic role. This would then predict that the higher NP would preferably have wide scope. But we lack concrete evidence for this assumption and so will not elaborate on it in this paper. As described in the main text, however, this same prediction is made by other principles and is tested in the experiment.
direct object and does not involve the surface subject, external argument, or topic, the surface subject, external argument, and topic principles make no particular predictions for preferences and so are inapplicable here. If there is a consistent preference for a particular interpretation, it would thus be an effect of some principle(s) other than these three.

The linear order and c-command principles can be applied to this construction, but only if their formulations are modified somewhat. Neither NP linearly precedes or c-commands the other, since one NP is contained within the other. However, the linear order principle could be reformulated so that it refers to the linear order of just the quantifier terms or of just the head nouns of the quantified phrases. Either reformulation leads to the prediction that the higher NP will preferably have wide scope, so that a WSI interpretation is preferred. These reformulations do not alter the linear order principle's predictions regarding the previous experiments' active and passive constructions.

The c-command principle can similarly be reformulated to refer to the c-command relations of just the quantifier terms or of just the head nouns; again, this leads to the prediction that the higher NP will preferably have wide scope and a WSI interpretation will be preferred. The problem with this reformulation, however, is that it no longer makes any predictions regarding the actives and passives, since in those constructions neither quantifier term/head noun c-commands the other. A more complicated reformulation is therefore required. Two can be suggested: (i) the c-command principle refers to quantifier terms or head nouns when applied within an NP, but refers to quantified phrases when applied within a clausal domain; (ii) a broader dominate-or-c-command principle is operative that refers uniformly to quantified phrases. Leaving aside the question of the desirability of such reformulations, they both predict that the higher NP will preferably have wide scope (so that a WSI interpretation is preferred), while the predictions for actives and passives remain unchanged.

In sum, all the available hypotheses predict either no consistent scope preference or a preference for the higher NP to have wide scope.

Method

Twenty-four sets of experimental passages were prepared, each with eight versions. All stimuli are listed in Appendix B. In all sets, the quantifier sentence contained seven or eight words (depending upon whether the subject NP contained one or two words). The continuation sentence contained four or five words (constant within a set). All sentences were in the active voice.

In all but six sets, the thematic role assigned to the quantifier sentence's lower NP via the preposition was Theme; the exceptions were those in which the
preposition was from (Source), for (Beneficiary), and to (Goal). As Grimshaw (1990) notes, the position on the thematic hierarchy of these thematic roles relative to Theme is controversial.

Three independent variables were manipulated, each with two levels: ambiguity, quantifier order, and interpretation. There was no verb manipulation in this experiment.

In the previous experiments, the ambiguity variable referred to the comparison between ambiguous and similar unambiguous quantifier sentences. However, for the ambiguous quantifier sentences tested in this experiment, it is impossible to create a complete set of unambiguous counterparts that are similar in length and structure. Instead, “lexical bias” sentences were created that contained each or the, which are both known to have a strong bias to take wide scope (although whether the should be considered a quantifier term in all respects is an open question; see Fiengo & Higgonbotham, 1981; Heim, 1982; May, 1985). As shown in Table 3, every in the ambiguous sentences was replaced with each, in order to bias toward the interpretation in which the each-phrase has wide scope over the a-phrase, and a in the ambiguous sentences was replaced with the to create a bias toward the interpretation in which the the-phrase has wide scope over the every-phrase. The lexical bias conditions therefore allow us to observe interactions between structural and lexical factors. Because it is already clear from the unambiguous conditions and similar incompatibles of the previous experiments that participants are capable of performing the task at a high level of accuracy, it is not essential to have a fully unambiguous comparison in this experiment.

The quantifier order and interpretation variables were manipulated as in Experiments 1 and 2. Each ambiguous quantifier sentence was tested with both continuation sentences, while each lexical bias quantifier sentence was tested only with the continuation sentence that was appropriate for its lexical bias interpretation. Plausibility was also controlled for as in the previous experiments.

Sixty-eight of the filler items and all 20 practice items were the same as in Experiments 1 and 2. The remaining 12 fillers were similar incompatibles—structurally identical to the lexical bias versions of the experimental passages except that the continuation sentence did not match the lexical bias interpretation of the quantifier sentence. Only one version of each of these fillers was presented to all participants (three each corresponding to the four types of lexical bias stimuli).

Eight stimulus files were created. Each file contained the fillers and one member of each of the 24 sets of experimental passages (three members exemplifying each of the eight types). Five participants were tested on each file. The procedure was identical to that of Experiments 1 and 2. Forty MIT undergraduates were paid for their participation.
Results

Results are illustrated in Fig. 4, which shows that the pattern of preferences is unlike that of the previous experiments. Specifically, there was a strong preference for the WS2 interpretation, $min F'(1, 58) = 65.87, p < .0001$. There was no significant interaction of interpretation and ambiguity ($min F' < 1$), indicating that the WS2 preference held for both the ambiguous and lexical bias versions. However, there was a three-way interaction among ambiguity, interpretation, and quantifier order, $F_1(1, 39) = 5.00, p < .05, F_2(1, 23) = 5.73, p < .05$. As Fig. 3 shows, this interaction is due primarily to the difference in frequency of WS2 responses between the ambiguous “Every . . . a” versions (56.7%) and the corresponding lexical bias “Every . . . the” versions (81.7%).

Two different post hoc analyses were conducted to explore the possible bases of this WS2 preference. First, a separate analysis of just the data from the stimulus items in which the lower NP of the quantifier sentence was assigned the Theme role produced virtually the same pattern of means as in Fig. 3. A second analysis included animacy of the lower noun as a factor (the higher noun was inanimate in nearly all the experimental stimuli). This analysis revealed no main effect of animacy and no reliable interactions with any other factor. Thus, variations in features of the NPs do not appear to have determined the response patterns.

Results for the similar incompatible filler passages were in accord with the results for the lexical bias passages. Rather than responding No to all the similar incompatibles, participants tended to respond Yes when the continuation sentence fit the WS2 interpretation of the quantifier sentence (57.9%) but not when it fit the WS1 interpretation (29.6%).

![Figure 4: Compatibility judgments for continuation sentences in Experiment 3.](image-url)
Discussion

The results for the ambiguous items show that the lower NP was preferably assigned wide scope. The same lower NP preference was also observed for the lexical bias items, indicating that the preference is so strong that it can overcome the scope tendencies of particular quantifier terms. These results are contrary to the predictions of all the principles under consideration, since those principles predicted either no preference or a preference for the higher NP. In particular, these results indicate that the reformulated linear order and c-command principles, which predicted a preference for the higher NP, are invalid or, at least, are inapplicable within complex NPs. Accounting for the preference for the lower NP to have wide scope requires that some additional principle, perhaps applying only within complex NPs, be formulated.

Given that none of the available principles predicted a preference for the lower NP, it is important to obtain additional evidence for this preference. Experiment 4 examines two of the lexical bias quantifier sentences (containing “Each . . . a” and “Every . . . the”) in more detail. Unlike the current experiment, which presented only those continuations that were compatible with the lexical biases of each and the, Experiment 4 pairs these quantifier sentences with both WS1 and WS2 continuation sentences. This allows for a more complete assessment of the lower NP preference for these quantifier sentences, as well as of the influence of particular quantifier terms. (The results for the similar incompatibles in the current experiment are not sufficient for this assessment, because they were not matched for lexical content.)

We focus on the quantifier order in which a/the occurs in the lower NP because the magnitude of the lower NP preference for the ambiguous “Every . . . a” items was not as great as for the ambiguous “A . . . every” items or for the lexical bias items. This weaker preference may stem from some infelicity associated with an indefinite or non-specific NP occurring in a deeply embedded position. Experiment 4, by testing lexical bias quantifier sentence with both continuation sentences, can demonstrate effects of a lower NP preference that are independent of effects of any awkwardness that may result from an embedded a-phrase.

EXPERIMENT 4: COMPLEX NP FOLLOW-UP

Method

Materials were adapted from those of Experiment 3. There were 24 sets of experimental passages, each with four versions. As the example in Table 4 illustrates, two independent variables were manipulated in the stimuli. The first
was quantifier pair, “Each...a” and “Every...the”. In the “Each...a” condition, the lexical bias of the quantifiers conflicts with the preference for WS2 interpretations for this construction, because the leftward quantifier each strongly tends to take wide scope. By contrast, the quantifiers in the “Every...the” condition further support the WS2 interpretation, because the rightward quantifier the strongly tends to take wide scope. The second variable was interpretation (WS1 vs. WS2, as in the ambiguous versions of Experiment 3).

Fillers were the same as in Experiment 3, except that the 12 similar incompatible fillers were no longer needed to test the alternate interpretation of the lexical bias items, because both continuations were tested for the experimental items in this experiment. These items were therefore modified to contain a variety of quantifier terms and, like other fillers, were not analyzed. Four stimulus files were created; eight participants were tested on each file, following the same procedure as in the previous experiments. Thirty-two MIT undergraduates were paid for participation.

Results and discussion

The compatibility judgment data are illustrated in Fig. 5. As in Experiment 3, there was an overall preference for the WS2 interpretation, as indicated by a significant main effect of interpretation, \( \min F'(1, 53) = 14.40, p < .001 \). In addition, there was a significant Quantifier pair \( \times \) Interpretation interaction, \( \min F'(1, 53) = 12.54, p < .01 \), due to the WS2 preference being stronger for the
"Every ... the" versions, in which the lexical bias supports the WS2 interpretation, than for the "Each ... a" versions, in which the lexical bias supports the WS2 interpretation, than for the "Each ... a" versions, in which the lexical bias conflicts with the WS2 interpretation. For the "Each ... a" versions alone, the preference for the WS2 interpretation was significant although at a marginal level, $F_1(1, 31) = 3.32, p < .10, F_2(1, 23) = 3.83, p < .10$.

The major result is that there is a preference for the lower NP to have wide scope for the "Each ... a" condition. This is so even though each, in the higher NP, is a quantifier term with a strong bias toward wide scope and even though a in the lower NP might be somewhat awkward. However, the fact that the preference is weaker for the "Each ... a" condition than for the "Every ... the" condition may be attributable to one or both of those factors. As noted earlier, the wide scope preference for the lower NP requires that some additional processing principle be devised which is sensitive to particular structural properties of complex NPs.

The preference for the lower NP is not unprecedented in the linguistics literature. May (1977) and Reinhart (1983), following earlier investigators, report informal judgments favoring wide scope for the lower NP. They take this as evidence that wide scope for the higher NP is ill formed according to core grammatical principles (although May allows for some idiosyncratic cases to be well formed though marked). This seems extreme, however, since informants generally judge both scope interpretations to be permissible (at least when both are plausible). And, in fact, in later work, May (1985) allows for both

---

9This holds for all the ambiguous and lexical bias versions tested here except perhaps for the "Every ... the" version. Many informants report that the WS1 interpretation is impossible or highly strained for such sentences. It is possible that the converging structural and lexical processing influences favoring the WS2 interpretation have been grammaticalized so that it is the only well-formed interpretation.
interpretations to be fully well formed, although he does not attempt to explain why scope on the lower NP would be preferred by the processing system.\textsuperscript{10}

May (1977) and Reinhart (1983) also present informal judgments indicating that for complex NPs containing more than two NPs, such as \textit{every exit from a freeway to some California city}, the very lowest NP (\textit{some California city}) preferably has wide scope. Thus the principle(s) determining scope resolution for complex NPs may specifically favor the lowest phrase, rather than a phrase at any level of embedding within the entire complex NP.

GENERAL DISCUSSION

Although various types of information may contribute to the resolution of quantifier scope ambiguities, including lexical information and real-world knowledge, this investigation has focused upon structural information. Various principles for how structural information is utilized have been proposed in the previous literature, and the experiments reported here set out to gain evidence for or against those proposals. The results indicate that no single one of the principles can account for the full range of interpretation preferences across various constructions. Nor does it seem possible at present to devise a single new principle that accounts for the full range. While future linguistic and psycholinguistic research may reveal such a single principle, the current results point to the conclusion that multiple structural principles interact in determining preferences.

Major results

Although the current data do not uniquely identify the set of operative principles, they do serve to constrain that set. To review, Experiments 1 and 2 indicate that at least one principle favors wide scope for the leftward quantified phrase of a passive and at least one other principle favors wide scope for the rightward phrase. For actives, however, these same two principles both favor

\textsuperscript{10}According to May (1985, pp. 67–77), the two interpretations of the complex NP ambiguity correspond to different logical form (LF) structures (which are derived by syntactic movement of the quantified phrases from their surface structure positions). However, there is no evident reason at present why one LF structure would be preferred. In May's framework, the two interpretations of the active ambiguity correspond to the same LF structure and are distinguished only at a higher semantic level whose format is dissimilar to syntactic structure. The same holds for the passive ambiguity. This itself suggests that the processing principles which determine resolution of the complex NP ambiguity may be quite different from those applying to the active and passive ambiguities.
wide scope for the leftward phrase. Candidates for the principle favoring wide scope for the leftward phrase in passives include the linear order, surface subject, c-command, the topic principles; candidates for the other principle are the external argument and thematic hierarchy principles.

Further, an influence of verb type was observed. Of all the principles under consideration, only the thematic hierarchy principle can account for this influence. In addition, there was an effect of order of quantifier terms. This cannot be explained by any of the previously proposed principles, but we argued that it can be explained by the single reference principle, according to which an a-phrase in surface subject or topic position is preferably interpreted as referring to just a single entity. This is a reasonable principle since this interpretation is simpler and since the processor, upon encountering such an a-phrase, does not have foreknowledge of whether the sentence contains a subsequent quantifier that could induce multiple reference of the a-phrase. However, the influence of the single reference principle is weak, serving just to modulate the effects of the other principles.

Experiments 3 and 4 investigated the complex NP construction. None of the available principles can explain the preference for the lower NP to have wide scope. Some other principle, not yet precisely formulated but sensitive to the structural properties of this construction, must be invoked. Further, the preference for the lower NP is directly contrary to the linear order and c-command principles, suggesting that those principles either are entirely invalid or are inoperative within certain domains.¹¹

⁻¹¹ An additional point can be made here concerning the interpretation of the results of all these experiments. When the quantifier sentence’s a-phrase has narrower scope than the every-phrase, we have treated the plural continuation sentence as most appropriate (e.g., Every kid climbed a tree. The trees were full of apples). Actually, the narrow-scope a-phrase can refer either to a non-specific set of entities, of indeterminate size (consistent with the plural continuation sentence), or to one specific entity (consistent with the singular continuation sentence). Following Fodor and Sag (1982), Hccim (1982), and Enc (1991), we assume that the narrow-scope a-phrase is construed as non-specific. The non-specific construal is the more general case for the narrow-scope a-phrases, and indeed there seems no reason why the processor would initially adopt the specific construal unless context or the specific content of the sentence supported it (which is not so for the sentences examined here). Furthermore, the narrow-scope interpretation with specific construal is logically equivalent to the wide-scope interpretation. It would be odd for the processor to build a representation of narrow scope that includes an explicit marker for specificity, rather than build a representation for wide scope that does not require an explicit specificity marker.

The pattern of results across experiments supports this assumption. A clear preference for single reference was observed for the actives in the “A... every” order (Experiment 1) and for the complex NPs in the “Every... a” order (Experiments 3 and 4). There is no evident reason why, for either of these cases, there should be a preference for the specific construal over the non-specific construal, whereas reasonable principles can be developed to account for these as wide-scope preferences. Furthermore, for the actives in the “Every... a” order, there was a higher frequency of single reference responses for the stimuli with perception verbs than for those with action verbs. The thematic hierarchy principle can explain this in terms of narrow versus wide scope; it is not clear how another principle would explain it in terms of specific versus non-specific construal.
Whatever the exact set of interacting principles is, they all refer to the syntactic or semantic positions within a domain no larger than that of a single clause. This distinguishes our claim from previous proposals that have allowed for the interaction of multiple principles. In all those proposals, only a single principle referred to structural positions of clausemate phrases, while the other principles concerned phrases occurring in different clauses (Lakoff, 1971; Moran, 1988; VanLehn, 1978) or the scope tendencies of particular quantifier terms (Fodor, 1982; Ioup, 1975; Kroch, 1975; Moran, 1988; VanLehn, 1978). These other principles are reasonable, but the present work, by showing that there is more than one principle that is sensitive to the positions of clausemate phrases, indicates that the determinants of scope preference are even more complex than has been generally presumed.

The demonstration of effects of structurally based principles is contrary to the view, suggested in Katz (1980), that preferences can be explained entirely at a pragmatic level. It is possible to develop pragmatic-level versions of the topic and single reference principles, but not for any of the other principles—and, indeed, the results indicate that some of these other principles are operative. Further, if these structurally based principles influence preferences, then it would seem that scope is represented at some level of semantic structure, not simply at a pragmatic level, for otherwise it is hard to see how structural position could have an effect (see also Jackendoff, 1981).

The experiments did not show any clear effects of the supposedly greater tendency of every to take wide scope over a. This suggests that principles referring to structural positions have a stronger influence on preferences than do lexical tendencies, at least for these two quantifiers. More extreme differences in lexical biases, as between each and a and between every and the in Experiment 4, may have a somewhat greater role in modulating structural effects.12

Finally, the results leave open the question of whether quantifier scope resolution is a mandatory process. VanLehn (1978) and Fodor (1983) suggest that, unlike lexical and syntactic processing, scope processing is optional or at least can be delayed. The present experimental method may well have encouraged participants to resolve scope ambiguities early—prior to fully processing the continuation sentence—but it is not clear whether people always do this, especially when they do not know that the comprehension of subsequent material

12The lack of strong quantifier term effects is contrary to Moran’s (1988) proposal. Moran claimed that every has a tendency to take wide scope over a, but rather than attribute this to properties of the individual quantifier terms he proposed that logically weaker interpretations are preferred over logically stronger interpretations. A similar proposal could be developed from Kempson and Cormack (1981), although they do not take this step themselves. (A weaker interpretation is one which is implied by, but does not imply, the stronger interpretation. Wide scope of every over a is weaker than wide scope of a over every, since the former interpretation includes the case, expressed in the latter interpretation, in which there is just one specific entity referred to by the a-phrase.)
or the task response is dependent upon adoption of one particular scope interpretation.

**Processing mechanisms**

In exploring how the various scope ambiguity resolution principles would be implemented in a processing mechanism, it is useful to compare our principles with those proposed by Frazier and Fodor (1978) for syntactic ambiguity resolution (see also Fodor & Frazier, 1980; Rayner, Carlson, & Frazier, 1983). Similar to our claim, Frazier and Fodor argued that it is necessary to refer to more than one principle (e.g., minimal attachment and local association) operating together in order to account for initially preferred structures across the entire range of syntactic ambiguities. They further show how each of their principles might be derived from particular properties of the organization of the processing mechanism. We have not attempted here to derive our scope principles from the organization of the processor, but we agree that this will be an important explanatory goal once the principles are more clearly specified. One possible result is that seemingly independent resolution principles will be shown to derive from or be motivated by common aspects of the processor.

It is already possible, however, to point to a difference between our principles and those of Frazier and Fodor which leads to contrasting views about the overall form of processing. According to Frazier and Fodor, for any one type of syntactic ambiguity (as distinguished by syntactic structure and by lengths of constituents), only one principle is determinative of the preferred structure. As they show, this is compatible with a processing mechanism in which just one structure is ever initially considered for an ambiguity, because the case never arises in which different principles apply to a particular ambiguity and thus could favor alternative structures.

By contrast, the current results suggest that for a particular scope ambiguity, more than one principle may jointly determine the preferred interpretation. Further, it is possible for different principles to favor distinct interpretations, so that, as for passives, there is either no consistent preference or a weak one. This suggests that more than one interpretation is initially considered.

Two general classes of parallel processing mechanisms, drawn from work on syntactic processing and differing from the type of mechanism favored by Frazier and Fodor, can capture the interaction of the scope principles. According to ranked-parallel models (Gibson, 1991; Gorrell, 1989; Kurtzman, 1985), complete representations of the various possible interpretations for the scope ambiguity are always built in parallel. These representations are ranked, based on how well they satisfy the various principles. Then, just one representation (generally the highest-ranked one) is selected and the other(s) discarded. If no representation is ranked sufficiently higher than the others, then one can be chosen at random.
The second class of mechanism is based on competition, as developed by MacWhinney (1987; MacWhinney & Bates, 1989; see also the related approaches of MacDonald, 1993; McClelland & Kawamoto, 1986). The various principles would all apply (compete) during the building of a representation. If the principles (whose influences can be weighted) collectively favor one particular representation, then only that representation is built. If the principles conflict and do not collectively favor a particular representation with sufficient strength, then both representations are built, with one subsequently selected (perhaps randomly). An alternative possibility, if the principles conflict, is that one principle (randomly determined) initially overrides the other, so that only one representation is built. However such a model is elaborated, when there is conflict among principles, there is at least an implicit consideration of parallel interpretations, even if only a single complete representation gets built.

While it is possible that processing mechanisms at different levels of linguistic analysis are organized in distinct ways, in fact much current psycholinguistic research suggests that processing mechanisms at all levels initially consider parallel interpretations for ambiguities. This view is well established for lexical ambiguities (Seidenberg et al., 1982; Swinney, 1979). The parallelism claim for syntactic ambiguities (contra Frazier & Fodor) is more controversial, but it appears to be steadily gaining evidence in its favor (Altmann & Steedman, 1988; Gorrell, 1989; Kurtzman, 1985; MacDonald, 1993). Our claim that quantifier scope ambiguities are resolved by comparing initial interpretations in light of multiple principles is therefore a claim that the scope-processing mechanism shares certain basic characteristics with processing mechanisms at other linguistic levels.

More direct evidence for parallelism in scope ambiguity resolution and for a specific parallel processing mechanism requires the development of on-line methods for scope ambiguities. Following the model of research on lexical and syntactic ambiguities, where on-line methods are already being pursued, it will be especially important to investigate the time-course by which structural principles, lexical biases, and real-world knowledge are used and interact (cf. Altmann & Steedman, 1988; Ferreira & Clifton, 1986; MacDonald, 1993; Rayner et al., 1983; Trueswell, Tanenhaus, & Kello, 1993). For example, evidence for early effects of real-world knowledge would strengthen the claim for parallelism, because such effects can reasonably occur only if real-world knowledge influences the selection from an initial set of parallel choices.

References


**Appendix A: stimuli for Experiments 1 and 2**

The stimuli are given in the active “Every... a” quantifier sentence, with a WS2 continuation sentence. Items 1–16 contain action verbs, 17–32 perception verbs. Passive versions of the items were used in Experiment 2. Half of the continuation sentences began with the determiner *the*, and half *this* in their singular form and *these* in their plural form. This manipulation had no effect on judgments.

1. Every actor used a prop./The prop was on the stage.
2. Every shopper squeezed a melon./This melon was large and green.
3. Every policemen hit a suspect./The suspect was fat and sweaty.
4. Every dog chased a squirrel./This squirrel was small and frightened.
5. Every soldier saluted a flag./The flag was above the fort.
6. Every hiker scaled a cliff./The cliffs were wet and slippery.
7. Every pirate attacked a ship./This ship was loaded with treasure.
8. Every kid climbed a tree./The tree was full of apples.
9. Every thief grabbed a sack./This sack was full of jewels.
10. Every cowboy cornered a steer./The steer was angry and confused.
11. Every electrician tested a circuit./This circuit was old and faulty.
12. Every trainer rewarded an elephant./The elephant was learning a trick.
13. Every maid scrubbed a hallway./This hallway was long and narrow.
14. Every robber stole a car./This car was new and expensive.
15. Every student rode a bus./The bus was old and dirty.
16. Every scout cleaned a mural./The mural was in the park.
17. Every fox saw a chicken./The chicken was in the henhouse.
18. Every boy watched a player./This player was in the stadium.
19. Every agent observed a spy./The spy was in enemy territory.
20. Every sailor sighted an airplane./This airplane was near the ship.
21. Every detective noticed a clue./The clue was in the diary.
22. Every hound scented a raccoon./This raccoon was under the house.
23. Every executive heard a speech./The speech was about new products.
24. Every drunk saw a crime./This crime was in the park.
25. Every student heard a lecture./The lecture was about nuclear energy.
26. Every guard watched a prisoner./This prisoner was planning an escape.
27. Every child spotted a bird./The bird was on a rooftop.
28. Every antelope smelled a lion./This lion was in the clearing.
29. Every tourist saw a statue./The statue was in a courtyard.
30. Every reporter overheard a rumor./This rumor was about the mayor.
31. Every doctor observed a patient./The patient was thin and pale.
32. Every accountant detected an error./This error was in the ledger.

Appendix B: stimuli for Experiments 3 and 4

Experiment 3 quantifier sentences are shown in the ambiguous “Every . . . a” order with the WS2 continuation sentence. The stimuli were identical in Experiment 4 except that the quantifier pairs “Each . . . a” and “Every . . . the” replaced the Experiment 3 quantifiers.

1. I saw every picture of an acrobat./The acrobat was very good.
2. Amy has every package from a student./The student is very kind.
3. The secretary overheard every comment about a professor./The professor was out of town.
4. Joan looked at every sketch of a dancer./The dancer was very graceful.
5. Sam played every song about a cowboy./The cowboy was very brave.
6. John bought every book about a poet./The poet was very interesting.
7. The library has every map of a state./The states are quite beautiful.
8. The governor met every representative of a labor union./The labor union was very militant.
9. The manager read every evaluation of a salesman./The salesman was performing well.
10. Jack photocopied every article about a hostage./The hostage was kept in a basement.
11. The manager responded to every request from a worker./The worker was very upset.
12. We saw every advertisement for a show./The show looked great.
13. The executive read through every report on a product./The product was innovative.
14. The commissioner has every message from a farmer./The farmer was becoming impatient.
16. The tourist took every pamphlet about a cruise. / The cruise looks wonderful.
17. I read every story about a terrorist. / The terrorist was quite violent.
18. The reporter heard every rumor about a candidate. / The candidate was in trouble.
19. The treasurer has every contribution from a neighbor. / The neighbor is generous.
20. George has every photograph of an admiral. / The admiral was quite famous.
21. The teacher graded every essay about an emperor. / The emperor was very interesting.
22. The church has every painting of a bishop. / The bishop was highly respected.
23. Mary read every review of a documentary. / The documentary sounded interesting.
24. The IRS questioned every contributor to a charity. / The charity was under investigation.