JOURNAL OF MEMORY AND LANGUAGE **38,** 283–312 (1998)
ARTICLE NO. ML972543

Modeling the Influence of Thematic Fit (and Other Constraints) in On-line Sentence Comprehension

Ken McRae

University of Western Ontario, London, Canada

Michael J. Spivey-Knowlton

Cornell University

and

Michael K. Tanenhaus

University of Rochester

The time-course with which readers use event-specific world knowledge (thematic fit) to resolve structural ambiguity was explored through experiments and implementation of constraint-based and two-stage models. In a norming study, subjects completed fragments that ended in the ambiguous region of a reduced relative clause (The crook arrested/by/the/detective). Completion proportions up to and including *the* were influenced by thematic fit. The results were simulated using a competition model in which independently quantified syntactic and semantic constraints simultaneously influenced interpretation. Predictions were then generated for a self-paced reading task using model parameter values established by the off-line simulations. The pattern of reading times matched the predictions of the constraint-based version of the model but differed substantially from a one-region delay garden-path version. In addition, a garden-path model with a very short delay simulated the data better than the one-region delay model, but not as closely as the constraint-based version. The experiment and modeling illustrate that thematic fit is computed and used immediately in on-line sentence comprehension. Furthermore, the modeling highlighted the difficulty of interpreting sentence comprehension experiments without both quantifying the relevant constraints and implementing the mechanisms involved.

Real-time language comprehension requires that comprehenders access and integrate a variety of types of knowledge, including knowledge that is primarily linguistic and knowledge that is primarily conceptual (Clifton, 1993; Swinney, 1979; Tanenhaus & Trueswell, 1995). Although there is little disagreement about the integrative nature of comprehension, models of sentence processing make different claims about the time-course with which different types of information exert their influence. In particular, models differ in their claims about what information is used *initially* to structure the input when readers and listeners encounter one of the numerous temporary syntactic ambiguities that arise because linguistic input is presented sequentially. The goal of the present research was to use experimental data in conjunction with computational modeling to evaluate claims made by different models about the time-course with which semantic/conceptual information about thematic fit is used in on-line sentence comprehension.

One of the major theories of sentence pro-

This work was supported by NSERC Grant OGP0155704 to the first author, and NIH Grant HD27206 to the third author. We thank Keith Rayner and two anonymous reviewers for helpful comments on an earlier draft.

Address correspondence and reprint requests to Ken McRae, Department of Psychology, Social Science Centre, University of Western Ontario, London, Ontario N6A 5C2, Canada. E-mail: kenm@sunrae.sscl.uwo.ca.

0749-596X/98 $25.00

cessing over the last 15 years has been the garden-path model developed by Frazier and colleagues (Frazier, 1987). This theory maintains that syntactic category information plays a privileged role in initially structuring the linguistic input. Principles defined over syntactic representations determine an initial syntactic structure. Conceptual information is delayed relative to grammatical information, and is used to evaluate and, if necessary, revise that initial structure. For example, Ferreira and Clifton (1986) argued that ''If the syntactic processor (or parser) is modular, it should initially construct a syntactic representation without consulting nonsyntactic information sources . . . Notice, however, that the modular view does not imply that this higher-level information is never consulted by the language processor. It is important to distinguish between *initial* and *eventual* [original emphasis] use of nonsyntactic information.''

In contrast, constraint-based models assume that multiple syntactic alternatives are evaluated using both linguistic and non-linguistic sources of constraint. The comprehension system continuously integrates all relevant and available information in order to compute the interpretation that best satisfies those constraints (Bates & MacWhinney, 1989; Jurafsky, 1996; McClelland, St. John, & Taraban, 1989; MacDonald, Pearlmutter, & Seidenberg, 1994a; Spivey-Knowlton & Sedivy, 1995; Spivey-Knowlton, Trueswell & Tanenhaus, 1993). Thus constraint-based models predict that conceptual information that is correlated with different syntactic alternatives can play a central role in guiding even the earliest moments of language comprehension in general, and ambiguity resolution in particular.

THEMATIC FIT

The assignment of thematic roles to noun phrases has served as an important empirical domain for evaluating the use of conceptual information in sentence processing. In semantic/conceptual terms, a verb's thematic roles describe the semantic role or mode of participation played by an entity in the activity or event denoted by the verb (cf. Carlson & Tanenhaus, 1988). In many of the classic attachment ambiguities that have formed the basis for much of the contemporary sentence processing literature, the alternative syntactic structures involve different thematic role assignments. For example, in the main clause/reduced relative clause ambiguity exemplified by the fragment ''The cop arrested . . .'', ''The cop'' would play the role of agent if the fragment was completed as a main clause, for example, ''The cop arrested the crook.'', whereas it would play the role of theme or patient in a reduced relative clause, for example, ''The cop arrested by the FBI agent was convicted for smuggling drugs.''[1]

An extensive literature, beginning with Rayner, Carlson and Frazier (1983), has examined the time-course with which information about thematic fit is used in syntactic processing. Rayner et al. monitored eye movements while subjects read sentences containing reduced relative clauses and dative verbs such as *send,* as illustrated in (1a) and (1b).

(1a) The florist sent the flowers was very pleased.

(1b) The performer sent the flowers was very pleased.

These sentences differ in that florists are more typical flower-senders than are performers, whereas performers are more likely to receive flowers than are florists (although many florists must be sent flowers so that they can then send them to their customers). Rayner et al. found that first-pass reading times at the disambiguating verb phrase (*was very pleased*) were longer compared to main clause controls. However, total reading times were shorter for the sentences with more typical recipients. They concluded that initial syntactic commitments do not take into account plausibility information; however, such information can

[1] A reduced relative clause can be of other forms as well. For example, there is no overtly stated agent in, ''The man cooked the steak thought it tasted like shoe leather.''

be used by a ‘‘Thematic Processor’’ to evaluate, and if necessary, revise initial structural commitments.

Ferreira and Clifton (1986) more directly manipulated thematic fit in an eye-tracking study by varying the animacy of the first noun phrase, using materials such as those in (2).

(2a) The witness examined by the lawyer turned out to be unreliable.
(2b) The evidence examined by the lawyer turned out to be unreliable.

Because animacy is a primary characteristic of agents, these materials provided a stronger manipulation of thematic fit than did Rayner et al.’s (1983, but cf. Trueswell, Tanenhaus & Garnsey, 1994; Burgess & Lund, 1997). In addition, thematic fit was manipulated at the point of ambiguity, the verb in the reduced relative, whereas Rayner et al.’s plausibility manipulation was contingent on the noun phrase that followed the ambiguous verb. Ferreira and Clifton reasoned that if readers could use information about thematic fit to influence initial syntactic commitments, then reduced relatives with inanimate agents should be substantially easier to process than those with animate agents.

Contrary to this hypothesis, Ferreira and Clifton (1986) found that reading times at the prepositional phrase (*by the lawyer*) were longer for reduced than for full relatives and the magnitude of this difference was not influenced by animacy. In addition, reading times at the ambiguous verb were longer for sentences with inanimate noun phrases. Ferreira and Clifton argued that this pattern of results indicated that readers were aware of the poor fit between the inanimate noun phrases and the agent role leading to the increased reading times at the verb, but were unable to use this information to resolve the ambiguity in favor of a reduced relative, resulting in a garden-path effect when the sentence was disambiguated as a relative clause.

Subsequent studies that have manipulated thematic fit by varying animacy have strongly qualified Ferreira and Clifton’s (1986) conclusions. For example, Trueswell et al. (1994) conducted two eye-tracking studies, modeled on Ferreira and Clifton (1986, Experiment 1), but with more strongly biasing materials. They found a clear interaction between animacy and syntactic ambiguity in reading times at the disambiguating phrase. Burgess, Tanenhaus and Hoffman (1994) found that the effectiveness of animacy constraints interacted with the availability of parafoveal constraints. When sentences were presented two words at a time with the verb and preposition presented together, inanimate subject noun phrases completely eliminated processing difficulty for reduced relatives when the verb was followed by a short preposition that strongly biased a reduced relative interpretation (e.g., *by*), but it had weaker effects when a longer preposition was used. When one-word presentation was used, preventing parafoveal viewing of the preposition, animacy effects were weaker and delayed. In addition, MacDonald, Pearlmutter, and Seidenberg (1994b) and Trueswell and Tanenhaus (1994) argued that contextual effects for reduced relatives are modulated by the relative frequency with which an ambiguous *ed* verb is used in the past tense, as it is in a main clause, versus as a passive participle, as it is in a reduced relative clause. Ferreira and Clifton’s verbs had a strong past tense bias. Trueswell (1996) confirmed this prediction in a self-paced reading study. Inanimate noun phrases eliminated processing difficulty for reduced relatives with verbs whose *ed* forms are typically used as a passive participle but not with verbs whose *ed* forms are typically used as a past tense.

These studies are clearly consistent with constraint-based models in that animacy can have extremely rapid effects on ambiguity resolution, and its influence is modulated by other relevant constraints. However, it is premature to conclude that conceptually-based thematic information is influencing syntactic ambiguity resolution based on animacy effects alone. Animacy is marked morphologically in the grammars of many of the world’s languages and thus can be considered a grammatical property. Animacy is also a plausible selectional restriction in the sense of Chomsky

(1965). Selectional restrictions are semantic features that are lexically marked as being syntactically relevant for argument selecting lexical items, such as verbs. Caplan, Hildebrandt, and Waters (1994) argue that selectional restrictions are the only semantic factors that influence initial syntactic ambiguity resolution. Similarly, in treatments such as that of Schlesinger (1995), thematic roles are viewed as primarily linguistic constructs that are composed of a few core semantic features. Animacy is a prime candidate for a core feature of the agent role.

To examine the time-course with which conceptually-based information is used in ambiguity resolution, it is necessary to manipulate thematic fit while holding constant features such as animacy that arguably have special linguistic status. Several studies in the literature have included manipulations of this type. Stowe (1989, Experiment 2) found that thematic fit influenced readers' ability to resolve a closure ambiguity, and Boland, Tanenhaus, Garnsey, and Carlson (1995) found that thematic fit affected readers' gap filling strategies in sentences with fronted *wh*-phrases. However, because these experiments used cumulative word-by-word reading with simultaneous grammaticality (Stowe) or stop-making-sense judgments (Boland et al.), claims of immediate information use can be challenged because the secondary judgment tasks produce long reading times that might reflect the use of decision strategies.

In a self-paced reading study that did not feature a secondary task, Taraban and McClelland (1988, Experiment 2) manipulated the fit between the expected and actual thematic role in a postverbal prepositional phrase, as in (3).

(3a) The janitor cleaned the storage area with the broom because of many complaints.

(3b) The janitor cleaned the storage area with the solvent because of many complaints.

These sentences differ only in that *broom* (3a) is a more typical instrument in this context than is *solvent* (3b). Reading times were longer in sentences like (3b) on the word following the target noun (*because*). Although this effect is rapid, it can be argued that it should have occurred at the noun itself because the information differentiating the sentences was available at that point.

McRae, Ferretti, and Amyote (1997), Pearlmutter and MacDonald (1992), and Tabossi, Spivey-Knowlton, McRae, and Tanenhaus (1994) have all found relatively weak and/or late effects of thematic fit on the resolution of the main clause/reduced relative clause ambiguity when animacy was held constant. More recently, however, Garnsey, Pearlmutter, Meyers and Lotocky (1996) found that the plausibility of an NP as the theme of a verb in sentences with noun phrase/sentence complement ambiguities had rapid effects on ambiguity resolution for those verbs which occurred equally often with a noun phrase or a sentence complement (as determined by norms), but not for verbs with strong argument structure preferences. This study further highlights the importance of examining thematic fit in conjunction with other relevant constraints.

In summary, although there is clear evidence for the rapid use of animacy information in ambiguity resolution, evidence regarding the time-course of the influence of thematic fit is less clear. Most studies have found either weak or delayed effects of thematic fit when animacy was held constant. These data are superficially consistent with either two-stage models in which information about thematic fit is delayed or with models in which thematic constraints are used immediately but effects appear delayed because they conflict with other constraints.

The central goal of the current work was to examine the time-course of thematic fit information with animacy held constant, comparing predictions of two-stage and multiple-constraint models. This was done by varying the fit between the initial noun in a reduced relative (''The crook/cop arrested by the detective . . .'') and the agent and patient roles associated with the verb (see McRae et al., 1997, for an account of how this process might oc-

cur). Both two-stage and constraint-based models predict that reduced relative clauses should be more difficult to process than unambiguous full relative clauses even when information about thematic fit is biased towards the patient role. However, the models make these predictions for different reasons. In two-stage models, processing difficulty occurs because thematic information is not used until after the parser has initially chosen the main clause structure. In constraint-based models, information about thematic fit interacts with other constraints that are likely to be biased toward the main clause. Thus there is not a signature data pattern that will distinguish two-stage from constraint-based models in the absence of explicit assumptions about how constraints are integrated.

A COMPETITION MODEL

In order to generate clear predictions, we used a competition-integration model developed by Spivey-Knowlton (1996) in which multiple constraints provide probabilistic support for possible syntactic alternatives. Tanenhaus, Spivey-Knowlton, and Hanna (in press) have used this architecture to conduct simulations of reading times in which the weights on the constraints were held constant while the constraint values varied. They showed that the conflicting results obtained in experiments manipulating animacy with reduced relatives are accommodated by taking into account differences in the strength of the constraints used in the materials in the different experiments. Similarly, the model has been used to simulate data patterns from recent experiments investigating the effects of discourse context on ambiguity resolution with reduced relative clauses (Spivey-Knowlton & Tanenhaus, 1997).

The competition-integration model assumes the availability of multiple syntactic alternatives that compete for activation. Constraints provide probabilistic evidence in support of the various alternatives. The weights among the constraints are normalized so that activation ranges between 0 and 1. The model iterates in that there are multiple processing cycles given each input. On each cycle, the evidence in support of the syntactic alternatives is computed. Competition ends when the activation of one alternative reaches a criterion. Processing time is assumed to be a linear function of the duration of competition. When all constraints strongly favor one alternative, competition is resolved quickly, leading to predictions of little or no processing difficulty. As the constraints become more balanced, competition time increases. Thus the model can be used to make quantitative predictions about the magnitude of ambiguity effects when a temporarily ambiguous sentence is compared against an unambiguous baseline. It is important to note that the integration model is not a model of how syntactic alternatives are generated and therefore cannot account for processing time that is associated with this process. However, the model **can** be used to contrast the central claims about thematic fit made by constraint-based and two-stage models. Constraint-based models predict that processing difficulty arising from ambiguity stems from competition among multiple constraints that apply simultaneously. In sentences such as (4), the effects of ambiguity on processing difficulty should be a function of the interaction among thematic fit and other available relevant constraints.

(4) The crook arrested by the detective was guilty of taking bribes.

In contrast, two-stage models predict that thematic fit, as well as any other semantic and lexically-specific constraint, should be delayed until after initial selection of the preferred syntactic structure. In order to evaluate these claims using the competition-integration model, it is necessary to (1) quantify the strength of thematic fit and the other plausible constraints with which it interacts, and (2) assign weights to the constraints.

Figure 1 shows a schematic of the model. We incorporated four constraints at the verb + by, *arrested by* in (4), that previous studies suggest are clearly relevant to our materials. The first is the thematic fit between the initial noun phrase and the verb-specific agent and patient roles of the ambiguous verbs. The sec-

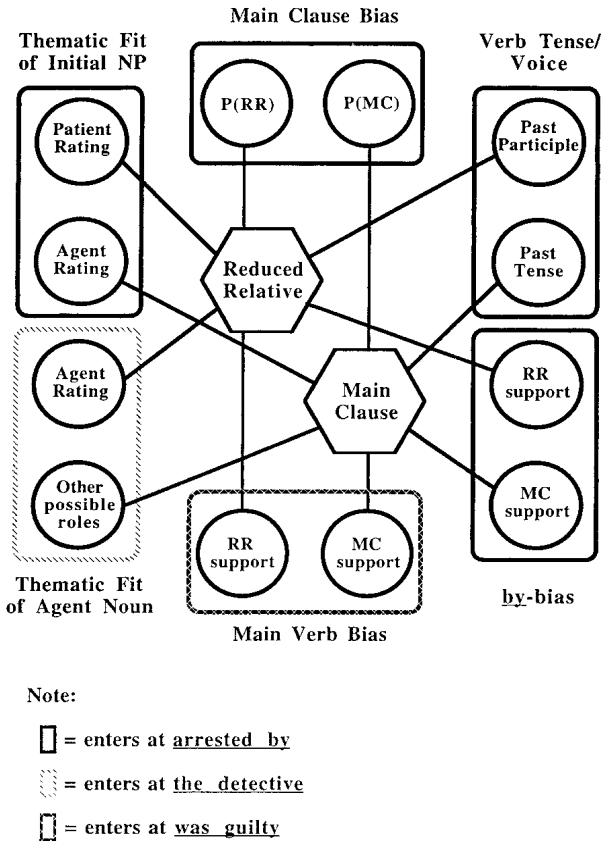

FIG. 1. A schematic of the competition model. Note that the thematic fit of the initial NP, the main clause bias, the *by*-bias, and the verb tense/voice constraint become operative at the verb + by region. The thematic fit of the agent NP comes into play at the agent NP region and the main verb bias becomes operative at the main verb region.

ond is the relative frequency with which the ambiguous verb occurs as a simple past tense and as a past participle, following Trueswell (1996). The more frequently the verb is used as a past participle, the stronger the support for the relative clause structure. It should be noted that past participle frequency as measured by Francis and Kucera (1982) is only a rough estimate of the tense bias associated with a verb's *ed* form. Nonetheless, it captures a significant portion of the item-specific variance associated with individual verbs in reduced relative constructions (MacDonald et al., 1994b; Trueswell, 1996; Spivey-Knowlton & Tanenhaus, 1997). The third is the preposition *by* following the verb. Previous research has shown that *by* provides a strong constraint in support of a relative clause: *by* after an *ed* verb is typically used to introduce an agent in a passive construction, and research has shown that it interacts with thematic and contextual constraints (Burgess et al., 1994; Spivey-Knowlton et al., 1993). Readers typically do not fixate separately on *by*, but instead process it parafoveally when fixating the preceding verb (Spivey-Knowlton & Tanenhaus, 1997; Trueswell et al., 1994). Therefore, for the purpose of studies in which *by* is displayed with the verb, the *by* constraint can be treated as being at least partially available as the verb is being processed.

The fourth constraint was a configurational bias favoring the main clause over the relative clause. A sentence-initial sequence of ''noun phrase verbed'' is typically the beginning of a main clause (Bever, 1970; MacDonald et al., 1994a; Tabossi et al., 1994). For present purposes, we remain agnostic about whether the main clause bias is best characterized at the structural level (e.g., Frazier & Fodor, 1978; Mitchell, Cuetos, Corley, & Brysbaert, 1995; Stevenson, 1994) or whether it emerges from other more local constraints, including argument structure preferences or other lexically-triggered constraints (e.g., Juliano & Tanenhaus, 1994; MacDonald et al., 1994a; Trueswell, 1996). Treating the clause bias as a separate constraint allows the model to remain neutral between multiple-constraint models that include probabilistic constraints defined in terms of syntactic categories (e.g., Jurafsky, 1996) and those that do not. Crucially, it also allows us to test an implemented two-stage model in which an initial stage of competition is restricted to using configural information only. This can be accomplished by delaying the other constraints relative to the use of the main clause bias.

Additional constraints strongly supporting the reduced relative clause come into play later in the sentence. The thematic fit of the agent noun as an agent, manner, time, or location enters at the noun phrase following the verb, *the detective* in (4). Finally, a structural bias strongly favoring a reduced relative interpretation is provided by the main verb, *was guilty.*

Although the constraints that we incorporated into our model are established in the literature, there are others such as argument structure that were not directly included. To the extent that there is item-specific variability in the materials that can be captured only by argument structure, failing to include it as a separate constraint will diminish the ability of the model to account for item-specific variance. Later, we return in more detail to the question of how the decision to use some constraints and not others affects the generality of the conclusions that can be drawn from our modeling efforts. For now, it is important to keep in mind a general point about this style of modeling. The extent to which the modeling effort is compromised by not including item-specific constraints such as argument structure depends on two factors. The first is the degree to which the items in the experiment vary along these dimensions; the greater the variability, the more important it is to include the constraint. The second factor is the degree to which the constraint is subsumed by and/or correlated with other constraints. Inspection of the sentence completions suggested that our verbs had strong transitive biases, and tense frequency is correlated with transitivity (Trueswell, 1996). Therefore, the tense constraint likely captures some of the item-specific variability due to argument structure whereas the overall transitivity bias is likely to be captured by the main clause bias.

In order to quantify the constraints, we used a combination of typicality ratings and corpus analyses. Following McRae et al. (1997), thematic fit was quantified using typicality norms to determine the goodness of fit between the entity denoted by the noun (*crook* or *cop*) and the agent and patient roles of the verb (an arrest-agent and arrest-patient). The strengths of the verb tense constraint, the main clause bias, and the *by* constraint were determined by corpus analyses. The weights for the constraints were determined by collecting gated sentence completions for fragments that were compatible with either a main clause or a reduced relative clause. We then adjusted the weights to simulate the completion data. This provided a model of how constraints are integrated during off-line processing in the absence of time pressure. Under these conditions, all models predict that ambiguity resolution is influenced by multiple constraints, one of which is thematic fit. We then used the model with the same constraints and weights to predict on-line reading times for a self-paced reading study. The rationale for doing this is as follows. The use of on-line measures to address architectural questions in sentence processing is largely motivated by the assumption that architectural influences concerning the time-course of the application of different constraints, for example, those imposed by modular organization, are relevant to on-line processing but not to off-line processing. All models assume that multiple constraints are used in off-line tasks. Because of this assumption, off-line completions provide a principled and theory-neutral method of setting weights. In their strongest and simplest form, constraint-based models predict that processing difficulty in ambiguity resolution is a function of the strength and availability of the relevant constraints, with no architectural restrictions on their ordering. Therefore, the same weights used to simulate off-line processing for a particular set of materials should also predict on-line processing difficulty. In contrast, the two-stage garden-path model predicts that the on-line data should be better simulated by having the main clause bias apply prior to the other constraints.

NORMING STUDY: GATED SENTENCE COMPLETIONS

The norming study had two primary goals. The first was to establish that our manipulation of thematic fit would influence off-line sentence completions. The second was to provide off-line data for setting weights in the model. Subjects completed sentence fragments that began with an animate noun followed by a verb that was ambiguous between a past tense and a past participle. The animate noun concept was either a typical agent/atypical patient for the verb (henceforth called a good agent, e.g., The cop arrested) or a typical patient/

atypical agent (good patient, e.g., The crook arrested). Sentence fragments continued until either the ambiguous verb (The crook arrested), *by* (The crook arrested by), *the* (The crook arrested by the), or the agent noun completing the *by*-phrase (The crook arrested by the detective). We refer to this task as gated sentence completion because it roughly resembles the gating task developed by Grosjean (1980).

The primary hypothesis was that the frequency of reduced relative completions would be higher for sentence fragments that began with a good patient. The secondary hypothesis concerned the effects of *by*. In a sentence fragment of the form, ''The noun verbed by'', *by* can introduce phrases with a variety of thematic roles, such as ''The woman stood by the door.'' (location), ''The man reacted by instinct.'' (manner), or ''I left by nine o'clock.'' (time). Therefore it was hypothesized that the influence of thematic fit might continue to be evident in fragments gated at *by*. Finally, it was hypothesized that the effect of thematic fit should disappear for fragments gated at the agent because of ceiling effects; it should be difficult for subjects to complete a fragment ending at the agent noun without treating ''verbed by the noun'' as a reduced relative clause.

Method

Subjects

One hundred and twenty native English-speaking University of Rochester undergraduates participated for course credit. Forty-eight completed fragments ending at the verb. Twenty-four subjects completed fragments ending at each of the following: *by, the,* and the agent.

Materials

The verbs and associated nouns used in the sentence fragments were chosen on the basis of role/filler typicality norms described in McRae et al. (1997). The typicality of the nouns as fillers of the agent and patient roles of specific verbs were determined by asking 36 subjects questions that were presented exactly as follows.[2]

How common is it for a

crook _____
cop _____
guard _____
police _____
suspect _____

to **arrest** someone?

or

How common is it for a

crook _____
cop _____
guard _____
police _____
suspect _____

to be **arrested by** someone?

Subjects judged role/filler typicality on a 1–7 scale, where 1 corresponded to a very uncommon event and 7 to a very common one. From among the 96 normed verbs, 40 were chosen for which agenthood and patienthood ratings of two animate nouns were highly polarized (high agenthood rating and low patienthood rating, or vice versa). Agenthood and patienthood ratings for the good agents did not overlap, agenthood: $M = 6.3$, *range* $= 4.4 - 7.0$; patienthood: $M = 2.5$, *range* $= 1.3 - 4.1$, $t(39) = 23.32$, $p < .0001$. The same was true for the good patients, agenthood: $M = 2.0$, *range* $= 1.0 - 3.8$; patienthood: $M = 6.0$, *range* $= 4.0 - 7.0$, $t(39) = 22.76$, $p < .0001$. The good agents were chosen to be reasonable patients because they were used as such in the relative clauses of the self-paced reading study. That is, it is important that the sentences made sense when the good agent was used as the patient, even though the situa-

[2] We thank Maryellen MacDonald and Neal Pearlmutter who collaborated with us in the collection of these role/filler typicality norms.

tion described was less typical (e.g., a cop being arrested rather than doing the arresting).

For each item, four sentence fragments resembling the following were created.

The crook arrested
The crook arrested by
The crook arrested by the
The crook arrested by the detective

Two lists were constructed so that no subject saw any verb twice; the fragments in each list included half of the verbs with their good agent, and the other half with their good patient. One hundred, twenty filler fragments of varying lengths and forms were included in each list so that a minimum of 2 fillers separated target items. To dampen order effects, 12 versions of each list were created by randomly ordering the target items.

Procedure

Subjects were run individually in a paper and pen format. They were instructed to complete all fragments as grammatical and sensible sentences. They were asked to work quickly, to write down what first came to mind, and not to try to think of something clever or funny. It took about 40 minutes to complete the task.

Design

Analyses of variance were conducted on the percentage of reduced relative clause completions. The independent variables were thematic fit (good agent versus patient) and gate (the verb, *by, the,* and the agent). In the analysis by subjects (F_1), thematic fit was within and gate was between. In the analysis by items (F_2), both factors were within.

Results

Responses were scored for all fragments that were completed as syntactically congruous English sentences. Completions meeting this criterion were scored for whether the verb was used as a past tense verb in a main clause or as a past particle in a reduced relative clause. In addition, for some of the sentence fragments that were gated at the verb, the subject wrote ''himself'', ''herself'', or ''themselves'', as in ''The criminal convicted himself.'' These were grouped with the reduced relative completions because subjects were effectively treating the initial NP as a patient and expressing this interpretation by completing the fragment with ''himself'', thus coindexing it with the initial NP. The mean percentage of reduced relative clause completions by condition are presented in Figure 2.

Thematic fit interacted with gate, $F_1(3, 116) = 12.90$, $p < .0001$, $F_2(3, 117) = 39.19$, $p < .0001$. Planned comparisons revealed that the frequency of relative clause completions was significantly higher for good patient fragments that were gated at the verb, $F_1(1, 116) = 53.93$, $p < .0001$, $F_2(1, 117) = 89.95$, $p < .0001$, *by,* $F_1(1, 116) = 77.93$, $p < .0001$, $F_2(1, 117) = 236.37$, $p < .0001$, and *the,* $F_1(1, 116) = 21.84$, $p < .0001$, $F_2(1, 117) = 60.10$, $p < .0001$. However, due to ceiling effects, there was no difference when fragments were gated at the agent noun, $F < 1$ in both analyses. In addition, there was a main effect of thematic fit, in that 77% ($SE = 3\%$) of the fragments with good patients were completed as reduced relatives compared to 59% ($SE = 3\%$) of the fragments with good agents, $F_1(1, 116) = 99.75$, $p < .0001$, $F_2(1, 39) = 126.52$, $p < .0001$. Finally, the frequency of reduced relative completions varied by gate: 12% ($SE = 2\%$) at the verb; 74% ($SE = 3\%$) at *by;* 87% ($SE = 2\%$) at *the;* and 99% ($SE = 0\%$) at the agent; $F_1(3, 116) = 371.67$, $p < .0001$, $F_2(3, 117) = 805.03$, $p < .0001$.

Discussion

Thematic fit clearly influenced the probability of producing a reduced relative clause. This influence was evident in fragments that ended at the verb, *by,* and *the.* When the agent noun in the *by*-phrase was provided, fragments were consistently completed as reduced relatives regardless of thematic fit. Thus it is evident that people possess knowledge about who does what to whom in specific situations and they access this knowledge when generating completions in an off-line task.

The secondary prediction was also borne

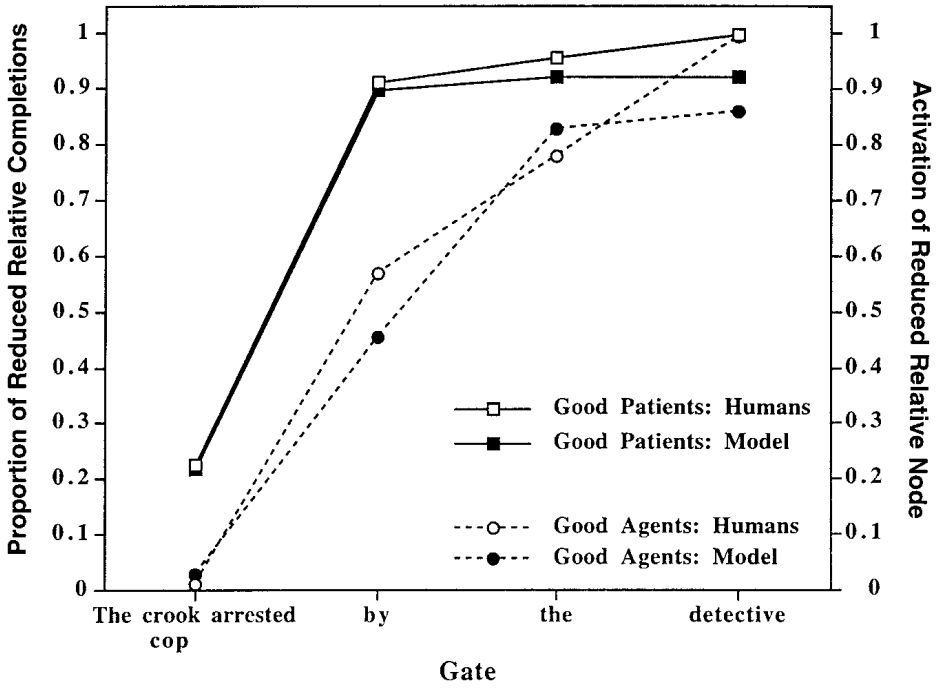

FIG. 2. Human and simulation results for fragment completions.

out; the effect of *by* following the verb was probabilistic. Although there was a sharp increase in the frequency of reduced relative completions at *by,* they were not at asymptote and were influenced by thematic fit. This suggests that many sentence-initial main clause/reduced relative clause ambiguities are not fully disambiguated semantically until the agent noun, and syntactically until the main verb of the sentence. This is particularly clear when the verb is optionally transitive, allowing a manner, location, time, or amount continuation. Thus, in studies that have demonstrated disambiguation effects at the prepositional phrase with reduced relatives (Ferreira and Clifton, 1986; Trueswell et al., 1994; Trueswell, 1996), the disambiguating information may be more appropriately viewed as semantic rather than as syntactic.

The Competition Model

The constraints. In order to model the fragment completion data, each of the relevant constraints was operationalized and assigned a numerical value. A summary of the constraints is provided in Table 1. Three constraints were included for fragments such as *The cop arrested:* a general bias toward the main clause; the thematic fit between the initial NP and the verb's relevant thematic roles (McRae et al., 1997); and the frequency of the verb appearing as a past participle versus a simple past tense (MacDonald et al., 1994b; Trueswell, 1996). Additional constraints were relevant for longer fragments, namely a probabilistic cue supporting the reduced relative from *by* and *the,* and the thematic fit of the noun in the *by*-phrase (good agent, poor location, manner, time, and amount).

The general main clause bias was derived from a corpus analysis reported in Tabossi et al. (1994). For the verbs used in that study, 92% of sentence-initial ''NP verbed'' sequences continued as a main clause, whereas 8% continued as a reduced relative; thus biases of .92 and .08 were used. Thematic fit of the initial NP was operationalized as an integer between 0 and 6 by subtracting 1 from the role/filler typicality ratings described above. Note that Table 1 presents this constraint using the 0–6 scale. The verb tense constraint represents the frequency with which the verb's *ed* form is used as a participle versus as a past tense and was derived from Francis and

TABLE 1

A Summary of the Constraints for the Fragment Completion Simulations

	Good patients		Good agents	
Constraint	Relative	Main	Relative	Main
Initial NP thematic fit	5.0 (.1)	1.0 (.1)	1.5 (.1)	5.3 (.1)
Main clause bias	.08	.92	.08	.92
Verb tense/Voice	.67 (.03)	.33 (.03)	.67 (.03)	.33 (.03)
by-bias	.8	.2	.8	.2
the-bias	.875	.125	.875	.125
Agent NP thematic fit	4.6 (.2)	1	4.6 (.2)	1

Note. Standard errors are in parentheses.

Kucera (1982) corpus frequency measures as shown in Equations 1 and 2, following Trueswell (1996) and Spivey-Knowlton and Tanenhaus (1997).

$$VTV(reduced) = (\log Part/\log Base)/((\log Part/\log Base) + (\log SP/\log Base)) \quad (1)$$

$$VTV(main) = (\log SP/\log Base)/((\log Part/\log Base) + (\log SP/\log Base)) \quad (2)$$

VTV(*x*) is the verb tense constraint for interpretation *x*, *Part* is the verb's frequency as a past participle, *SP* is the verb's frequency as a simple past, and *Base* is the verb's overall frequency, all according to Francis and Kucera. Verb tense/voice support averaged .67 ($SE = .03$, *range* = .26 − .96) for the reduced relative and .33 ($SE = .03$, *range* = .04 − .74) for the main clause, $t(39) = 5.36$, $p < .0001$. As a further indication of the strength of this bias, on a verb-by-verb basis, support was greater by at least .1 for the reduced relative than for the main clause for 21 of the 40 verbs, whereas the reverse was true for only 2 verbs.

For fragments ending at *by*, a *by*-bias was included. The Wall Street Journal and Brown corpora were searched for the 40 verbs used in the norming study (and the self-paced reading study) in order to calculate an independent estimate of the degree to which a *by*-phrase subsequent to *verbed* biases readers toward a reduced relative. There were 124 sentences in which *by* directly followed the *ed* form of the verb, of which 99 were *by*-phrases introducing agents in passive constructions. Thus the *by*-bias was set to .8 (99/124) for the reduced relative clause and .2 (25/124) for the main clause.

For fragments gated at *the*, a *the*-bias was added with strengths of .875 for the reduced relative and .125 for the main clause. These values reflect the fact that in the corpus analysis, 21 of the 24 *by*-phrases that began with *by the* were agentive. Note that this assumes that information provided by *the* is conditionalized on its syntactic environment. Tabor, Juliano, and Tanenhaus (1997) provide experimental and computational explorations of how lexical and category-based constraints interact, focusing on function words in different syntactic environments. For present purposes, it is sufficient to note that (1) the processing system is extremely sensitive to contingent frequencies such as these, and (2) how best to capture these contingent frequency effects remains an important theoretical and empirical issue.

Finally, an additional bias associated with the agent was included for fragments ending at the noun in the *by*-phrase. To quantify it, the fit between the agent noun and the verb's agent

role was measured. Role/filler typicality ratings were conducted in an identical fashion to those described earlier, using a new sample of 20 subjects who responded to questions such as ''How common is it for a detective to arrest someone?'' On a 0–6 scale, agenthood ratings averaged 4.6 and ranged from 1.7 to 5.9. This measure was used as support for a reduced relative interpretation. In contrast, the main clause was supported by a constant bias of 1 that represented the non-agenthood of the agent. The 1 reflected the fact that the agent NPs were poor manners, locations, times, and amounts.

Note that the constraints were conditionalized over different parts of the item set. For those operative at the verb, the main clause bias was the same for all 80 items (40 verbs crossed with good agent versus good patient), verb tense/voice differentiated among verbs, and thematic fit was based on the ratings for each noun-verb combination. In the longer fragments, the *by*-bias and the *the*-bias were constant across all of the items. Finally, thematic fit of the agent noun varied by verb because the same agent was used for the matched good agent and good patient fragments.

Normalized recurrence. The normalized recurrence framework developed by Spivey-Knowlton (1996) was used to model constraint integration at each fragment length. This framework is intended as a generic realization of a model in which (1) multiple constraints are combined to compute alternative interpretations in parallel and (2) the alternatives compete with one another during processing until one achieves criterion activation (see Spivey-Knowlton, 1996, and Spivey-Knowlton & Tanenhaus, 1997, for further discussion).

Information sources were integrated using a three-step normalized recurrence mechanism. First, each of the c informational constraints (three for fragments that ended at the verb) was condensed into its normalized probabilistic support for the a relevant competing alternatives (i.e., the main clause and reduced relative clause).

$$S_{c,a}(\text{norm}) = S_{c,a} / \sum_{a} S_{c,a} \qquad (3)$$

$S_{c,a}$ represents the activation of each constraint node (i.e., the c^{th} constraint that is connected to the a^{th} interpretation node). $S_{c,a}$(norm) is the same variable, but normalized within each constraint. Constraints were then integrated at each interpretation node via a weighted sum based on Equation 4.

$$I_a = \sum_{c} [w_c * S_{c,a}(\text{norm})] \qquad (4)$$

The activation of the a^{th} interpretation node is represented by I_a. The weight on the connection linking the c^{th} constraint node to interpretation node I_a is represented by w_c. Equation 4 was applied to each interpretation node and summed across all constraint nodes that fed into it. Finally, Equation 5 determined how the interpretation nodes sent positive feedback to the constraints commensurate with how responsible the constraints were for the interpretation node's activation. Note that the weights were equal in both directions.

$$S_{c,a} = S_{c,a}(\text{norm}) + I_a * w_c * S_{c,a}(\text{norm}) \qquad (5)$$

These three steps (Equations 3–5) were computed in sequence within each cycle of competition. Thus, as cycles of competition progress, the difference between the two interpretation nodes gradually increases.

Simulation of Sentence Completions

For simulating the completions, because the model attains a steady state only when the activation of one interpretation node reaches 0 and the other reaches 1, it is necessary to stop competition after a prespecified number of cycles and sample the activations of the interpretation nodes at that time. However, it is not clear how to determine the number of cycles that best simulate subjects' behavior in the completion task. Experience with the model showed that it produced reasonable results at approximately 30 cycles of competition. Therefore, we simulated completions at the verb using 20 to 40 cycles, with a step size of 2 (i.e., 20, 22, 24, . . . 40). For each number of cycles, to estimate the weights for

the main clause bias, thematic fit, and verb tense/voice, we tested 941,192 models by varying the weight of each constraint independently from .01 to .99 using a step size of .01. The activation of the reduced relative interpretation node was used to estimate the proportion of reduced relative completions. The root mean square error for completions at the verb was computed for each weight configuration as in Equation 6, where n is the number of points to be predicted.

$$error = \text{sqrt}[\sum (model\ prediction - human\ datum)^2/n] \quad (6)$$

For each number of competition cycles tested, the models were ranked on the basis of root mean square error (0 signifies perfect prediction). The weights for the three constraints were estimated by averaging the weights of the 10 best models from each number of cycles, so that they were averaged over 110 models. The resulting weights were .5094 main clause bias, .3684 thematic fit, and .1222 verb tense/voice. These weights predicted 22.0% reduced relative completions at the verb for good patients, as compared to the 22.4% human completions, and 2.7% for good agents, as compared to 1.0% human completions. Note that the weights are normalized. In fact, the last word in the fragment was always assigned an overall weight of 1, which was apportioned among the constraints that came into play at the word. In this way, we made the simplifying assumption that the weight assigned to the last word in a fragment was equivalent to the sum of the previous constraints. Therefore the *by*-bias, *the*-bias, and thematic fit of the agent noun were all given a weight of 1.

With these weights, the completions were simulated by averaging the activation of the reduced relative interpretation node after 20 to 40 cycles of competition with a step size of 1. Percentage of sentence completions and the activation of the reduced relative node are plotted in Figure 2. The model simulated the major trends in the data. Effects of thematic fit begin at the first fragment. As fragment length increased, both the reduced relative completions and the activation of the reduced relative node increased monotonically. The root mean square error was .071. The best predictions occurred with 31 cycles of competition and for which the root mean square error was .070.

In summary, a model was designed that featured parameters corresponding to the strength of bias each constraint provided for reduced relative and main clause interpretations, and these parameters were independently motivated and estimated. This procedure made it possible to assign weights to a multiple-constraints competition model that closely matched the completion data. This result is not particularly surprising given that the weights at the verb were treated as free parameters and that all current models of sentence processing view off-line ambiguity resolution as using multiple constraints. The crucial question is, however, whether or not these weights will predict on-line reading times. In the Experiment, the model was modified only as much as was necessary to accommodate the fact that a different task--self-paced reading--was being simulated. Thus the model was constrained in that the weights were taken from the off-line simulations. It was then used to generate on-line reading time predictions. In addition, the temporal sequence of the application of constraints was varied in order to compare predictions made by a constraint-based model and variations of a two-stage garden-path model.

EXPERIMENT: SELF-PACED READING

The purpose of this experiment was to collect on-line reading time data for reduced relative clauses that began with the fragments used in the gated completions. These data, in conjunction with different versions of the competition model, were then used to infer whether readers were using thematic fit immediately as predicted by the constraint-based model, or only after a delay, as predicted by two-stage models. Using an explicit model was critical to this endeavor because there is

no model-independent signature data pattern that provides definitive evidence concerning when a specific informational source begins to influence interpretation (Tanenhaus et al., in press). Part of the reason for the lack of such a data pattern is that the interaction among multiple constraints, along with an integration mechanism, must be taken into account in order to predict the effects of any single constraint on reading time for a particular region of a sentence.

Sentences containing reduced and unreduced relatives were created from the fragments used in the gated completions. Thus the items had the following structure (the complete set is listed in the Appendix).

(5a) The cop/arrested by/the detective/was guilty/of taking/bribes.

(5b) The crook/arrested by/the detective/ was guilty/of taking/bribes.

Unambiguous control versions of these sentences were constructed by inserting *who was* before the *ed*-verb, thus disambiguating it as a passive participle. The sentences were presented two words at a time in a moving window format. Two-word presentation was used because it more accurately mimics one key element of reading sentences like these than does single-word moving window presentation (Burgess, 1991). Crucially, eyetracking experiments have shown that subjects usually do not fixate on a short post-verbal preposition such as *by,* rather, it is typically processed parafoveally while the verb is being fixated (Trueswell et al., 1994). Similarly, readers typically fixate on the noun in the *by*-phrase and not on the determiner. The two-word presentation mimics this pattern in that verbed and *by* are presented simultaneously, as are the determiner and agent noun. Comparisons of reading time patterns using self-paced reading and free-field reading with reduced relative clauses like these have found that two-word presentation results in data patterns that are more similar to eye-tracking data than does one-word presentation (Trueswell & Tanenhaus, 1991, 1992; Burgess et al., 1994).

Relevant constraints. Data was collected for three two-word regions: verb + by, *arrested by* in (5); agent NP, *the detective;* and the main verb, *was guilty.* Therefore, regions of competition in the model were segmented accordingly.

Figure 1 shows the constraints used in the self-paced simulations. The weights of .5094 main clause bias, .3684 thematic fit, and .1222 verb tense/voice were identical to the completion simulations. Analogous to the completion simulations, the region currently being simulated was given a total weight of 1. In the self-paced study, *by* was presented with the verb so that its weight had to be combined with those of the other constraints that were directly associated with the verb. If *by* was assigned a weight equal to the sum of all the weights of the constraints associated with the verb, its normalized weight would be .5. This seems unreasonably high because data show that readers typically do one of two things when they fixate on a verb before a *by*. Readers either fixate the verb and pick up *by* parafoveally, often on a second fixation at the verb, or they fixate *by* after fixating the verb.[3] In either case, constraints from the verb will presumably have a stronger influence, at least initially, than constraints coming from *by*. Thus the *by*-bias was given a weight that reflected the fact that it was weighted equally to the sum of the verb biases (as it was in the off-line simulation) but was available for only about half of the reading time in the verb + by region by making it account for 25% of the weighting (i.e., its weight was .25). The main clause, thematic fit, and verb tense/voice biases accounted for the remaining 75% (by multiplying the completion weights by .75).[4]

At the agent NP region, information derived from the agent was introduced. This necessitated confronting the other difference between the off-line and on-line tasks, namely that *the*

[3] The probability of a reader making more than one fixation on our verbs appears relatively high given that the verbs averaged 8.1 letters, with 16 of 40 being 9 letters or longer, and 25 being at least 8 letters long.

[4] Similar patterns of predicted reduction effects are obtained with a normalized *by*-weight ranging from approximately .15 to .35.

NP was presented without the accompanying agent noun in the completion task, but with it in the reading task. Because eyetracking studies have demonstrated that readers typically do not fixate the noun and determiner separately, the biases associated with *the* and the noun must be considered together. In the offline simulation, the *the*-bias provided .875 support for the reduced relative and .125 for the main clause. On average, the agent noun provided a support of 4.6 for the reduced relative and 1 for the main clause, which when normalized equals .82/.18. Because thematic fit and the *the*-bias were similar, dividing the weight at the noun phrase between the two biases would have had a negligible effect on the performance of the model. Therefore we simply used the role/filler typicality ratings to assign item-specific biases to the agent NP. The agent NP constraint was given a weight of 1 and the weights were then normalized prior to competition in the region.

Finally, a bias associated with the syntactically disambiguating main verb was introduced, 1 for the reduced relative and 0 for the main clause. This constraint was also given a weight of 1 and all weights were again normalized prior to competition.

Processing. Competition continued at each region until one of the interpretation nodes reached a criterion level of activation, and activations were carried over to serve as the initial state for the following region. The criterion within a region was dynamic and was a function of the cycle of competition.

$$\text{dynamic criterion} = 1 - \Delta crit * cycle \quad (7)$$

In Equation 7, the constant that controlled the rate of change of the dynamic criterion is represented by $\Delta crit$. The current cycle of competition in a certain region is represented by *cycle*. According to Equation 7, as the duration of competition in a particular region increases, the criterion for stopping competition and moving to the next region becomes more lenient. Competition necessarily terminates in a region when the dynamic criterion becomes .5 because the activation of at least one of the interpretation nodes must be greater than or equal to .5. For example, if $\Delta crit = .01$, the maximum number of competition cycles is $.5/\Delta crit = .5/.01 = 50$. A dynamic criterion is necessary for modeling reading across multiple regions of a sentence because fixation durations are partially determined by a preset timing program (Rayner & Pollatsek, 1989; Vaughan, 1983). In other words, a reader will spend only so long on a fixation before making a saccade. Presumably, this same logic holds for self-paced reading in that readers attempt to resolve competition at each two-word segment for only so long before pressing the space bar for more information. It does not make sense for readers to expect ambiguity to be fully resolved at each point in a sentence; reading processes are presumably sensitive to the fact that language contains numerous local ambiguities that are typically disambiguated by subsequent input.

In the remaining simulations, cycles of competition were mapped onto differences in reading times between the reduced relatives and the full relatives. This procedure was used under the assumption that the unreduced versions of the sentences were unambiguous variations of the same structure, thus providing the proper baseline against which to measure the effects of competition. Thus reading time differences due to processes not stimulated by the model are factored out and there is no need to incorporate variables such as word length and frequency. This approach can be used because *who was* and *that was* unambiguously signal a reduced relative and the sentences were constructed so that the critical regions of the reduced and unreduced versions were identical. Variants of the model were used to generate predictions for a constraint-based model and a two-stage model with delayed thematic and lexical constraints. Predictions from each of these models are described in the next sections.

Constraint-based predictions. For simulating the self-paced reading experiment, it is not clear what $\Delta crit$ should be used. It is necessary to choose a $\Delta crit$ that is not too large so

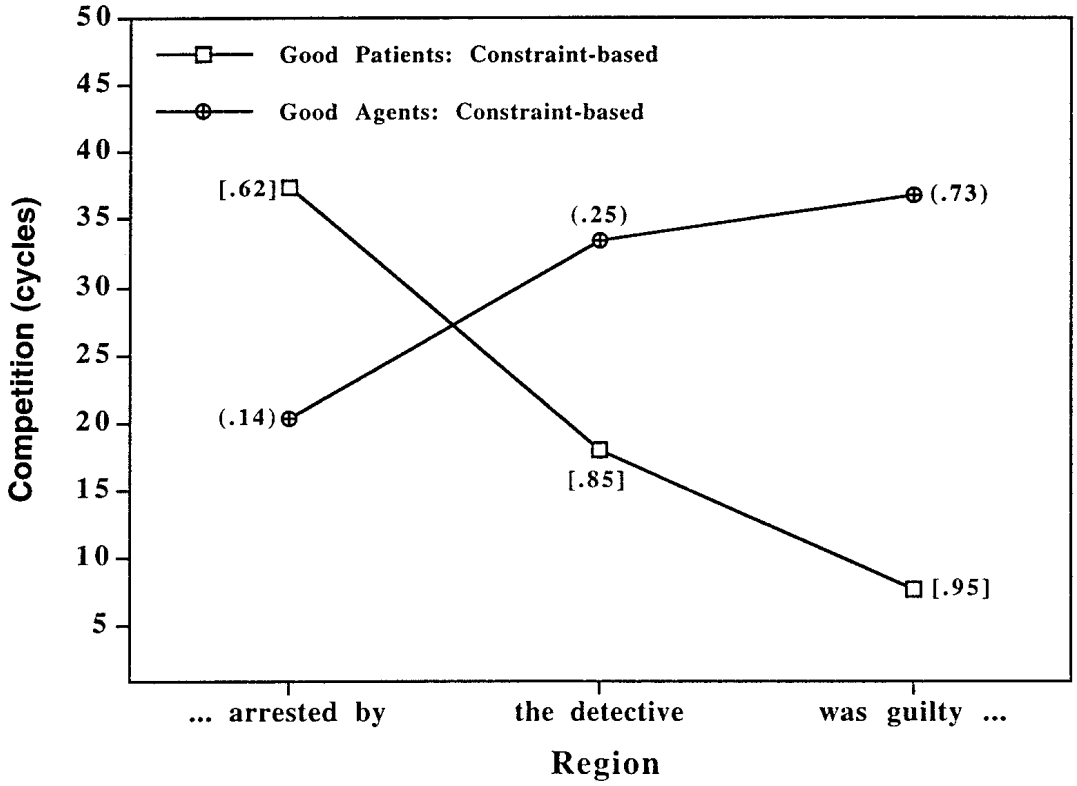

FIG. 3. Self-paced predictions derived from the constraint-based competition model. In this and all following model figures, the number beside each model datum is the mean activation of the reduced relative node after competition in that region for either (good agents) or [good patients].

that differences among items can emerge. If it is too large, competition stops too quickly in a region and ceiling effects result. Similar modeling by Spivey-Knowlton and Tanenhaus (1997), who used $\Delta crit = .01$ when stimulating eyetracking, can be used as a guide to choosing an appropriate $\Delta crit$. Given that reading times are slower in self-paced reading than in eyetracking experiments, a $\Delta crit$ somewhat less than .01 seems appropriate. To generate predictions for the self-paced reading study, constraint-based simulations were conducted by varying the $\Delta crit$ parameter from .006 to .009 with a step size of .0001.

The results of averaging these 31 simulations are shown in Figure 3. Plotted are the number of cycles of competition averaged over the 40 good patient or 40 good agent sentences. Note that one cycle of competition indicates an absence of ambiguity. The numbers in square brackets [good patients] and round parentheses (good agents) signify the mean activation of the reduced relative node at the end of competition in that region. It varies from 0, a confident main clause interpretation to 1, a confident reduced relative interpretation.

On the surface, the predicted reduction effects presented in Figure 3 might appear to directly oppose what would intuitively be predicted from a model that immediately integrates all relevant sources of information. In particular, intuition suggests that there should be no reduction effect for good patients in any region if thematic fit is a strong cue for comprehension. However, a closer investigation of the relevant constraints suggests otherwise. In the verb + by region, the clause bias supported the main clause interpretation. On the other hand, the *by*-bias supported the reduced relative as did verb tense/voice. Because the thematic fit of a good patient additionally supported the reduced relative, initial activations of the competing interpretation nodes tended to be similar. This resulted in substantial competition in the region and acti-

vation levels following competition tended to be close to .5. In the present simulation, the mean activation for good patients was .62 after the verb + by competition, indicating an interpretation that slightly favors the reduced relative. That activation rose to .86 in the agent NP region so that there was very little competition when the main verb was encountered.

Conversely, because thematic fit of a good agent supported the main clause in conjunction with the general clause bias, relatively brief competition ensued and activation values following competition tended to highly favor the main clause. The mean reduced relative activation for the good agents was .20, representing a confident main clause interpretation. After competition in the agent NP region, reduced relative activation for the sentences with good agents remained well below .5. Only following substantial competition in the main verb region did good agent sentences tend to settle toward a reduced relative interpretation.

Garden-path predictions. Two-stage garden-path model predictions were derived using a variation of the competition model in which all constraints except the purely structural main clause and main verb biases were delayed by one region. The remainder of the constraints can be classified as conceptual knowledge or as linguistic knowledge associated with specific lexical items, both of which are delayed in the garden-path model. The *by*-bias represents information about a specific lexical item and the probability that it introduces certain thematic roles. Verb tense frequency is item-specific information. Thematic fit is based on world knowledge of events.

For the simulations presented in Figure 4, the main clause bias was operative during competition in the verb + by region and was given a weight of 1. This has the effect of instantiating Frazier and Fodor's (1978) Minimal Attachment Principle as applied to the main clause/reduced relative ambiguity for a one-region window. Note that thematic fit and verb tense/voice were delayed by two words, given that the region included the verb and *by*. The *by*-bias was delayed by one word. At the agent NP region, thematic fit of the initial NP, verb tense/voice, and the *by*-bias were included with the main clause bias and their weights returned to those used in the constraint-based simulations. Thematic fit of the agent noun was delayed until the main verb region, at which time it was given a weight of 1 as it had been in the constraint-based simulations. Because the delayed information concerns the noun, the final word in the region, it corresponds to a one-word delay. This configuration captured the spirit of the garden-path model in that the nonstructural information was delayed by a region. Later, the effects of decreasing this temporal window are examined.

As in the constraint-based simulations, predictions were averaged over $\Delta crit$'s ranging from .006 to .009. Figure 4 presents a pattern of competition that differs substantially from the constraint-based model. (Note that the scales are identical for Figures 3 and 4.) In particular, the garden-path model predicts no reduction effect for good patient sentences in the verb + by region, followed by a large effect at the agent NP and another at the main verb. For good agents, a linear increase in reduction effects through the main verb region was predicted.

In summary, predictions were derived from constraint-based and two-stage models of ambiguity resolution. In this experiment, self-paced data was collected using the stimuli from the norming study in order to test these predictions on-line.

Method

Subjects

Forty University of Rochester undergraduate students were paid $5. All were native English speakers and had normal or corrected to normal vision.

Materials

Each fragment pair from the completion study was used to generate four sentences, as illustrated in the examples (5a) and (5b) with the third and fourth sentences created by inserting *who was* or *that was* prior to the past

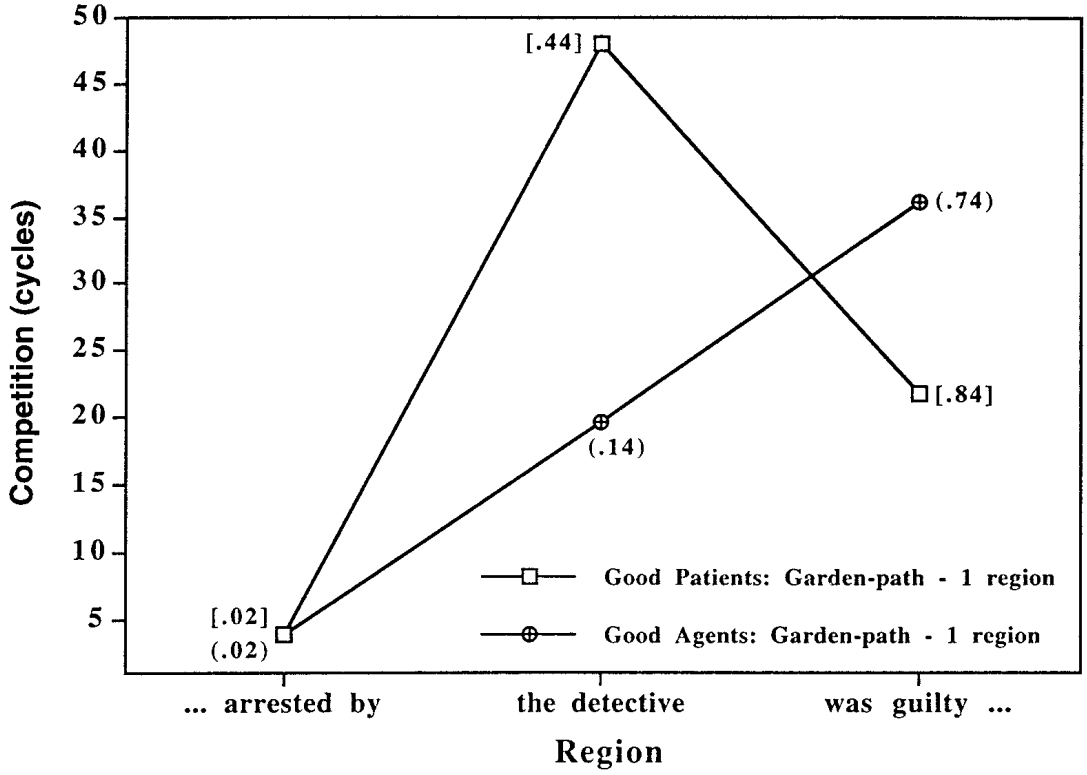

FIG. 4. Self-paced predictions as derived from the garden-path model when constraints other than the main clause and main verb biases were delayed by a region.

participle. A second sentence was added to each target in order to create a more meaningful and natural narrative and to decrease the proportion of reduced relative clauses. The complete set of 40 sentence pairs appears in the Appendix. Care was taken to create plausible, natural sounding sentences even when a good agent was used as the patient of the past participle. This is important because the presence of implausible sentences is unnatural and might cause subjects to reject thematic fit as a cue.

Four lists were constructed so that no subject saw any verb or critical noun phrase more than once. Each list contained all 40 verbs, half with good agents and half with good patients. Half of the target sentences included a reduced relative, and half retained the overt relative clause marker. Ninety-six two-sentence filler items were constructed and interleaved throughout the target sentences such that no two targets were adjacent, and each block of trials began with a filler so that 15% of the sentences contained a relative clause. Ten additional two-sentence practice items were also constructed that did not contain relative clauses.

Procedure

An experimental session began with the ten practice items. Each trial began with the trial number displayed in the upper left-hand corner of the screen. A button press revealed the sentence pair, with each letter replaced by a dash. With further button presses, subjects read sentences presented two words at a time in a moving-window format (Just, Carpenter, & Woolley, 1982). Following the practice session was a block of trials containing half of the 144 sentence pairs, including 10 good agents and 10 good patients. After a break, the remaining 72 were presented. Subjects were also instructed that they could take a break between items by delaying their button press when the trial number was visible. For each of the four lists, order of the two blocks was counterbalanced. The first line of each target item on the CRT display contained the critical regions plus one (up to ''of taking''

in 6). Approximately one third of the 144 trials were followed by yes–no comprehension questions. Subjects were instructed to read at a comfortable pace that best approximated their natural reading style. The experimental session lasted approximately 30 minutes. Sentences were presented on, and button pressing latency was measured by, an IBM-AT clone containing a Digitry timing board.

Design

Reading latencies were analyzed based on the word pairs presented to subjects. The three critical regions are shown below.

(6) The cop / arrested by / the detective / was guilty / verb + by / agent NP / main verb / of taking / bribes.

Analyses of variance were conducted on the reading latencies. There were two factors, each with two levels: thematic fit (good agent versus good patient); and reduction (unreduced versus reduced). Both factors were within subjects and items. Separate analyses were conducted on reading latencies for each of the critical two-word regions.

Results

Reading times for the four conditions in the three regions are presented in Figure 5. Thematic fit clearly interacted with the magnitude of the ambiguity (reduction) effects. For the sentences with good patients, reduction effects declined from 61 ms at the verb + by region, to 32 ms at the NP, and −1 ms at the main verb. In contrast for the sentences with good agents, the reduction effect increased from 24 ms at the verb + by region, to 42 ms at the NP and 44 ms at the main verb. The effects were examined in separate ANOVAs for each region.

Verb + by Region

There was an interaction between thematic fit and reduction, $F_1(1, 47) = 5.13$, $p < .03$, $F_2(1, 39) = 4.78$, $p < .04$. Planned comparisons revealed that subjects experienced difficulty with both types of sentences in that there was a 24 ms reduction effect with good agents, $F_1(1, 47) = 4.59$, $p < .04$, $F_2(1, 39) = 4.82$, $p < .05$, and a 61 ms effect with good patients, $F_1(1, 47) = 28.56$, $p < .0001$, $F_2(1, 39) = 26.65$, $p < .0001$. The interaction resulted from the larger reduction effect for the sentences with good patients. There was no main effect of thematic fit, $F < 1$ in both analyses. Finally, mean reading time was 42 ms shorter for unreduced sentences than for reduced sentences, $F_1(1, 47) = 27.50$, $p < .0001$, $F_2(1, 39) = 30.15$, $p < .0001$.

Agent NP Region

Thematic fit influenced reading times in that there was a 42 ms main effect, $F_1(1, 47) = 8.24$, $p < .007$, $F_2(1, 39) = 11.26$, $p < .002$. There was no thematic fit by reduction interaction, $F < 1$ in both analyses. Both planned comparisons were significant: there was a 44 ms reduction effect for good agents, $F_1(1, 47) = 12.97$, $p < .0009$, $F_2(1, 39) = 10.58$, $p < .003$, and a 32 ms reduction effect for good patients, $F_1(1, 47) = 7.26$, $p < .01$, $F_2(1, 39) = 5.92$, $p < .02$. Finally, there was a 38 ms main effect of reduction, $F_1(1, 47) = 15.28$, $p < .0004$, $F_2(1, 39) = 11.54$, $p < .002$.

Main Verb Region

Thematic fit and reduction interacted, $F_1(1, 47) = 5.84$, $p < .02$, $F_2(1, 39) = 4.77$, $p < .04$. Planned comparisons revealed a significant 49 ms reduction effect for good agent sentences, $F_1(1, 47) = 11.22$, $p < .002$, $F_2(1, 39) = 9.17$, $p < .005$, in contrast to a −1 ms reduction effect for good patient sentences, $F < 1$ in both analyses. The 14 ms main effect of thematic fit was not reliable, $F_1(1, 47) = 1.32$, $p > .2$, $F_2(1, 39) = 1.42$, $p > .2$. Finally, there was a marginally significant 24 ms main effect of reduction, $F_1(1, 47) = 4.64$, $p < .04$, $F_2(1, 39) = 3.31$, $p > .07$.

Discussion

Knowledge of the fit between a noun concept and the agent and patient roles of a specific verb exerted a strong influence on ambiguity resolution. Given that both the noun and verb concepts must be available for thematic fit to be computed, the interaction at the verb + by demonstrates rapid computation and in-

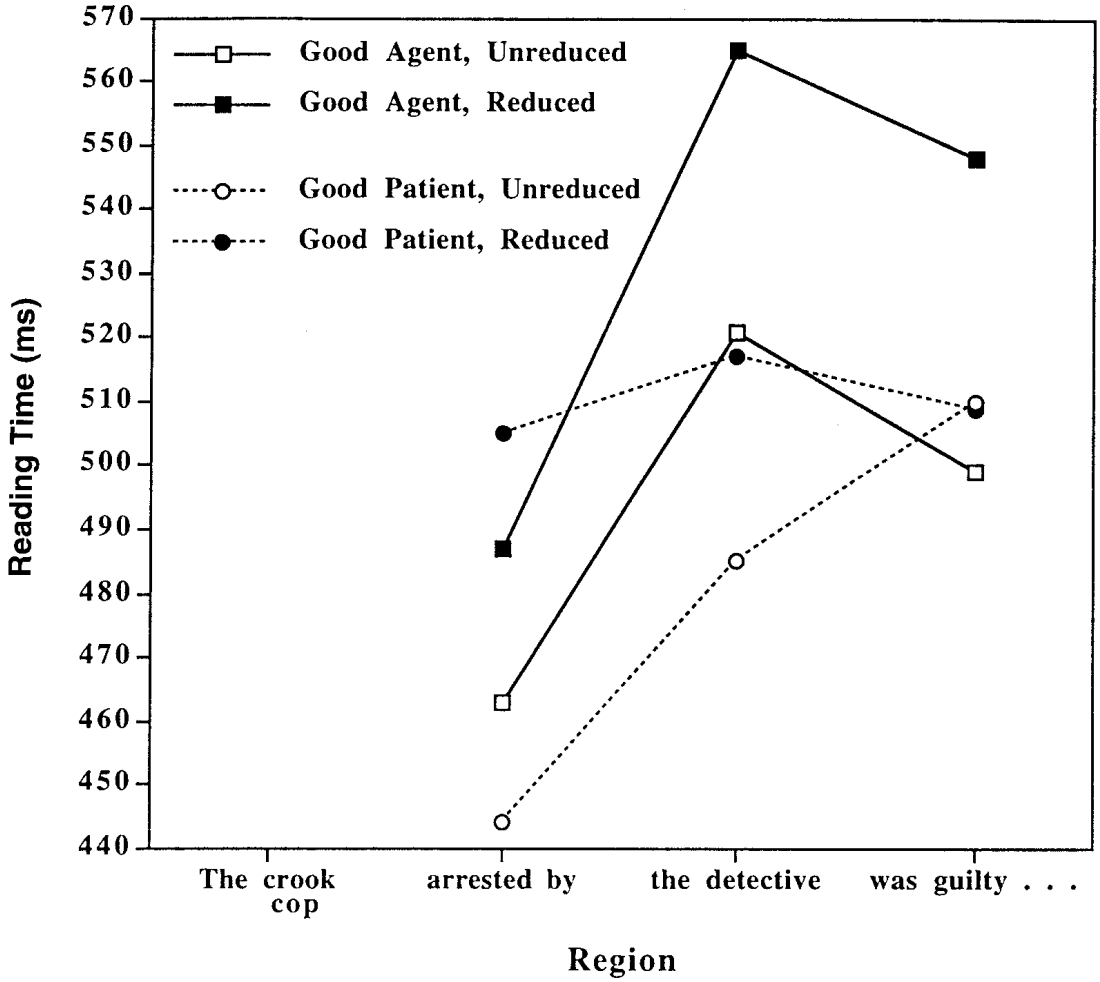

FIG. 5. Self-paced reading times for the Experiment.

tegration of this information. Because animacy was held constant across the manipulations of thematic fit, the effects are clearly due to conceptual knowledge and cannot be explained as selectional restriction effects (Caplan et al., 1994). These results support the hypothesis that thematic fit is computed using verb-specific conceptual representations, consistent with a conceptually-based theory of thematic roles such as that proposed by McRae et al. (1997).

Figure 6 shows the predictions of the constraint-based and two-stage models along with the reduction effects. Note that the numbers are slightly different than those in Figures 3 and 4. Rather than choosing a plausible range of $\Delta crit$'s as was necessary for generating predictions, the models were compared by using the 25 $\Delta crit$'s that minimized the root mean square error between the observed human reduction effects and each model's predictions.[5] It is clear from Figure 6 that the predictions of the constraint-based competition model closely resemble the human data, but the predictions of the garden-path model deviate substantially. A two-tailed independent groups *t*-test that used the 25 lowest root mean square errors as the values of the dependent variable

[5] To compute the root mean square error for the online simulations, it was necessary to convert the reduction effects in ms to the same scale as cycles of competition. This was accomplished using the scale of Figs. 6 and 7. That is, each reduction effect was divided by 90 (the scale was 0–90), multiplied by 59 (the scale was 1–60), and then 1 was added to the result because 1 cycle of competition corresponds to no reduction effect. Note that although this transformation is somewhat arbitrary, it is bias-free in terms of the alternative models tested. This is important because although the root mean square error is not highly informative in terms of any single model, it is valuable for comparing models.

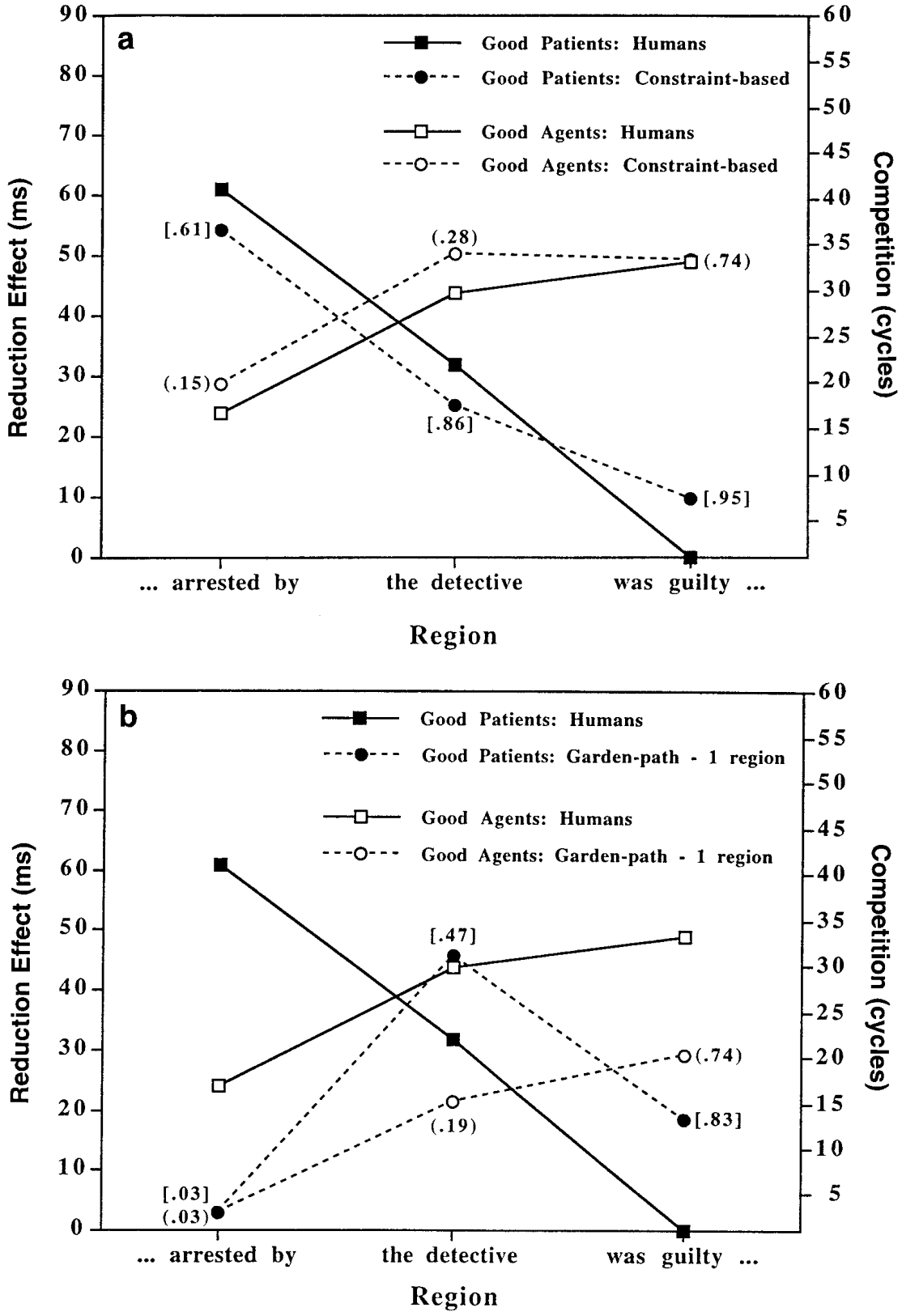

FIG. 6. Simulations of self-paced reading by (a) the constraint-based model, and (b) the one-region delay garden-path model.

confirmed that the constraint-based model ($M = 4.60$, $SE = 0.38$) was a better predictor of the self-paced reading data than was the two-stage model ($M = 19.35$, $SE = 0.20$), $t(48) = 195.97$, $p < .0001$. The best constraint-based models were those with $\Delta crit$ in the range

.0068 to .0092, somewhat below the .01 that Spivey-Knowlton and Tanenhaus (1997) used to simulate eyetracking. The best garden-path models occurred from .0122 to .0146.

Garden-Path Modeling with Shorter Delays

In the garden-path simulation, we assumed that all constraints except for the main clause and main verb biases would be delayed by a region (two words for the verb information, and one word for the *by* and agent noun information). It is possible, however, to argue that evaluation and revision are more rapid than that. Of course, as the delay between structural and non-structural constraints becomes smaller, the two-stage model begins to approximate a constraint-based model in which multiple constraints are applied simultaneously. Nonetheless, it is possible that a shorter delay version of a two-stage model would provide a better fit to the experimental data than a constraint-based model. This is just the kind of evidence that would provide support for the hypothesis that non-structural constraints are delayed. In order to evaluate this possibility, we simulated the self-paced reading task using a two-stage model that featured shorter informational delays. Information other than the main clause bias was delayed for four cycles at the verb + by, and thematic fit of the agent noun was delayed for four cycles in the agent NP region. Four cycles was chosen because this was the number required for the garden-path model to reach criterion in the verb + by region. If cycles of competition are mapped onto reduction effects, a reasonable estimate is that a four-cycle delay corresponds to approximately 10–25 ms. Simulations were again conducted across a wide range of $\Delta crit$'s. The 25 best short-delay two-stage simulations occurred with $\Delta crit$'s of .0109 to .0133. As shown in Figure 7, with a four-cycle delay, the short-delay model predicts the human data much better than does the garden-path model with a one-region delay. The major difference between the human data and the short-delay model is that the predicted reduction effects remain quite large for good patient sentences through the main verb; the model can not compensate for the delay of information. Although it performed better than the one-region delay model, a two-tailed independent groups *t*-test using the 25 best root mean square errors as the dependent variable showed that the constraint-based model ($M = 4.60$, $SE = 0.38$) was a better predictor than was the short delay model ($M = 5.34$, $SE = 0.15$), $t(48) = 9.11$, $p < .0001$. In addition, the best constraint-based model, which occurred at $\Delta crit = .0078$ had a root mean square error of 4.15, as compared to the best two-stage model, which occurred at $\Delta crit = .0122$ and had a root mean square error of 5.15.

We also implemented a two-stage model that maximized fit to the self-paced data by delaying information for only four cycles and treating all six weights as free parameters. When the weight and $\Delta crit$ spaces were searched, the best two-stage short-delay model had weights of .2966 main clause bias, .4661 thematic fit, .0254 verb tense/voice, .2119 *by*-bias, 1.1 agent noun thematic fit, and 2.6 main verb bias. The best fit was found at $\Delta crit = .0080$ and had a root mean square error of 2.77. The short-delay, weights as free parameters model overcame the delay of conceptual and lexically-specific information by increasing the initial NP thematic fit weight from .3684 to .4661 while decreasing the main clause bias weight from .5094 to .2996, increasing the agent noun thematic fit weight from 1 to 1.1, and more than doubling the main verb weight from 1 to 2.6. An important test of the consequence of these weight changes is whether this model is capable of simulating the off-line completion data. Table 2 presents the simulations of the completions at the verb. Because of the initial NP and main clause bias weight changes, no number of competition cycles allows the predicted proportion of reduced relative clause completions to approach the observed human data for the good patient items (.22).

Finally, the same method was used with the constraint-based model. When the weight and $\Delta crit$ spaces were searched, the best constraint-based model had weights of .4062 main clause bias, .2083 thematic fit, .0104 verb

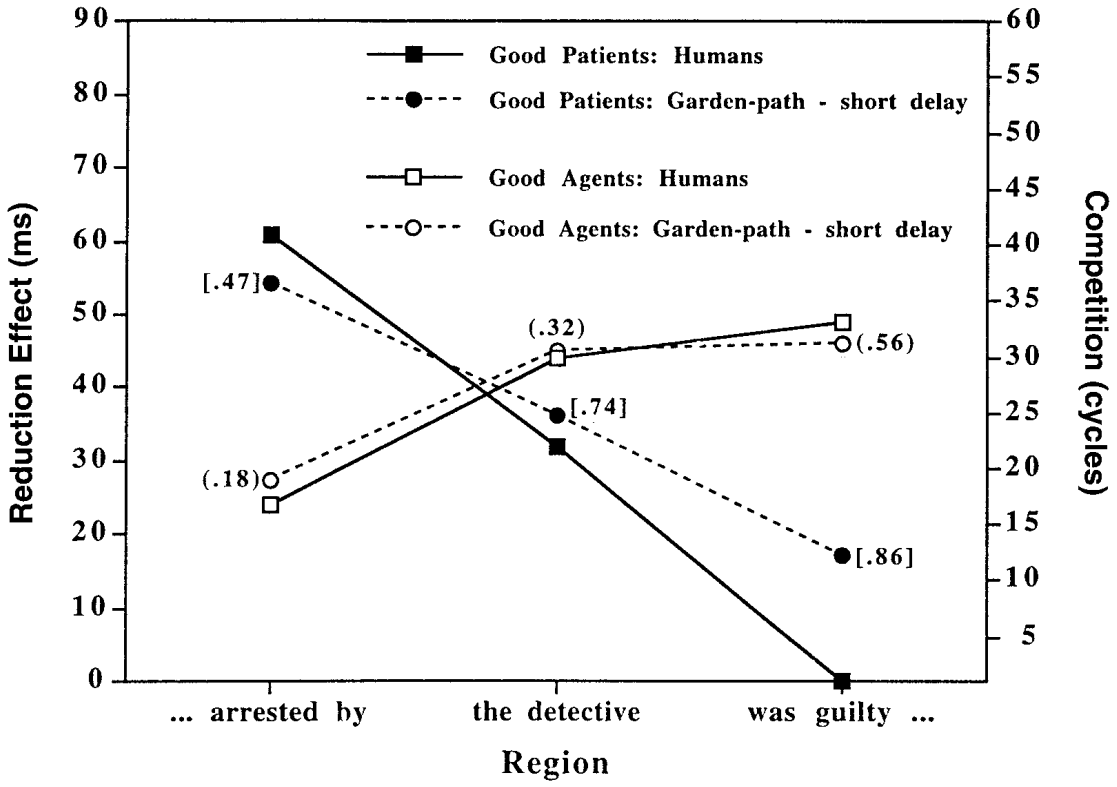

FIG. 7. Garden-path simulations of self-paced reading when constraints other than the main clause and main verb biases were delayed by 4 cycles of competition, or approximately 10–25 ms. Note the large predicted reduction effect at the main verb for the good patients.

tense/voice, .3750 *by*-bias, 1.1 agent noun thematic fit, and 2.1 main verb bias. The best fit was found at $\Delta crit = .0058$ and had a root mean square error of 2.61, which is comparable to that of the best free-weights short-delay garden-path model. When these weights were used to simulate the off-line completion data at the verb, the best fit was found at 12 iterations, when the model predicted completion percentages of .02 for the good agents and .22 for the good patients, as compared to the human completion proportions of .01 and .22 respectively.

TABLE 2

Using the Free Parameter, Short-Delay Garden-Path Model to Predict the Fragment Completions

Number of cycles	Activation of reduced relative node	
	Good patients	Good agents
1	.55	.18
2	.55	.15
3	.55	.12
4	.56	.10
5	.56	.08
10	.56	.03
20	.57	.02
30	.66	.02
40	.71	.01
50	.72	.01

Note. The human completion proportions were .22 for the good patients and .01 for the good agents.

In summary, the constraint-based model represents the null hypothesis, namely that the time-course with which constraints are used on-line is similar to how they are used off-line. In the absence of a compelling empirical reason to introduce the additional assumptions required by a two-stage model, there is no reason to reject the simpler constraint-based account.

GENERAL DISCUSSION

A combination of experiments and modeling was used to investigate the time-course with which semantic/conceptual information influences syntactic ambiguity resolution. Via typicality ratings, we determined the extent to

which an animate entity, as denoted by a noun, fits a semantic role of an event, as denoted by a verb. Knowledge about entities' participation as agents and patients in specific events had clear effects on syntactic ambiguity resolution, as measured by fragment completions and self-paced reading. Because thematic fit was varied independently of selectional restrictions, these results provide clear evidence that conceptually-based information is used immediately in ambiguity resolution.

The results were simulated using an architecture in which multiple constraints provided probabilistic support for competing alternatives. The constraint values were determined from a combination of ratings and corpus analyses and the weights were set to fit the completion data. We then used the model to make predictions about ambiguity effects in self-paced reading. With this procedure, we were able to evaluate the claim that multiple constraints are applied simultaneously, as well as contrast it with the claim that some constraints are delayed in on-line processing. No evidence for delayed use of constraints was found in that predictions of the constraint-based model provided a better fit to the reading time data than did variants of the model in which nonstructural constraints were delayed.

In conjunction with other recent work adopting a similar approach (Spivey-Knowlton & Tanenhaus, 1997, Tanenhaus et al., in press, Hanna, Spivey-Knowlton, & Tanenhaus, 1997), the studies presented here provide strong support for models of ambiguity resolution in which multiple constraints are combined probabilistically and continuously in order to compute the best interpretation from among grammatically consistent alternatives (MacDonald et al., 1994a; Spivey-Knowlton et al., 1993). The normalized recurrence architecture used here is one instantiation of a mechanism that embodies these properties. This architecture is also valuable in that it does not introduce biases in favor of constraint-based models. That is, there are many data patterns that would provide clear evidence for delayed use of constraints as long as the constraints and their weights are fixed. However, when weights are treated as free parameters as they often are in absence of explicit integration models, there is no data pattern that can provide unambiguous evidence for either delayed or immediate use of constraints because a particular data pattern will almost always be compatible with different interpretations.

In conclusion, it is important to consider both the strengths and limitations of the current approach to modeling the processing difficulty associated with syntactic ambiguity resolution. First, the use of local interpretation nodes and independent constraints abstracts away from important issues about how people learn and represent contingent constraints and knowledge about the world. As such, the models share some limitations with localist connectionist networks. Second, it is important to keep in mind that what we have implemented is a model of constraint integration during ambiguity resolution. Thus the model cannot account for sources of variation in processing time that are due to generating syntactic alternatives. This is clearly an important limitation. However, it is also important to note that ambiguity resolution is in itself a central component of language comprehension. Moreover, it has served as the primary testing ground for evaluating conflicting claims about parsing theories, typically under conditions where a temporarily ambiguous structure is compared to an unambiguous baseline. This is precisely the domain in which the model is most appropriate. Finally, the success of this style of modeling is going to depend crucially on the extent to which the constraints and biases are independently motivated. The work presented here clearly captured some but not all of the relevant constraints for the reduced relative construction, and some of the procedures for estimating the constraint values could be improved in future work (see Hanna et al., 1997, for steps in this direction). Although a great deal of work is required to do so, providing constraints and weights as we have done makes it possible to conduct explicit quantitative tests of some of the crucial assumptions

made by alternative theories of sentence processing.

APPENDIX

The 40 target items for the self-paced reading study are presented along with the agenthood (A) and patienthood (P) role/filler typicality ratings. Each (a) item is the typical agent/atypical patient and each (b) item is the typical patient/atypical agent.

1. a) The postman carried by the paramedics was having trouble breathing. He suffered from asthma, but this attack was worse than usual. A: 6.8 P: 1.6
   b) The newborn carried by the paramedics was having trouble breathing. It looked as though something was stuck in her throat. A: 1.2 P: 6.2
2. a) The waitress served by the trainee was displeased with his attitude. He complained to the manager about her lack of enthusiasm. A: 6.8 P: 2.5
   b) The customer served by the trainee was displeased with his attitude. He complained to the manager about her lack of enthusiasm. A: 1.5 P: 7.0
3. a) The reporter interviewed by the editor was unfit for the job. Although his resume looked promising, the editor found him to be immature and egotistical. A: 6.8 P: 2.7
   b) The candidate interviewed by the editor was unfit for office. Although the interview went well, many of his promises turned out to be lies. A: 2.6 P: 6.6
4. a) The employer fired by the owner was jobless for several months. He had worked no other jobs, and now found himself without a recommendation. A: 6.1 P: 2.4
   b) The employee fired by the owner was jobless for several months. He had worked no other jobs, and now found himself without a recommendation. A: 1.9 P: 6.4
5. a) The cop arrested by the detective was guilty of taking bribes. He was let off because no one would testify against him. A: 6.7 P: 1.6
   b) The crook arrested by the detective was guilty of taking bribes. They had been following him for months and finally had the evidence to convict him. A: 1.2 P: 5.9
6. a) The boss hired by the corporation was perfect for the job. In addition to his great resume, he was a damn nice guy. A: 6.7 P: 2.9
   b) The applicant hired by the corporation was perfect for the job. In addition to his great resume, he was a damn nice guy. A: 1.3 P: 5.6
7. a) The coach instructed by her mentor had made similar career decisions. They both ended up at Ivy League schools. A: 6.7 P: 2.1
   b) The novice instructed by her mentor had made several judgment errors. She was still learning to cope with the fast pace of the game. A: 1.6 P: 5.3
8. a) The monster frightened by the flames was moving away as quickly as it could. He ran back to Frankenstein's castle where he felt safe. A: 6.4 P: 1.8
   b) The baby frightened by the flames was moving away as quickly as it could. Her mother picked her up and tried to calm her down. A: 2.0 P: 5.5
9. a) The hunter shot by the teenager was only thirty years old. He left his wife and child penniless. A: 6.9 P: 2.8
   b) The deer shot by the teenager was only used as a trophy. He was the biggest buck killed that year. A: 1.0 P: 6.4
10. a) The inspector interrogated by the FBI was found guilty of drug trafficking. He had been obtaining cocaine from a narcotics officer. A: 6.3 P: 1.6
    b) The suspect interrogated by the FBI was found guilty of drug trafficking. He had been obtaining cocaine from a narcotics officer. A: 1.7 P: 5.8
11. a) The farmer slaughtered by the mercenaries was completely surprised by their attack. He had been providing information to the government for months, but had no idea that anyone

suspected him. A: 5.9 P: 1.4

b) The pig slaughtered by the mercenaries was completely eaten in one meal. They had been without food for a number of days. A: 1.0 P: 6.8

12. a) The knight rescued by the soldiers had been humiliated by his captors. They had forced him to beg for mercy on several occasions. A: 6.0 P: 1.7

b) The hostage rescued by the soldiers had been humiliated by his captors. They had forced him to beg for mercy on several occasions. A: 1.4 P: 5.2

13. a) The organizer invited by the trustees gave a speech at the luncheon. This was the third time she had organized this conference in the last five years. A: 5.0 P: 3.4

b) The guest invited by the trustees gave a speech at the luncheon. She presented her firm's proposal concerning new methods of mass transit. A: 1.9 P: 6.4

14. a) The psychoanalyst hypnotized by his therapist cried like a baby. Clearly, the abuse he had suffered as a child was still with him. A: 5.8 P: 2.0

b) The patient hypnotized by his therapist cried like a baby. Clearly, the abuse he had suffered as a child was still with him. A: 1.3 P: 5.3

15. a) The kidnapper tortured by the CIA should have known better than to kidnap a senator's daughter. He was never seen again. A: 5.7 P: 1.6

b) The slave tortured by the CIA should have known better than to conspire against a U.S.-supported South American dictator. He was never seen again. A: 1.3 P: 5.6

16. a) The judge sentenced by the state was associated with organized crime. He ended up spending ten long years in a maximum security prison. A: 6.9 P: 1.3

b) The criminal sentenced by the state was associated with organized crime. He ended up spending ten long years in a maximum security prison. A: 1.3 P: 6.7

17. a) The lawyer questioned by the sergeant was being overly defensive. The sergeant's doubts about her honesty increased throughout the interview. A: 6.5 P: 2.9

b) The witness questioned by the sergeant was being overly defensive. The sergeant's doubts about her honesty increased throughout the interview. A: 2.0 P: 6.7

18. a) The doctor cured by the treatment had actually developed it in the first place. It was quite ironic that he had benefited so directly from it. A: 6.8 P: 3.8

b) The patient cured by the treatment had actually been diagnosed as terminal by several doctors. Needless to say, he was ecstatic as he left the hospital. A: 1.4 P: 6.1

19. a) The pirate terrorized by his captors was freed when it became obvious that he knew nothing. It was an ordeal that he would not forget. A: 6.5 P: 2.2

b) The victim terrorized by his captors was freed when it became obvious that he knew nothing. It was an ordeal that he would not forget. A: 1.4 P: 6.6

20. a) The juror convicted by the judge had leaked restricted information to the press. Somehow, a television reporter had gotten to him during the trial. A: 6.6 P: 1.3

b) The criminal convicted by the judge had leaked restricted information to the press. Somehow, a television reporter had gotten to him during the trial. A: 1.4 P: 5.9

21. a) The teacher graded by the principal scored highly in all areas. He had basically been successful at everything he attempted. A: 6.9 P: 2.6

b) The pupil graded by the principal scored highly in all areas. He had basically been successful at everything he attempted. A: 2.3 P: 6.8

22. a) The witness accused by the lawyer was trying to cover for his wife. Although he knew she had committed

the murder, he felt that he couldn't live without her. A: 6.4 P: 3.2

b) The suspect accused by the lawyer was trying to cover for his wife. Although he knew she had committed the murder, he felt that he couldn't live without her. A: 3.3 P: 6.8

23. a) The stripper entertained by the comedian laughed despite her troubles. He was good at finding humor in even the most mundane activities. A: 7.0 P: 2.2

b) The audience entertained by the comedian laughed despite their troubles. He was good at finding humor in even the most mundane activities. A: 1.7 P: 6.7

24. a) The committee evaluated by the reviewers stood alone above all others. Their recommendations were not only socially responsible, but fit within the budget as well. A: 6.3 P: 3.9

b) The applicant evaluated by the reviewers stood alone above all others. Not only was she appropriately educated for the job, but she also had relevant experience. A: 3.3 P: 5.7

25. a) The witness recognized by the photographers was trying to get away before they took any pictures. They had been waiting by the backdoor of the courthouse for several hours. A: 6.1 P: 3.9.

b) The celebrity recognized by the photographers was trying to get away before they took any pictures. He felt unwell and didn't want to be bothered. A: 3.8 P: 6.8

26. a) The snake devoured by the tribesmen had been roasting over the coals all afternoon. It was the centerpiece of their ritual meal to celebrate the new moon. A: 6.2 P: 3.9

b) The rabbit devoured by the tribesmen had been roasting over the coals all afternoon. It was the centerpiece of their ritual meal to celebrate the new moon. A: 3.3 P: 5.7

27. a) The donkey kicked by the farmer had refused to keep working. The farmer did not enjoy punishing the animal, but the work had to be done. A: 6.1 P: 3.4

b) The wimp kicked by the farmer had refused to keep working. The farmer did not enjoy punishing his son, but the work had to be done. A: 1.6 P: 5.4

28. a) The hangman executed by the government had been convicted of treason. They discovered that he had faked the execution of a leading subversive. A: 5.1 P: 3.9

b) The martyr executed by the government had been convicted of treason. The movement he had been leading gained momentum from his death. A: 2.7 P: 4.0

29. a) The babysitter punished by her mother had been stealing clothes from the family for whom she babysat. The mother was extremely embarrassed. A: 4.4 P: 2.1

b) The child punished by her mother had been stealing candybars from the corner drugstore. The mother was extremely embarrassed. A: 1.5 P: 5.8

30. a) The artist studied by the woman was known throughout the world. He had been a leader of the pop art movement for three decades. A: 5.7 P: 3.1

b) The newborn studied by the woman was known to have Down's syndrome. Like most Down's children, he was cute, lovable and good-natured. A: 2.8 P: 4.0

31. a) The priest worshipped by his followers was ignorant of their strife. He often turned away the homeless when they sought shelter in his church. A: 6.7 P: 4.1

b) The goddess worshipped by her followers was ignorant of their strife. She often wished they would just go away and leave her alone. A: 1.5 P: 6.7

32. a) The professor taught by the specialists was better skilled than the others. Her exceptional training had allowed her to get a job at Stanford. A: 6.6 P: 2.6

b) The trainee taught by the specialists was better skilled than the others. He moved to a management position

much more quickly than normal. A: 1.6 P: 6.3

33. a) The policeman lectured by the committee was told that he would be suspended. He had been secretly videotaped while beating an Asian man. A: 5.9 P: 2.3
    b) The freshman lectured by the committee was told that he would be suspended. He had been caught vomitting in the dormitory bathrooms on three different occasions. A: 1.9 P: 5.4
34. a) The hunter captured by the troopers had been hunting on someone else's land. To catch him in the act, they had set up a trap two days earlier. A: 6.0 P: 2.0
    b) The fugitive captured by the troopers had been on the run for eight weeks. He had travelled from New York to Arizona before being recognized. A: 2.6 P: 5.1
35. a) The bellboy lifted by the fireman was unable to breathe. He put him over his shoulder and ran out of the burning building. A: 6.5 P: 2.6
    b) The infant lifted by the fireman was unable to breathe. He put him against his shoulder and ran out of the burning building. A: 2.5 P: 5.9
36. a) The lion chased by the zookeeper had escaped from its cage. The zoo had been evacuated and all gates were secured. A: 6.6 P: 2.6
    b) The mouse chased by the zookeeper had escaped from its cage. It was a rare Arabian mouse with orange and grey fur. A: 3.1 P: 5.5
37. a) The principal dismissed by the superintendent was happy to be leaving his job. He was tired of dealing with self-indulgent teenagers. A: 6.3 P: 2.0
    b) The student dismissed by the superintendent was happy to be leaving the room. He was tired of being accused of being the school's major drug source, even thought he in fact was. A: 1.5 P: 5.9
38. a) The auditor investigated by the government was accused of bribing public officials. He denied all charges. A: 6.2 P: 2.2
    b) The gangster investigated by the government was accused of bribing public officials. He denied all charges. A: 3.5 P: 6.4
39. a) The patrolman searched by the guard was tired of this nightly routine. He felt that after working there for 20 years he should be trusted. A: 6.3 P: 1.4
    b) The prisoner searched by the guard was tired of this nightly routine. He felt that after being there for 20 years he should be trusted. A: 1.7 P: 6.2
40. a) The scientist examined by the coroner had been strangled. There was no physical evidence for the cops to go on. A: 6.7 P: 2.3
    b) The defendant examined by the coroner had been strangled. There was no physical evidence for the cops to go on. A: 2.7 P: 5.8

REFERENCES

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In MacWhinney & Bates (Eds.), *The crosslinguistic study of sentence processing.* New York: Cambridge U. Press.

Bever, T. G. (1970). The cognitive basis for linguistic structure. In J. R. Hayes (Ed.), *Cognitive development of language.* New York: Wiley.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language,* **34,** 774–806.

Burgess, C. (1991). *Interaction of semantic, syntactic, and visual factors in syntactic ambiguity resolution.* Unpublished Ph.D. dissertation, University of Rochester, Rochester, NY.

Burgess, C., & Lund, K. (1997). Modeling parsing constraints with high-dimensional context space. *Language and Cognitive Processes: Special Issue on Lexical Representations in Sentence Processing,* **12,** 177–211.

Burgess, C., Tanenhaus, M., & Hoffman, M. (1994). Parafoveal and semantic effects on syntactic ambiguity resolution. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (pp. 96–99). Hillsdale, NJ: Erlbaum.

Caplan, D., Hildebrandt, N., & Waters, G. S. (1994). Interaction of verb selectional restrictions, noun anim-

acy, and syntactic form in sentence processing. *Language and Cognitive Processes,* **9,** 549–585.

Carlson, G. N., & Tanenhaus, M. K. (1988). Thematic roles and language comprehension. In W. Wilkins (Ed.), *Thematic relations* (pp. 263–288). New York: Academic Press.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Clifton, C. (1993). Thematic roles in sentence parsing. *Canadian Journal of Experimental Psychology,* **47,** 222–246.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language,* **25,** 348–368.

Francis, W., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.

Frazier, L. (1987). Theories of syntactic processing. In J. Garfield (Ed.), *Modularity in knowledge representation and natural language processing* (pp. 559–586). Cambridge, MA: MIT Press.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition,* **6,** 291–325.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, **37,** 58–93.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, **28,** 267–283.

Hanna, J. E., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1996). *Integrating contextual and sentential constraints in ambiguity resolution.* Manuscript in preparation.

Juliano, C., & Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject-object attachment preference. *Journal of Psycholinguistic Research,* **23,** 459–471.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science,* **20,** 137–194.

Just, M., Carpenter, P., & Woolley, J. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General,* **111,** 228–238.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994b). Syntactic ambiguity resolution as lexical ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 123–154). Hillsdale, NJ: Erlbaum.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994a). The lexical nature of syntactic ambiguity resolution. *Psychological Review,* **101,** 676–703.

McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed approach. *Language and Cognitive Processes,* **4,** 287–335.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes: Special Issue on Lexical Representations in Sentence Processing,* **12,** 137–176.

Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). The linguistic tuning hypothesis: Further corpus and experimental evidence. *Journal of Psycholinguistic Research,* **24,** 469–488.

Pearlmutter, N. J., Daugherty, K., MacDonald, M. C., & Seidenberg, M. S. (1994). Modeling the use of frequency and contextual biases in sentence processing. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (pp. 699–704). Hillsdale, NJ: Erlbaum.

Pearlmutter, N., & MacDonald, M. C. (1992). Plausibility and syntactic ambiguity resolution. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 498–503). Hillsdale, NJ: Erlbaum.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing. *Journal of Verbal Learning and Verbal Behavior,* **22,** 358–374.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ: Prentice Hall.

Schlesinger, I. M. (1995). *Cognitive space and linguistic case.* Cambridge: Cambridge University Press.

Spivey-Knowlton, M. J. (1996). *Integration of visual and linguistic information: Human data and model simulations.* Unpublished doctoral dissertation, University of Rochester, Rochester, N.Y.

Spivey-Knowlton, M. J., & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition,* **55,** 227–267.

Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1997). *Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency within an integration-competition framework.* Manuscript submitted for publication.

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology,* **37,** 276–309.

Stowe, L. A. (1989). Thematic structures and sentence comprehension. In G. N. Carlson & M. K. Tanenhaus (Eds.). *Linguistic structure in language processing.* (pp. 319–357). Boston: Kluwer Academic Publishers.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research,* **23,** 295–322.

Swinney, D. A. (1979). The resolution of indeterminacy during language comprehension: Perspectives on modularity in lexical, structural, and pragmatic pro-

cessing. In G. B. Simpson (Ed.), *Understanding word and sentence.* Amsterdam: North Holland.

Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes: Special Issue on Lexical Representations in Sentence Processing,* **12,** 211–272.

Tabossi, P., Spivey-Knowlton, M. J., McRae, K., & Tanenhaus, M. K. (1994). Semantic effects on syntactic ambiguity resolution: Evidence for a constraint-based resolution process. In C. Umilta & M. Moscovitch (Eds.), *Attention & performance XV* (pp. 589–616). Cambridge, MA: MIT Press.

Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language,* **27,** 597–632.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of cognition and perception.* San Diego, CA: Academic Press.

Tanenhaus, M. K., Spivey-Knowlton, M. J., & Hanna, J. E. (in press). Modeling the effects of discourse and thematic fit in syntactic ambiguity resolution. In M. Crocker, M. Pickering, & C. E. Clifton (Eds.), *Architectures and mechanisms for language processing.* Cambridge: Cambridge University Press.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework of syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155–180). Hillsdale, NJ: Erlbaum.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language,* **35,** 566–585.

Trueswell, J. C., & Tanenhaus, M. K. (1991). Tense, temporal context and syntactic ambiguity resolution. *Language and Cognitive Processes,* **6,** 303–338.

Trueswell, J. C., & Tanenhaus, M. K. (1992). Consulting temporal context in sentence comprehension: Evidence from the monitoring of eye movements in reading. *Proceedings of the 14th Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language,* **33,** 285–318.

Vaughan, J. (1983). Control of fixation duration in visual search and memory search: Another look. *Journal of Experimental Psychology: Human Perception and Performance,* **8,** 709–723.

(Received April 29, 1997)
(Revision received September 24, 1997)