Available online at www.sciencedirect.com
SCIENCE DIRECT®

Cognitive Psychology 47 (2003) 164–203

Cognitive Psychology

www.elsevier.com/locate/cogpsych

The misinterpretation of noncanonical sentences

Fernanda Ferreira

Department of Psychology and Cognitive Science Program, Michigan State University, Psychology Research Building, East Lansing, MI 48824-1117, USA

Accepted 31 October 2002

Abstract

Research on language comprehension has focused on the resolution of syntactic ambiguities, and most studies have employed garden-path sentences to determine the system's preferences and to assess its use of nonsyntactic sources information. A topic that has been neglected is how syntactically challenging but essentially unambiguous sentences are processed, including passives and object-clefts—sentences that require thematic roles to be assigned in an atypical order. The three experiments described here tested the idea that sentences are processed both algorithmically and heuristically. Sentences were presented aurally and the participants' task was to identify the thematic roles in the sentence (e.g., Who was the do-er?). The first experiment demonstrates that passives are frequently and systematically misinterpreted, especially when they express implausible ideas. The second shows that the surface frequency of a syntactic form does not determine ease of processing, as active sentences and subject-clefts were comprehended equally easily despite the rareness of the latter type. The third experiment compares the processing of subject- and object-clefts, and the results show that they are similar to actives and passives, respectively, again despite the infrequent occurrence in English of any type of cleft. The results of the three experiments suggest that a comprehensive theory of language comprehension must assume that simple processing heuristics are used during processing in addition to (and perhaps sometimes instead of) syntactic algorithms. Moreover, the experiments support the idea that language processing is often based on shallow processing, yielding a merely "good enough" rather than a detailed linguistic representation of an utterance's meaning.

E-mail address: fernanda@eyelab.msu.edu.

doi:10.1016/S0010-0285(03)00005-7

1. Introduction

Much of the research that has been conducted over the last 20 years or so on sentence comprehension has addressed a rather specific question: How is syntactic ambiguity resolved (e.g., Ferreira & Clifton, 1986; Ferreira & Henderson, 1990, 1991; Frazier & Rayner, 1982; Hoeks, Vonk, & Schriefers, 2002; MacDonald, 1994; Schneider & Philips, 2001; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Traxler, Pickering, & Clifton, 1998)? The question is significant because often a string of words can be given more than one syntactic analysis. For example, in *Mary put the book on the table onto the shelf*, the string *on the table* either could refer to the book's destination (the incorrect analysis) or it might modify the word *book* (the ultimately correct interpretation) (Boland & Boehm-Jernigan, 1998; Britt, 1994; Ferreira & Clifton, 1986; Speer & Clifton, 1998; Tanenhaus et al., 1995). The assumption is that the syntactic ambiguity must be resolved properly so that a correct interpretation for the sentence can be computed. According to what are known as "syntax-first" or "two-stage" models of sentence comprehension (or parsing, a term that focuses particular attention on the mechanism that assigns syntactic structure to sentences), the ambiguity is resolved first on a purely syntactic basis, without the help of information from nonsyntactic sources such as immediate discourse and visual context, real-world knowledge, or even lexical and prosodic constraints (Ferreira & Clifton, 1986; Ferreira & Henderson, 1990; Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983; Steinhauer, Alter, & Friederici, 1999). In this view, a sentence is hard to comprehend when its syntactic form is incompatible with the parser's initial biases for creating syntactic structure. According to what are known as "constraint-based" or "interactive" models, the different syntactic alternatives are activated in proportion to the evidence for them. Sentences are difficult to understand when a syntactic analysis for which there initially appears to be a great deal of evidence turns out to be incorrect—as in the example given above (MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus et al., 1995; Trueswell, Tanenhaus, & Kello, 1993).

The back-and-forth between the syntax-first and the constraint-based camps has gradually pruned down the set of questions being investigated in the field to a very small subset of all the critical issues one might explore concerning language processing. Presumably, all researchers are interested in how a person arrives at an interpretation of a sentence or some other significant chunk of linguistic material. The focus on syntactic analysis has arisen because it is assumed that meaning-based representations are built on syntactic frames (e.g., Frazier & Clifton, 1996; MacDonald et al., 1994). Thus, in the example above, the initial, incorrect syntactic analysis on which *on the table* is sister to the verb *put* supports the (wrong) interpretation that *on the table* is the destination for the book. For the correct interpretation to be obtained, a different syntactic structure must be computed. The focus on parsing has also led to a heavy reliance on what are termed "online" measures as the gold standard for psycholinguistic research. The disputes between the syntax-first and constraint-based approaches concern the time-course of activation of different sources of information, and those are most obviously resolved by examining moment-by-moment changes in processing load as a sentence unfolds over time.

The combination of these two influences—the emphasis on parsing and the use of online measures—has led to a situation in which few studies of "sentence comprehension" actually include serious measures of people's interpretations of sentences. Although researchers assume that a complex sentence is assigned the semantic interpretation supported by the syntactic frame (Frazier & Clifton, 1996; MacDonald et al., 1994), little direct evidence for that assumption has actually been collected (but see Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, Bailey, & Ferraro, 2002; and Sanford, 2002 for evidence that interpretations can be inconsistent with syntactic form). True, most experiments require participants to read sentences so that they can later answer "comprehension questions" about them. However, typically the questions tap into superficial features of the sentences and often data on question-answering accuracy are not collected. As a result, we lack information about the sorts of meanings people actually derive for the sentences they are shown in psycholinguistic experiments, and we know little about how those interpretations were created. To begin to remedy this situation, the research reported here attempts to measure the actual content of people's interpretations rather than just the time that was required to construct them. The task used in the three experiments requires participants to identify the agent or patient/theme of an action. The result is a measure of some aspects of the semantic representation constructed for a sentence, as well as its stability.

It is important to focus on the content of the interpretations because there is evidence that under some circumstances comprehenders do **not** obtain the meaning consistent with a sentence's true content. For example, Fillenbaum (1971, 1974) conducted a number of studies demonstrating that people tend to recall sentences in a "normalized" form—they changed the sentences' meaning to make them sensible and conventional rather than strange or anomalous. For example, most participants paraphrased *Don't print that or I won't sue you* to mean that if some item were printed, the result would be a lawsuit. The classic Moses illusion (Erickson & Mattson, 1981; Kamas, Reder, & Ayers, 1996) reveals the same phenomenon: If someone is asked *How many animals of each kind did Moses take on the ark?*, the person tends to overlook the problem with the question and answers "two" (see also Barton & Sanford, 1993; Sanford, 1999, 2002). Duffy, Henderson, and Morris (1989) made a similar point with a very different methodology. They presented sentence contexts using rapid serial visual presentation (RSVP), and the task of the participants was to name the sentence-final word. Duffy et al. observed facilitation relative to an appropriate baseline when participants were presented with "moustache" following a context such as *The barber who watched the woman trimmed the...* The relevant finding for the present purposes was that the same amount of facilitation was observed for *The woman who watched the barber trimmed the...*, in which the semantic relations among the content words as specified by the form of the sentence do not lead to the meaning that presumably underlies the facilitation (see also Morris, 1994). In all of these cases, it appears that the comprehender's analyses of the sentences were too superficial to support the correct interpretations (Sanford, 1999, 2002).

In a study of people's interpretations of elliptical verb phrases, Garnham and Oakhill (1987) asked participants to read sequences such as *The elderly patient had* *been examined by the doctor. The child/nurse had too*, and the participants' task was to answer whether the doctor examined the child or the nurse. Garnham and Oakhill also varied whether an adjunct phrase occurred following the by-phrase of the first sentence (e.g., *The elderly patient had been examined by the doctor during the ward round*) in order to test whether question-answering accuracy was affected by the delay between receipt of the critical arguments of the full passive and the question that probed their representation. Garnham and Oakhill found that when the content of the elliptical verb phrase (VP) was plausible (that is, when the doctor examined a child instead of examining a nurse), the error rate was 8% for the regular condition and 11% for the delayed condition. In contrast, when the content of the elliptical VP was implausible, the error rate for the same two conditions was 25 and 39%. A similar story emerged from the reaction time data: People were faster in the plausible conditions, faster with no intervening adjunct phrase, and the cost of delay was much greater for the implausible texts. Thus, this study again indicates that people's interpretations of sentences are sometimes incompatible with their actual content. Our schematic knowledge states that it is more plausible that a doctor would examine a child than that he or she would examine a nurse, and that semantic information seems strong enough to override the need for further structural analysis.

A sentence's syntactic complexity seems to influence the extent to which misinterpretations occur. The example sentences that have been mentioned as leading to illusions of comprehension are typically not simple, active, declarative clauses. For example, the item *Don't print that or I won't sue you* (Fillenbaum, 1971, 1974) consists of two clauses, both of which include negation. Work by Wason (1959, 1972) demonstrated that sentences containing negation (e.g., 57 *is not an even number*) are harder to comprehend than affirmative sentences (see also Just & Carpenter, 1976; Just & Clark, 1973). The sentences that induce the Moses illusion involve a fairly complex type of wh-question that places the critical item (e.g., Moses) in a position usually reserved for presupposed information. Hornby's (1974) study of picture–sentence verification showed that comprehenders are more likely to overlook presupposed information than focused information in a sentence (see also Cutler & Fodor, 1979, for a study making a similar point). Some of the Duffy et al. materials involved embedding of a relative clause inside the main clause, and subject relatives are known to be more difficult than object relatives (King & Just, 1991; Sheldon, 1974). Finally, the Garnham and Oakhill sentences are stated in the passive voice, so not only does the comprehender have to assign thematic roles in an atypical order, he must do so in a full clause and then use those assignments to fill out elided material. This observation regarding complexity is critical, because whereas there is evidence that sometimes misinterpretations occur with challenging structures, there is also work showing that people are generally willing to accept wildly implausible meanings of simple sentences. For example, MacWhinney, Bates, and Kliegl (1984) asked participants to identify the agent in sentences such as *the eraser bites the turtle*. MacWhinney et al. found that English speakers relentlessly chose the first noun phrase as the agent (the sentences' main verbs were all of the agent–patient variety) no matter how strange the result (and a similar finding will be reported here). It appears, then, that syntactic complexity increases the chances that a sentence will be misinterpreted.

Clearly, then, much remains to be learned about people's comprehension of basic, unambiguous sentences. The current study examined how people assign semantic roles (Fillmore, 1968) to the various concepts in a sentence, and how they maintain those assignments. Consider the sentence *the dog was bitten by the man*. The syntax of the sentence together with the argument structure (Grimshaw, 1990; MacDonald et al., 1994) of the verb *bite* specifies that the man is the agent of the action and the dog is the theme or patient. On the other hand, schematic knowledge is consistent with the opposite thematic role assignments, and at least intuition suggests that the sentence is difficult to understand. It is as if the correct assignment of roles is repeatedly challenged by the interpretation more consistent with world knowledge.

As described earlier, standard models of ambiguity resolution have little to say about how such examples are interpreted, because these sentences do not create classic garden-paths. Moreover, the standard models assume that interpretations are built on syntactic frames, and therefore they have trouble accounting for some of the misinterpretations summarized earlier. It appears, then, that existing models of comprehension are inadequate. What they are missing is an architectural component that can account for cases in which people engage in shallow, incomplete, and even inaccurate processing of utterances. In other domains of cognition such as reasoning and decision making, it has been argued that it would be in the interests of the organism to engage in heuristic processing, at least in some situations (Gigerenzer, 2000; Gigerenzer, Todd, & ABC Research Group , 1999). Gigerenzer and his colleagues point out that all organisms make decisions with limited time and resources. Models of rational choice which assume "unbounded rationality" are unrealistic because the computations that are assumed to take place are often far too burdensome for real creatures operating in demanding environments. Moreover, as Gigerenzer et al. suggest, organisms must have some sort of *stopping rule*; that is, once they have come to the point where they subjectively sense that they can make a decision, they should terminate the search for additional information, particularly if it is costly to collect. Gigerenzer and his colleagues propose that people use "fast and frugal heuristics." Fast and frugal heuristics tell an organism how to search efficiently for information, terminate explorations, and then make a decision quickly. These heuristics are in contrast to algorithmic computations (e.g., Bayesian calculations) that might be guaranteed to yield a correct solution but which are exceedingly costly to undertake. Comparable (but less radical) proposals include "optimization under constraints" (Anderson & Milson, 1989) and satisficing (Simon, 1956).

Similarly, in computer science as well as some areas of psychology and zoology, an approach known as Sequential Decision Making has been gaining increasing attention, and it too is compatible with the notion of heuristic processing in certain cognitive domains (Henderson, Falk, Minut, Dyer, & Mahadevan, 2001; Houston & McNamara, 1999; Kaelbling, Littman, & Cassandra, 1998; Mahadevan & Cornell, 1992). These models assume that decision making is a sequential process, involving a series of episodes in which an agent (artificial or biological) selects the action appropriate for its goals. The problem is made difficult by perceptual uncertainty, time and resource limitations, and constraints on the extent to which the history of previous decisions is stored and maintained. For example, an organism that must try to understand a spoken sentence is engaged in a complex sequential decision making task, because the words are received one at a time, and decisions about how to interpret them must be made both in the face of uncertainty about the current input (because of environmental noise, errors in the signal), lack of knowledge about upcoming material, and only partial information about its past decisions (Tyler & Warren, 1987). A basic assumption of the Sequential Decision Making approach (which is based on the theory of Markov Decision Processes; Puterman, 1994) is that agents therefore store little information about past states and create quite shallow representations of the domains relevant to their decisions. This framework has been applied to a wide variety of problems including gaze control in human vision (Henderson et al., 2001; Minut & Mahadevan, 2001), robot navigation (Mahadevan, Theocharous, & Khaleeli, 1998), and language processing (Crocker & Brants, 2000).

As has been argued thus far, language comprehension in many cases is shallow and incomplete. To account for this fact, it would be useful for the field of language comprehension to adopt an approach similar to that taken in Sequential Decision Making and the Fast and Frugal Heuristics models. The central idea is that people use simple heuristics to process sentences in addition to the syntactic algorithms that have been carefully evaluated in recent studies of processing. Based on just the results of the experiments that will be reported here, it is not yet possible to specify how the product of the heuristics is coordinated with the output of the syntactic algorithms. Eventually it will be important to know whether the two operate in parallel, whether the heuristics are used only when the algorithms become hard to apply, or whether algorithms are used only when the comprehender lacks confidence in the interpretation created from the heuristics. All three possibilities are plausible, but unfortunately at present there is no work to allow us to to distinguish among them with any confidence. Nevertheless, the present study makes a significant contribution: The results of the experiments provide compelling evidence that some type of heuristic processing clearly takes place. The implication is that no theory of language comprehension can be considered complete or even adequate without considering the role heuristics might play.

One model that has explicitly considered the role of heuristics is the Late Assignment of Syntax Theory or LAST (Townsend & Bever, 2001). In LAST, people are assumed to essentially process sentences twice. One analysis yields a "quick and dirty" interpretation based mainly on semantic associations and syntactic habits. The critical syntactic habit for our purposes is the comprehender's strong tendency to assume that the subject of a sentence is also the agent of some action and the object of the same sentence is the patient or theme (the so-called NVN strategy). This habit arises because an overwhelming number of sentences in English conform to this particular pattern (Bever, 1970). Townsend and Bever refer to this analysis as a "pseudo-parse," because the language comprehension system does not make use of its rich store of syntactic knowledge to assign a syntactic structure systematically and from left to right. Instead, a handful of simple heuristics yield a "gut" level representation of the sentence's meaning. A second more time-consuming algorithmic analysis also takes place. If the algorithmic parser has time to finish all its computations, then it will output the correct interpretation of the sentence. If it does not, and if the pseudo-parse has produced a wrong analysis, the sentence will be misunderstood. And even if the true parse finishes but yields a structure inconsistent with the pseudo-parse, the comprehension system will have to reconcile the two, and this is one potential source of error in understanding.

Heuristics are a central component of analysis-by-synthesis models, of which LAST is a particularly good example. In analysis by synthesis approaches, heuristics are used to create a "best guess" about the input, and then syntactic procedures are used to try to generate that input (Fodor, Bever, & Garrett, 1974; Garrett, 2000; Townsend & Bever, 2001). The pseudo-parser is the component that uses heuristics (like the NVN strategy as well as semantic associations) to create a preliminary hypothesis. The true parser then uses that preliminary hypothesis to constrain its search space, so that when it calls up its syntactic procedures it need only consider a small subset of the formal possibilities. Garrett (2000) has proposed a model of sentence comprehension which takes an even more radical analysis-by-synthesis approach (compared with LAST). Garrett argues that sentence comprehension uses the language production system. As he points out, it is already uncontroversial in the area of language production to assume that the opposite is true: Researchers in language production have postulated the existence of an editor which checks output before it is articulated to make sure it conforms to the rules of language (Dell & Reich, 1981; Levelt, 1989). The editor is not a special purpose device; instead, the speaker simply parses his or her own output before sending it out to the articulators. Conversely, Garrett proposes that the sentence comprehension system uses the production system to allow meaning and other types of conceptual relations to constrain parsing. On this view, when a comprehender analyses a sentence, he or she uses the language production system to generate a structure and to evaluate her hypothesis about the structure of the sentence. An intriguing implication of Garrett's proposal is that one might expect that the same mistakes that can arise in production can derail comprehension as well. For instance, the production system sometimes creates word exchange errors, as in *Murder is a form of suicide* (Garrett, personal communication). This type of mistake arises at the point when the production system uses meaning to begin to establish grammatical relations (Bock & Levelt, 1994; Garrett, 1975). It is an interesting property of this mechanism that it seems insensitive to the anomaly that results from the word exchange; that is, the production system seems unaware that the utterance does not match the speaker's semantic intention. If the comprehension system does indeed use language production mechanisms to perform analysis by synthesis, it is possible that sentences such as *the dog was bitten by the man* are misinterpreted because the production system has trouble keeping track of the bindings between grammatical positions and thematic roles. It is certainly intriguing that for both the real speech error example *Murder is a form of suicide* and the comprehension example *the dog was bitten by the man* it is difficult to detect that the result is semantically anomalous.

The idea that heuristics are used during comprehension allows us to account for cases in which people sometimes process sentences shallowly (Ferreira et al., 2002; Oakhill & Garnham, 1996; Oakhill, Garnham, & Vonk, 1989; Sanford, 2002). It is only the algorithmic parse that is guaranteed to yield a correct analysis, but it takes longer to apply the algorithms than to use a few simple heuristics (for details, see Simons & Levin, 1997). Therefore, there may be situations in which people do not obtain the algorithmically based analysis at all. Perhaps the person was under time pressure, as often happens in normal conversations; or it might be that the comprehender's processing system terminated its operations because the pseudo-parse reached some criterion level of confidence (as argued in Christianson et al., 2001, for reanalysis processes). In these cases, if the associations that the heuristics are based on are inconsistent with the algorithmically based analysis, then a sentence could very well be misinterpreted.

It is clear that a useful heuristic for processing in English would be one which assumes that sequences consisting of a subject followed by a verb and then an object should be interpreted as agent-verb-patient. It is important to realize that this description glosses over some important details. It is not clear that the specific thematic role that get assigned is critical. A sentence such as *Mary likes Tom* is just as easy to understand as *Mary kicked Tom* even though in the *likes* example the thematic roles involved are not agent and patient (with verbs of mental state such as *like*, the role assigned to the subject of an active is experiencer and the role assigned to the object is theme). A way around this problem is to invoke the notion of a proto-agent and a proto-patient (Dowty, 1991). The idea is that in any sentence there is one argument that is most agent-like and another that is most patient-like. Dowty explicitly defines the features typically correlated with the two roles; for example, agents are often animate and are entities capable of instigating events. The most accurate way to describe the NVN strategy, then, is to say that an entity that has many of the features of an agent should be bound to the subject position, and one that has many of the characteristics of a patient should be bound to the object position of any given sentence.

To begin to investigate the question of how heuristics might be used during language processing, three experiments were conducted investigating the processing of easy and difficult unambiguous sentences.[1] In the first experiment, active and passive sentences of various types were presented aurally to participants, and the participants' task was to identify explicitly the concept corresponding to some thematic role. The second experiment compared two infrequent structures—subject-clefts (e.g., *It was the man who bit the dog*) and passives—in order to assess whether it is the frequency of the form itself that is important, or if instead what matters is whether thematic roles are assigned in the most frequent way (agent to subject and patient to object—that is, according to the NVN strategy). The third experiment examined subject-clefts and object-clefts (e.g., *It was the dog that the man bit*) in order to obtain further information about what kind of frequency information is critical to the sentence comprehension system. In the General Discussion, I consider the implications of these data for models of comprehension.

[1] Clearly, the sentences are not entirely without ambiguity, because, for instance, any sentence-initial NP can be given many syntactic analyses. The relevant sense of "unambiguous" here is that the sentences do not produce any traditional type of garden-path.

2. Experiment 1

Participants listened to active and passive sentences and at the end of each named out loud either the agent or the patient. Decision accuracy and reaction times were recorded. This type of task has been used in other studies of language comprehension, particularly those that have been conducted within the framework of the Competition Model (Bates, Devescovi, & D'Amico, 1999; Bates, McNew, MacWhinney, Devescovi, & Smith, 1982; MacWhinney & Bates, 1989; MacWhinney et al., 1984). The measure has yielded a large body of important work concerning how different languages are comprehended, which lends some validity to the use of the thematic role decision task in the present experiment. (See Bates et al., 1999, for a discussion of various criticisms of this task.) One difference between the agent–decision task used by proponents of the Competition Model and the task used here is that here both thematic roles will be tested. Nevertheless, because only one role is probed on any given trial, the logic of both paradigms is similar.

2.1. Method

Participants. A total of sixty-three undergraduates attending Michigan State University participated in the experiment in exchange for partial credit in their Introductory Psychology courses. Seven of the participants were excluded from the data analyses because they were non- native speakers of English,[2] five because they could not understand the task (as indicated by their performance on the practice trials), and three because they did not speak loudly enough to reliably set off the voice-activated relay switch. Thus, the data from 48 participants were included in the data analyses.

Materials. There were 72 experimental items, each describing simple transitive events. Each could appear in one of four versions (the complete set of items is given in Appendix A). The item was either active or passive, and the two arguments were arranged in one order or the other. Twenty-four of the items were reversible but highly biased: The two arguments could be swapped but one arrangement was much more plausible than the other (e.g., *the dog bit the man/the man bit the dog*). Another twenty-four items were nonreversible: One arrangement of the objects produced not just implausible but semantically anomalous meanings (e.g., *the mouse ate the cheese/the cheese ate the mouse*). Items were nonreversible because one argument was animate and the other inanimate. A final set of 24 items was symmetrical: The arguments could be swapped, and the two arrangements were equally plausible (e.g., *the woman visited the man/the man visited the woman*).

The semantic properties of these (and many other) sentences were assessed in a separate normative study involving 100 participants. All were native speakers of

[2] The practice in our laboratory is to allow those who are non-native speakers and are taking courses that require them to obtain experimental credits to sign up for the experiments we conduct. We then eliminate their data rather than restricting their participation.

English, and none participated in this experiment or the other two. The purpose of this norming study was to assess the properties of a large number of stimuli that were candidates for various experiments in the laboratory, including the ones for the present experiments. To reduce the time-burden on participants, the stimuli were distributed over 10 lists, and any one participant responded to just one of those lists. The first five lists included all 72 of the experimental items and all the fillers. Half of the 24 biased, reversible items and half of the 24 nonreversible items were presented in the plausible versions and the other half were presented in the implausible versions. The 24 symmetrical items were included as well, half in one arrangement and half in the other. The other five lists were simply the complement of the first five (i.e., if an item in lists 1–5 occurred in its plausible version, it was given in its implausible version for the lists 9–10). All experimental items were presented in their active forms. Any one list required a participant to respond to 120 sentences. Participants were told to read each item slowly and carefully, and they were warned that some of the sentences were implausible. They were told to rate each sentence on a scale from 1 to 7, where "1" meant that the sentence was so implausible as to be anomalous and "7" meant that the sentence described an extremely likely event (examples were provided). The means and standard deviations for the six critical conditions are given in Table 1.

As can be seen, the stimuli had the appropriate semantic properties for the experiments. The implausible sentences were rated as far less plausible than the plausible sentences, and the two symmetrical versions were rated almost exactly the same.

The 72 experimental items were combined with a total of 144 filler items (also evaluated in the normative study). The fillers were sentences of a variety of different syntactic types. One half of the fillers were constructed so that they could be probed regarding either the location in which some event took place or the color of some critical object (e.g., *Dave fed Alpo to the brown dog at the park*). The other 72 were written so either a temporal period or the action could be probed (e.g., bicycles were banned by the authorities in May).

The 72 experimental items in their four different versions (active/passive by plausible/implausible), the 144 filler sentences, and the practice sentences were recorded by a female native speaker of American Midwestern English and the sentences were digitized at a rate of 10 kHz. Each sentence was stored as a speech file for presentation over headphones. An individual participant heard a given item only once and

Table 1
Mean plausibility ratings and standard deviations for experimental sentences used in Experiments 1 and 2

Sentence type	Mean rating from 1 (implausible) to 7 (plausible)/standard deviations
Biased reversible, plausible (*the dog bit the man*)	5.82 / .93
Biased reversible, implausible (*the man bit the dog*)	2.08 / 1.00
Nonreversible, plausible (*the mouse ate the cheese*)	6.85 / .40
Nonreversible, implausible (*the cheese ate the mouse*)	1.86 / 1.01
Symmetrical, one order (*the woman visited the man*)	2.95 / 1.03
Symmetrical, other order (*the man visited the woman*)	2.98 / 1.09

thus heard just one version of any experimental sentence. Across experimental items, a participant heard an equal number of actives and passives, and an equal number of plausible and implausible sentences. In total, a participant responded to 216 sentences: 24 that were biased but reversible, 24 nonreversibles, 24 symmetricals, and 144 fillers. For half the 72 experimental items, the participant was required to identify the agent; for the remaining 36 experimental items, the participant identified the patient. For the 144 fillers, the participant made 36 decisions regarding a color, 36 regarding a temporal interval, 36 concerning a location, and 36 about the action. The 216 items were presented in a random order, generated separately for each participant in the experiment.

Procedure. The experimental session began with instructions. Participants were told that they would listen to sentences, and after each, make a decision. They were informed that the decisions were of six different types, and the types were described and illustrated with a sample sentence. These six types were: (1) DO-ER (corresponding to agent): Participants were told that in *the man fixed the dress*, the man would be the do-er, because he is the one who did the action. (2) ACTED-ON (corresponding to patient or theme): In the same sentence the dress is the thing acted-on, because it is the dress that is acted on by the do-er. (3) ACTION: For the same sentence the correct response would be "fix" or "fixing" (both variants were scored as correct), because fixing is the action described in the sentence. (4) LOCATION: Given *the bank was across the street*, the correct response would be "across the street," because that is where one of the things mentioned in the sentence is located. (5) COLOR: For *I saw a bright yellow Volkswagon yesterday* the correct response would be "yellow," because that is the color of an object provided in the sentence. Finally, (6) WHEN: For the same sentence, the correct response would be "yesterday," because the sentence states that that is when the event took place.

The participants were then read 12 example sentences from the instructions sheet, and after each they had to make two decisions for each of the six types. The experimenter corrected any errors and asked the participants whether they understood the rationale for each response. If participants provided a satisfactory response, the experimenter set up the practice session. The practice session was identical to the experimental portion, and included twelve unique sentences presented aurally (two for each of the six decisions). Four of these sentences were passives and the rest were actives. If a participant made more than two errors, that person was excused from the rest of the experiment (five people were dismissed based on this criterion). If the participant made fewer than two errors, the experimental session was initiated.

The participant sat in front of a computer and wore headphones throughout the experiment. A trial began with a message to the participant to begin the trial when ready, and the message was displayed on a video monitor. After a button on a response panel was pushed, a sentence was played out over headphones. Immediately after the sentence ended, one of the six prompts appeared on the screen in Arial 24-point font, and the participant's task was to provide the appropriate response out loud. The onset of vocalization triggered a voice-activated relay, and thus decision times were automatically stored in a computer file. The participant's response was written down by the experimenter so that the trial could be scored off-line as either correct or incorrect. Participants were tested individually, and each experimental session lasted between 45 min and 1 h (depending on how much time participants chose to take between trials).

Design. The experiment employed a $2 \times 2 \times 2$ within-participants design: the syntactic form of the sentence was either active or passive, the sentence's meaning varied (plausible vs. implausible for the biased reversible items, plausible vs. anomalous for the nonreversible items, and order 1 vs. order 2 for the symmetrical items), and the participant either made a decision about the agent or patient/theme. Each of the three types of sentences—reversible but biased, nonreversible, and symmetrical—was analyzed separately. The primary dependent measure was accuracy, but decision times were analyzed as well.

2.2. Results and discussion

Reaction times shorter than 300 ms and longer than 7500 ms were eliminated. These criteria resulted in the removal of less than 3% of the data. Means were computed for all eight conditions for both participants and items, and analyses of variance were performed with both participants ($F1$) and items ($F2$) as random effects. All effects are significant at $p < .05$ unless otherwise indicated.

Reversible, biased sentences. The percentage of decisions that were correct in each of the eight conditions is shown in Table 2. For agent decisions, accuracy was the same for plausible and implausible active sentences (99% for both). Thus, people had no trouble stating that the man was the agent in a sentence such as *the man bit the dog*. With passives, participants were less accurate overall (81%), and less accurate for sentences that were implausible (74% vs. 88%). For patient decisions, participants were less accurate in response to passives, but for the two sentence types accuracy was equivalently higher in the two plausibility conditions.

The overall $2 \times 2 \times 2$ ANOVA revealed a significant three-way interaction, $F1(1,47) = 5.23$, $SEM = 2.1\%$, $F2(1,23) = 5.43$, $SEM = 2.7\%$. Collapsing over the agent vs. patient decision, there was also an interaction between form and meaning, $F1(1,47) = 5.63$, $SEM = 2.4\%$, $F2(1,23) = 17.89$, $SEM = 2.1\%$. For actives, participants were about as accurate with plausible sentences (98%) as with implausible sentences (95%). For passives, participants were more accurate with plausible than implausible sentences (90% vs. 80%). There was also a significant two-way interaction between form and decision type, $F1(1,47) = 14.19$, $SEM = 2.4\%$, $F2(1,23) = 18.65$, $SEM = 2.0\%$. For actives, participants were better at identifying the agent than the patient (99 and 94% respectively); for passives, the opposite was true (81 and 89%). This pattern is likely due to a tendency for people to be more accurate making decisions about the first argument of the sentence. There was no interaction between meaning and decision type, both F's < 1. There was a significant main effect of form, $F1(1,47) = 28.91$, $SEM = 3.0\%$, $F2(1,23) = 78.85$, $SEM = 2.4\%$ and of meaning, $F1(1,47) = 24.60$, $SEM = 1.8\%$, $F2(1,23) = 11.26$, 2.9%, but no main effect of decision type, $F1 \leqslant 1$, $F2(1,23) = 2.5$, $SEM = 2.4\%$, $p > .10$.

The decision time data (correct trials only) will be discussed as well, but it is important to keep in mind that they are of secondary importance. (The decision time

Table 2
Accuracy data and correct decision times (in ms), Experiment 1

Decision	Accuracy (in %)		Decision times (ms)	
	Agent	Patient	Agent	Patient
Reversible sentences (e.g., *the dog bit the man*)				
Actives				
Plausible	99	97	1596	1773
Implausible	99	91	1855	2005
Passives				
Plausible	88	92	1990	1927
Implausible	74	85	2075	2292
Nonreversible sentences (e.g., *the mouse ate the cheese*)				
Actives				
Plausible	100	98	1520	1851
Implausible	96	91	1946	2114
Passives				
Plausible	87	87	1826	1960
Implausible	82	84	2047	2202
Symmetrical sentences (e.g., *the woman visited the man*)				
Actives				
Version 1	94	97	1737	2032
Version 2	95	87	1771	2057
Passives				
Version 1	82	86	2211	2139
Version 2	76	83	2104	2172

data are less central in this and the other two experiments for two reasons: first, because accuracy is rather low in some conditions; second, as stated in Section 1, the fundamental purpose of the experiments is to obtain information about the actual content of people's interpretations rather than the time that was required to construct them.) For agent decisions, participants took longer with implausible sentences, but only for the actives. Participants required more time with the passives overall, but plausibility had little effect on the decision times. For patient decisions, participants took longer with passives than with actives, and they were faster when the meaning of the sentence was plausible. The three way interaction among form, meaning, and decision type was marginal by participants and significant by items, $F1(1,47) = 2.64$, $SEM = 66$ ms, $p = .10$, $F2(1,23) = 4.17$, $SEM = 63$ ms. None of the two-way interactions was significant, all p's $> .15$. Each of the main effects was highly significant: Participants took longer to make decisions about passives than actives (2071 vs. 1807 ms), $F1(1,47) = 27.52$, $SEM = 71$ ms, $F2(1,23) = 94.49$, $SEM = 42$ ms; longer with implausible sentences than plausible ones (2057 vs. 1822 ms), $F1(1,47) = 24.37$, SEM = 67 ms, $F2(1,23) = 49.48$, $SEM = 44$ ms; and they took longer to make patient decisions than agent decisions (1999 vs. 1879 ms), $F1(1,47) = 5.63$, $SEM = 71$ ms, $F2(1,23) = 14.97$, $SEM = 66$.

Participants had more difficulty identifying thematic roles both when a sentence was syntactically challenging, as in a passive, and when it was implausible. Interestingly, decisions were affected by plausibility even for active sentences, but in general, the effect of plausibility was larger for passives than actives. Certainly, overall, there was a major cost for the passive; apparently, even college students have some difficulty interpreting this particular noncanonical structure.

Nonreversible sentences. The data are shown in Table 2. Consider first just the accuracy data. For agent decisions, and for both active and passive forms, accuracy was lower when the sentence was implausible, and performance overall was worse with passives than with actives. For the patient decisions, plausibility had some effect on performance with the two structures, and again, accuracy was lower with passive forms. As the pattern of means suggests, there was no three-way interaction, $F1 < 1$, $F2(1, 23) = 1.52$, $SEM = 1.6\%$, $p > .20$. None of the two-way interactions was significant either, all p's $> .05$. There was no main effect of decision type (91% accuracy for agent decisions and 90% for patient decisions), both p's $> .15$. Thus, the only significant main effects were of structure (96% accuracy with actives and 85% with passives), $F1(1, 47) = 23.90$, $SEM = 3.3\%$, $F2(1, 23) = 46.94$, $SEM = 2.2\%$ and of meaning (93% accuracy with plausible sentences and 88% with implausible sentences), $F1(1, 47) = 9.66$, $SEM = 2.1\%$, $F2(1, 23) = 14.32$, $SEM = 1.9\%$.

The reaction time results are similar. Overall, participants took longer to make correct thematic decisions for passives than for actives (2009 vs. 1858 ms), $F1(1, 47) = 9.12$, $SEM = 71$ ms, $F2(1, 23) = 13.46$, $SEM = 81$ ms, and for implausible sentences than for plausible ones (2077 vs. 1789 ms), $F1(1, 47) = 22.11$, $SEM = 87$ ms, $F2(1, 23) = 26.96$, $SEM = 72$ ms. One difference, though, is that in the latency data there was a main effect of decision type: Participants took longer to make patient decisions than agent decisions (2032 vs. 1835 ms), $F1(1, 47) = 12.85$, $SEM = 78$ ms, $F2(1, 23) = 11.83$, $SEM = 90$ ms. None of the two-way interactions was significant, nor was the three-way, all p's $> .10$.

Thus, there is a cost for the passive structure overall, even when the arguments are nonreversible. Since the critical studies conducted by Slobin (1966), it has generally been believed that nonreversible sentences can be understood by going directly to the semantic roles without an intervening syntactic structure. However, it is possible that the overall cost for the nonreversible passive was observed here because the task in this experiment specifically required comprehenders to identify whatever the agent or patient was for a sentence as stated, even if the result made no sense (an approach that has been used with quite fruitful results within the context of the Competition Model, for example). An unfortunate side effect of this requirement might have been that it was not possible for participants to use the semantic information maximally and in a typical fashion, because atypically, the sentences sometimes expressed anomalous events.

To examine this possibility, a separate experiment was conducted in which participants received only plausible, nonreversible sentences in active or passive form and had to identify the agent or patient. All the fillers from the main experiment were included with these 24 experimental items. A total of 44 participants was tested, and the data from 40 were included in the analyses (four people did not understand the task in the practice session and therefore were dismissed from the experiment). The results were clear: For the active sentences, participants were 100% accurate at identifying the agent and 99% accurate at identifying the patient. For the passives, accuracy levels in the same two conditions were 90 and 95%, respectively. The interaction between syntactic form and decision type was marginal by participants, $F1(1,39) = 3.46$, $p < .07$, $SEM = 1.6\%$ and not significant by items, $F2(1,23) = 2.41$, $p > .10$, $SEM = 1.8\%$. There was a significant main effect of syntactic form, $F1(1,39) = 14.24$, $SEM = 1.9\%$, $F2(1,23) = (1,23) = 11.89$, $SEM = 2.0\%$, but no effect of decision type, both p's $> .20$. Reaction times associated with correct trials were 1699 ms for identifying the agent of the active, 1422 ms for identifying the patient of the active, 1916 ms for identifying the agent of the passive, and 1695 ms for identifying the patient of the passive. Both the main effects of structure ($F1(1,39) = 8.69$, $SEM = 76$ ms, $F2(1,23) = 18.05$, $SEM = 69$ ms) and decision type ($F1(1,39) = 35.29$, $SEM = 93$, $F2(1,23) = 72.67$, $SEM = 70$ ms) were significant, but the interaction was not, both F's < 1.

It appears, then, that even when participants deal only with nonreversible and plausible sentences (and a number of fillers, of course), there is still a cost for the passive sentences. Accuracy overall was high (over 90%), and in particular, higher than in the main experiment, but decisions were still influenced by syntactic complexity. Thus, even sentences which theoretically could be processed just lexically are difficult to understand because comprehenders seem to automatically use their syntactic knowledge—both their knowledge of syntactic associations such as the NVN strategy as well as syntactic algorithms which actually compute the correct and detailed syntactic form.

Symmetrical sentences. The data are shown in Table 2. Note that the third variable in the $2 \times 2 \times 2$ ANOVA performed on these data is arrangement of arguments rather than plausibility, because both versions were shown in norms to be equally plausible. (The first variable is syntactic form and the second is agent vs. patient/theme decision, as in the previous analyses.) The pattern of results is now familiar: Participants were equally accurate at identifying the agent and patient of the active, and performance overall was high (95 and 92%, respectively). Accuracy was lower for the passive, and lower for identifications of the agent than the patient (79% vs. 85%) of those passives. The interaction between syntactic form and decision type was significant by participants, $F1(1,47) = 5.63$, $SEM = 2.5\%$ but not by items, $F2(1,23) = 1.20$, $p > .25$, $SEM = 3.8\%$. There was also a main effect of syntactic form, $F1(1,47) = 25.80$, $SEM = 3.2\%$, $F2(1,23) = 17.92$, $SEM = 4.2\%$: Decisions were more accurate for actives (93%) than for passives (82%). There was no main effect of decision type, both F's < 1.

There was an effect of the order variable: Accuracy was 85% for the order arbitrarily designated as "one," and 90% for the order arbitrarily labeled as "two," $F1(1,47) = 6.75$, $SEM = 2.5\%$, $F2(1,23) = 3.09$, $p < .09$, $SEM = 3.6\%$. In other words, although it appears that it is as likely that a customer would thank a clerk as the contrary, these participants apparently thought otherwise. This result is unexpected given that the norms did not reveal any significant differences between the two orders, but of course it is a different task to judge the plausibility of a sentence than to identify individual thematic roles. This order variable did not participate in either of the possible two-way interactions, all p's $> .20$. However, the three-way interaction was reliable by participants, $F1(1,47) = 5.78$, $SEM = 2.2\%$, $F2(1,23) = 3.32$, $p < .08$, $SEM = 2.5\%$.

The results for decision times are shown in Table 3. Latencies were longer for passives than for actives, $F1(1,47) = 20.81$, $SEM = 80$ ms, $F2(1,23) = 32.28$, $SEM =$ 71 ms. Participants required about 2156 ms to respond to the passives, and 1899 to respond to the actives. Decision times for passives were not affected by decision type (agent or patient), but active decisions were, $F1(1,47) = 7.29$, $SEM = 77$ ms, $F2(1,23) = 9.68$, $SEM = 62$ ms for the interaction between syntactic form and decision type. All other possible effects were not significant, all F's < 1. The data for the symmetrical sentences, then, are consistent with the hypothesis that passives are difficult to understand.

The main findings from this experiment are the following. First, for all sentence types—reversibles, symmetricals, and (contrary to Slobin, 1966) nonreversibles—passives are more difficult to understand than actives. The nature of the errors people made also makes clear that the problem people have with the passive is keeping track of the bindings between syntactic positions and thematic roles: On every trial of the experiment, when people provided the wrong answer in the decision task, their mistake was to give the other concept in the sentence—the argument filling the other thematic role. In addition, it is more difficult to maintain algorithmically based thematic role assignments in the passive when the resulting interpretation is inconsistent with schematic knowledge (Rumelhart, 1980). This can be seen in the worse performance on implausible, reversible passives compared with actives, and for symmetrical passives compared with symmetrical actives. In both these cases, people's schemas suggest an alternate set of thematic roles, and perhaps because the NVN strategy used by the pseudo-parser reinforces those alternative assignments, syntactically computed thematic roles in passives start to break down.

The results of this experiment make clear why it seems necessary to propose that people use both simple heuristics and syntactic algorithms to understand sentences. A model that included only the algorithmic component would be inadequate because the experiment shows that people misinterpret passive sentences fairly often. Moreover, their mistake seems to be to misassign the thematic roles, as if the NVN strategy (which is merely a syntactic association and not a syntactic procedure) was used on the input rather than the right, detailed syntactic operations. A heuristics-only model cannot be right either, because people more often than not do get the right analysis of the sentence. Thus, with enough time (and recall that passives almost invariably take longer to process than actives) people often compute the real structure for the passive and therefore successfully interpret it.

3. Experiment 2

The first experiment showed that a noncanonical syntactic structure such as the passive is difficult to comprehend. The goal of the second experiment was to determine whether frequency of surface form is important, or if rather what matters is whether thematic roles are assigned to syntactic positions in the most common way (i.e., in accordance with the NVN strategy). Two noncanonical structures were compared: the subject-cleft and the passive. Consider *it was the man who bit the dog* and *the dog was bitten by the man*. The first sentence, the subject-cleft, occurs quite rarely (Nelson, 1997; Quirk, Greenbaum, Leech, & Svartvik, 1985). It is also more syntactically complex than the ordinary active, because (according to transformational theories of grammar), the phrase *the man* has been moved from its initial position as subject of *bit* (thus the term "subject-cleft"). The moved phrase leaves behind a trace, and the trace must be coindexed with the moved constituent. In addition, the cleft structure is a device for focusing information in discourse (Ball, 1994; Delahunty, 1984; Nelson, 1997): The moved phrase is the informational focus, and the sentence carries a strong presupposition that the event described in the rest of the sentence (e.g., biting of a dog) is already established in the discourse. Thus, another reason why the subject-cleft might be complex is that it requires a particular discourse context to make it felicitous.

The interesting question, then, is whether the subject-cleft behaves more like an active or a passive sentence in these particular experiments that employ a thematic role decision task. It is like a passive in that it is infrequent and tightly discourse constrained, but it is similar to an active because thematic roles ultimately are assigned to constituents in a frequent and typical order (agent and then patient). To evaluate the difficulty of the subject-cleft, one of two strategies could be adopted: One is to compare the subject-cleft to the active in the same experiment, and the other is to compare the subject-cleft to the passive. It is not practical to include all comparisons in a single experiment, because the design would be exceedingly complex. Therefore, the decision was made to run Experiment 2 as a comparison between subject-clefts and passives. This choice was made because intuition suggested that subject-clefts would be quite easy, and so an experiment comparing actives and subject-clefts might simply have yielded a null effect (indeed, a cross-experiment comparison will demonstrate just that). This second experiment, then, had the same design as Experiment 1, except that the active condition was replaced with a subject-cleft condition. The materials were otherwise identical, as was the task.

3.1. Method

Participants. A total of 44 undergraduates attending Michigan State University participated in the experiment in exchange for partial credit in their Introductory Psychology courses. Four of the participants were not included in the data analyses because they were non-native speakers of English, two because they could not understand the task (as indicated by their performance on the practice trials), and two because they did not speak loudly enough to reliably set off the voice-activated relay switch. Thus, the data from 32 participants were included in the data analyses.

Materials. The materials for this experiment were identical to the ones used for Experiment 1, except that the active versions were modified to be subject-clefts (for both practice and nonpractice items). Thus, for each of the 72 experimental sentences, a sentence such as *the man bit the dog* was changed to *it was the man that bit the dog*.

Procedure. The procedure was the same as that used for the second experiment.

Design. The experiment employed a $2 \times 2 \times 2$ within-participants design: The syntactic form of the sentence was either subject-cleft or passive, the sentence's meaning varied (plausible vs. implausible for the biased reversible items, plausible vs. anomalous for the nonreversible items, and order 1 vs. order 2 for the symmetrical items), and the participant made a decision either about the agent or patient/theme. Each of the three types of sentences—reversible but biased, nonreversible, and symmetrical—was analyzed separately.

3.2. Results

Reaction times shorter than 300 ms and longer than 7500 ms were eliminated. These criteria resulted in the removal of less than 4% of the data. Means were computed for all eight conditions for both participants and items, and analyses of variance were performed with both participants ($F1$) and items ($F2$) as random effects. All effects are significant at $p < .05$ unless otherwise indicated.

Reversible, biased sentences. The percentage of decisions that were correct in each of the eight conditions is shown in Table 3. For agent decisions, accuracy was the same for plausible and implausible subject-cleft sentences (97% for both). Thus, people had no trouble stating that the man was the agent in a sentence such as *It was the man who bit the dog*. With passives, participants were less accurate overall (72%), and less accurate for sentences that were implausible (68% vs. 77%). For patient decisions, a similar pattern held: high accuracy in the subject-cleft condition, and no effect of plausibility (95 and 93% for the plausible and implausible conditions, respectively); lower accuracy overall for the passives, and a large effect of plausibility (87% accuracy in the plausible condition, 74% in the implausible condition). Comparing performance in the agent and patient conditions, it can be seen that there was a tendency for participants to be more accurate with the first encountered role: For actives performance was (nonsignificantly) better for the agent than for the patient, and for passives the opposite was true.

The ANOVA revealed a significant two-way interaction between syntactic structure and plausibility, $F1(1,31) = 7.48$, $SEM = 3.1\%$, $F2(1,23) = 12.37$, $SEM = 2.4\%$: The effect of plausibility was absent for subject-clefts but large for passives. There was also a significant interaction between form and thematic role decision, $F1(1,31) = 5.67$, $SEM = 3.2\%$, $F2(1,23) = 3.75$, $p > .06$, $SEM = 4.0\%$: Performance on subject-clefts was equally good for both thematic roles, but for passives, performance was much worse with agents than patients. The three-way interaction in this experiment was not significant, both F's < 1. The main effect of structure was significant, $F1(1,31) = 30.26$, $SEM = 4.9\%$; $F2(1,23) = 50.58$, $SEM = 3.8\%$, the effect of plausibility was marginal, both p's $> .08$, and there was no main effect of thematic role decision, both F's < 1.

Decision times were longer overall for passives than for subject-clefts (2004 vs. 1640 ms), $F1(1,31) = 31.36$, $SEM = 92$ ms, $F2(1,23) = 17.09$, $SEM = 127$ ms, longer for implausible sentences than for plausible sentences (1904 vs. 1740 ms), $F1(1,31) = 6.13$, $SEM = 94$ ms, $F2(1,23) = 6.69$, $SEM = 94$ ms, and longer for

Table 3
Accuracy data and correction decision times (in ms), Experiment 2

Decision	Accuracy (in %)		Decision times (ms)	
	Agent	Patient	Agent	Patient
Reversible sentences (e.g., *the dog bit the man*)				
Subject-clefts				
Plausible	97	95	1411	1593
Implausible	97	93	1570	1984
Passives				
Plausible	77	87	1879	2076
Implausible	68	74	1960	2102
Nonreversible sentences (e.g., *the mouse ate the cheese*)				
Subject-clefts				
Plausible	100	95	1309	1638
Implausible	95	87	1634	1974
Passives				
Plausible	98	89	1823	1969
Implausible	76	74	2181	2288
Symmetrical sentences (e.g., *the woman visited the man*)				
Subject-clefts				
Version 1	98	93	1523	2003
Version 2	98	90	1434	1853
Passives				
Version 1	77	88	2057	2209
Version 2	71	80	2099	2253

patient decisions than for agent decisions (1939 vs. 1705 ms), $F1(1,31) = 12.75$, $SEM = 92$ ms, $F2(1,23) = 8.76$, $SEM = 102$ ms. There was a significant interaction between structure and plausibility by participants only, $F1(1,31) = 4.18$, $SEM = 77$ ms, $F2(1,23) = 1.03$, $SEM = 136$ ms.

Nonreversible sentences. The results are shown in Table 3. Accuracy overall was lower for passives than for subject-clefts (84% vs. 94%), ($F1(1,31) = 14.43$, $SEM = 3.7\%$, $F2(1,23) = 31.56$, $SEM = 2.5\%$). Accuracy was lower for implausible compared with plausible sentences (83% vs. 95%), $F1(1,31) = 25.94$, $SEM = 3.5\%$, $F2(1,23) = 27.60$, $SEM = 3.4\%$. Decisions about the agent were made more accurately than those about the patient (92% vs. 86%), $F1(1,31) = 6.05$, $SEM = 3.6\%$, $F2(1,23) = 8.62$, $SEM = 3.0\%$. The only other significant effect was the interaction between syntactic form and meaning, $F1(1,31) = 7.01$, $SEM = 3.1\%$, $F2(1,23) = 7.26$, $SEM = 3.0\%$. For subject-clefts, accuracy was higher for plausible sentences than for implausible sentences (97% vs. 91%); the same was true for passives, but the difference was much larger (93% vs. 75%).

Decision times for the nonreversible sentences were longer overall for passives than for subject-clefts (2065 vs. 1639 ms), $F1(1,31) = 43.08$, $SEM = 92$ ms, $F2(1,23) = 13.83$, $SEM = 156$ ms, longer for implausible sentences than for plausible ones (2019 vs. 1685 ms), $F1(1,31) = 25.73$, $SEM = 93$ ms, $F2(1,23) = 23.29$, $SEM = 102$ ms, and longer for patient decisions than agent decisions (1967 vs.

1737 ms), $F1(1,31) = 14.12$, $SEM = 87$ ms, $F2(1,23) = 8.05$, $SEM = 100$ ms. No other effects were significant.

Symmetrical sentences. The results are shown in Table 3. As in the previous experiment, accuracy was different for the two versions even though in the norming study they were judged to be equally plausible. Accuracy overall was 89% for the version 1 sentences and 85% for the version 2 sentences, $F1(1,31) = 4.13$, $SEM = 2.9\%$, $F2(1,23) = 5.41\%$, 2.5%. More importantly, accuracy was higher for subject-clefts than for passives (95% vs. 79%), $F1(1,31) = 22.41$, $SEM = 4.7\%$, $F2(1,23) = 90.78$, $SEM = 2.3\%$. Accuracy was the same overall for agent and patient decisions, both F's < 1, but the interaction between syntactic structure and decision type was significant, $F1(1,31) = 8.85$, $SEM = 4.0\%$, $F2(1,23) = 14.15$, $SEM = 3.1\%$. For subject-clefts, participants were more accurate making decisions about the agent than about the patient (98% vs. 91%). For passives, the opposite was true (74 and 84%).

Decision times were longer for passives than for subject-clefts (2154 vs. 1703 ms), $F1(1,31) = 41.89$, $SEM = 99$ ms, $F2(1,23) = 17.63$, $SEM = 130$ ms, and longer for patient decisions than agent decisions (2080 vs. 1778 ms), $F1(1,31) = 10.86$, $SEM = 129$ ms, $F2(1,23) = 10.37$, $SEM = 111$ ms. Decision times did not differ for the two versions, both F's < 1. The only other significant effect was an interaction between syntactic form and decision type, $F1(1,31) = 5.69$, $SEM = 88$ ms, $F2(1,23) = 3.47$, $SEM = 151$ ms, $p < .08$. For both subject-clefts and passives, participants took longer to make patient than agent decisions. However, the difference was larger for subject-clefts (1928 vs. 1479 ms) than for passives (2231 vs. 2078 ms).

The main results of this experiment replicated those from Experiment 1. Performance was worse for passive sentences, particularly when they expressed implausible events. In addition, Experiment 2 demonstrates that a structure may be rarely encountered but still properly understood. The subject-cleft is an infrequently used form, yet comprehenders had little difficulty identifying its constituent thematic roles accurately. It appears that what makes a structure canonical is not its frequency, but rather whether thematic roles are assigned in a standard order (proto-agent before proto-patient) or an unconventional order (proto-patient before proto-agent). This result is exactly as predicted by a model that incorporates the NVN strategy as a basic heuristic for language processing. To reinforce this point, the results of the last two experiments were directly compared.

3.3. Between-experiment comparison between actives and subject-clefts

Experiment 2 showed that the subject-cleft is a much easier structure for comprehenders to process than is the passive. In addition, an informal examination of the numbers seems to suggest that the subject-clefts are no more difficult than regular actives. To formalize this comparison, the active conditions from Experiment 1 were compared with the subject-cleft conditions from Experiment 2. Only the data from the first 32 participants from Experiment 1 were included, so that the experiments would have an equal number of participants. The result is a design much like the ones used in Experiments 1 and 2, except that the structure variable is now between-participants rather than within. Because this analysis is exploratory, and to maximize the chances of finding any differences, the three different sentence types—biased reversible, nonreversible, and symmetrical—were combined.

The overall accuracy for actives and subject-clefts was 96 and 98%, respectively. Thus, any numerical difference actually disfavors the more common structure, but this difference was not significant (both p's $> .25$). Indeed, the only significant effect was that agent decisions were made more accurately than patient decisions (98% vs. 94% correct), $F1(1,62) = 9.95$, $SEM = 1.4\%$. In addition, as a comparison of Tables 2 and 3 reveals, decision times were no slower for subject-clefts than for actives. Therefore, all the available evidence suggests that people can maintain their thematic role assignments as easily in the subject-cleft as in a regular active sentence.

At least on its face, this finding that subject-clefts are every bit as easy to process as actives is incompatible with strongly discourse-based models of processing and processing ease. For example, Hoeks et al. (2002) have argued that reading times for some types of apparently syntactically complex sentences are long not because of their grammatical form but because they have a particularly demanding topic-comment structure. Hoeks et al. found that sentences that presupposed the existence of two topics were harder to understand than those that presupposed just one, but this difference disappeared in a discourse context that established two entities as possible topics. The results here comparing actives and subject-clefts are inconsistent with this argument. The subject-cleft strongly presupposes the existence of a topic; for example, in *It was the man who bit the dog*, the presupposed topic is that an event of dog-biting took place, and what needs to be filled in is the agent. In contrast, the active has a much weaker focus-presupposition structure (Rochemont, 1986; Rooth, 1992; Selkirk, 1984). Yet, these sentences were equivalently easy to process. Of course, the sentences in the current experiments were presented aurally whereas those in Hoeks et al. were shown visually; moreover, the tasks used are very different (thematic role decision versus self-paced reading in Hoeks et al.). Nevertheless, it is still worth noting that a task that taps into the semantic processing of sentences does not reveal any difference between actives and active-clefts, even though the latter contains presuppositions that can only be satisfied in a particular discourse context.

4. Experiment 3

The preceding experiment demonstrated that subject-clefts behave just like regular active sentences even though they are much less frequent. This finding in turn suggests that what is important to the language processing system is whether the assignment of thematic roles in the sentence conforms to the NVN/agent-patient strategy. Therefore, one would expect object-clefts to be as difficult as passives, because they require thematic roles to be assigned in a way that violates that strategy: proto-patient to subject and then proto-agent to object. In addition, if it is true that what makes a given English structure difficult to process is its deviation from the standard agent-patient template, then this experiment directly comparing subject-clefts to object-clefts will also provide an opportunity to see how the results of the previous experiments replicate, and to obtain more data concerning the interpretation of simple and complex sentences that are either plausible or implausible.

4.1. Method

Participants. A total of 49 participants from the same subject pool as in the previous experiments participated in exchange for partial credit in their Introductory Psychology courses. Three people were excluded from the data analyses because they were not native speakers of English, five because they did not understand the task (as revealed by their performance on the practice trials), and one because he or she did not speak loudly enough to consistently set off the voice-activated relay switch. The data from the remaining 40 participants were included in these analyses.

Materials. The materials for this experiment were the same as those used in Experiment 2, except that the passive versions were modified to be object-clefts (for both practice and nonpractice items). For example, a sentence such as *the dog was bitten by the man* was changed to *it was the dog the man bit*.

Procedure. The procedure was the same as the one used in Experiments 1 and 2.

Design. The experiment employed a $2 \times 2 \times 2$ within-participants design. The syntactic form of the sentence was either subject-cleft or object-cleft, the sentence's meaning varied (plausible vs. implausible for the biased reversible items, plausible vs. anomalous for the nonreversible items, and order 1 vs. order 2 for the symmetrical items), and the participant made a decision either about the agent or patient/theme. Each of the three types of sentences—reversible but biased, nonreversible, and symmetrical—was analyzed separately. Reaction time data include only trials on which participants answered the question correctly.

4.2. Results

Reaction times shorter than 300 ms and longer than 7500 ms were eliminated. These criteria removed less than 3% of the data. Means were computed for all eight conditions for both participants and items, and analyses of variance were performed with both participants ($F1$) and items ($F2$) as random effects. All effects are significant at $p < .05$ unless otherwise indicated.

Reversible, biased sentences. The percentage of decisions that were correct in each of the eight conditions is shown in Table 4. There was a two-way interaction between syntactic form and plausibility, $F1(1,39) = 11.01$, $SEM = 3.4\%$, $F2(1,23) = 13.20$, $\text{SEM} = 3.1\%$. Performance was better for subject-clefts than for object-clefts (94% vs. 76%; for the main effect of structure, $F1(1,39) = 46.58$, $SEM = 3.9\%$, $F2(1,23) = 53.77$, $SEM = 3.5\%$), and accuracy did not depend on plausibility for the subject-clefts (3% cost for implausibility) whereas it did for the object-clefts (19% cost). There was also a two-way interaction between syntactic form and thematic role decision, $F1(1,39) = 5.99$, $SEM = 3.9\%$, $F2(1,23) = 6.68$, $SEM = 3.5\%$. For actives, people were 7% more accurate at making agent than patient decisions; for passives, the reverse was true. The only other significant effect was a main effect of plausibility, $F1(1,39) = 23.46$, $SEM = 3.2\%$, $F2(1,23) = 14.05$, $SEM = 3.9\%$.

Table 4
Accuracy data and correction decision times (in ms), Experiment 3

Decision	Accuracy (in %)		Decision Times (ms)	
	Agent	Patient	Agent	Patient
Reversible sentences (e.g., *the dog bit the man*)				
Subject-clefts				
Plausible	98	93	1378	1683
Implausible	97	89	1503	1888
Object-clefts				
Plausible	80	90	2131	2218
Implausible	64	68	1922	2348
Nonreversible sentences (e.g., *the mouse ate the cheese*)				
Subject-clefts				
Plausible	100	95	1301	1868
Implausible	94	82	1911	2342
Object-clefts				
Plausible	88	94	1784	2216
Implausible	55	62	2220	2256
Symmetrical sentences (e.g., *the woman visited the man*)				
Subject-clefts				
Version 1	98	95	1474	1851
Version 2	94	89	1682	1939
Object-clefts				
Version 1	79	78	2222	2309
Version 2	71	78	1971	2320

Accuracy overall in the plausible condition was 90%, and in the implausible condition it was 80%. There was no three-way interaction between syntactic form, plausibility, and thematic role decision.

Decision times (correct trials only) were longer overall for object-clefts than for subject-clefts (2154 vs. 1613 ms), $F1(1,39) = 34.98$, $SEM = 129$ ms, $F2(1,23) = 48.79$, $SEM = 133$ ms, and longer for patient decisions than agent decisions (2034 vs. 1733 ms), $F1(1,39) = 7.93$, $SEM = 151$ ms, $F2(1,23) = 7.60$, $SEM = 147$ ms. There were no significant interactions (all p's $> .05$).

Nonreversible sentences. The results are shown in Table 4. The accuracy data were very similar for these sentences as for the biased, reversible sentences. The interaction between syntactic form and meaning was significant, $F1(1,39) = 31.59$, $SEM = 2.9\%$, $F2(1,23) = 26.08$, $SEM = 3.2\%$. The cost of implausibility was only 10% for subject-clefts but 33% for object-clefts. There was also the same two-way interaction between syntactic form and thematic role decision, $F1(1,39) = 8.40$, $SEM = 3.8\%$, $F2(1,23) = 8.32$, $SEM = 3.7\%$. People were 9% more accurate to make agent decisions when the sentence was a subject-cleft, and 7% more accurate to make patient decisions when the sentence was an object-cleft. No other interactions were significant, all p's $> .50$. There was a main effect of syntactic form in the expected direction (93% accuracy for subject-clefts and 75% for object-clefts), $F1(1,39) = 52.78$, $SEM = 3.5\%$, $F2(1,23) = 48.68$, $SEM = 3.6\%$, and a predictable effect of plausibility (94% for plausible items and 73% for implausible items),

$F1(1,39) = 62.96$, $SEM = 3.8\%$, $F2(1,23) = 79.943$, $SEM = 3.3\%$. There was no main effect of thematic role decision (both p's $> .50$).

Decision times were longer overall for object-clefts than for subject-clefts (2119 vs. 1855 ms), $F1(1,39) = 3.61$, $SEM = 196$ ms, $.05 < p < .10$, $F2(1,23) = 9.26$, $SEM = 226$ ms, longer for implausible sentences than plausible ones (2182 vs. 1792 ms), $F1(1,39) = 4.98$, $SEM = 247$ ms, $F2(1,23) = 14.61$, $SEM = 207$ ms, and overall longer for patient decisions than agent decisions (2170 vs. 1804 ms), $F1(1,39) = 8.56$, $SEM = 177$ ms, $F2(1,23) = 12.90$, $SEM = 179$ ms. None of the interactions was significant, all p's $> .20$.

Symmetrical sentences. The results are shown in Table 4. The only significant effects for these items were two main effects. The first is the by-now quite familiar effect of syntactic form: Accuracy was higher for actives (94%) than for passives (77%), $F1(1,39) = 37.85$, $SEM = 4.0\%$, $F2(1,23) = 57.63$, $SEM = 3.1\%$. As in the previous experiment, even though these sentences were constructed to be equally plausible for both arrangements of the arguments (according to the norming data described in Experiment 1), there was a main effect of order, $F1(1,39) = 7.62$, $SEM = 2.3\%$, $F2(1,23) = 4.49$, $SEM = 3.1\%$: People were more accurate with order 1 (88%) than order 2 (83%). All interactions were insignificant, p's $> .10$.

For decision times, there was a main effect of structure, $F1(1,39) = 19.70$, $SEM = 149$ ms, $F2(1,23) = 27.64$, $SEM = 153$ ms. Decisions were made faster for subject-clefts (1736 ms) than for object-clefts (2205 ms). There was also a main effect of thematic role decision, $F1(1,39) = 4.84$, $SEM = 172$ ms, $F2(1,23) = 8.97$, $SEM = 120$ ms. Patient decisions were made more slowly than agent decisions (2201 vs. 1947 ms). No other effects were significant, all p's $> .10$.

The findings from Experiment 3 complement those obtained from the other two experiments. Object-clefts are different from passives because they do not have the passive verb morphology, nor is the agent incorporated into a structure using a prepositional phrase headed by *by*. But in an object-cleft such as *It was the dog the man bit*, the proto-patient precedes the proto-agent, and so the structure is incompatible with the NVN strategy. Yet, despite the many important syntactic differences between these object-clefts and passives, they are processed similarly. What the results of Experiment 3 suggest, then, is that regardless of the surface details of the sentence's structure, the parser finds it easier to assign a correct interpretation if it is compatible with the NVN strategy. Thus, as has been argued by Townsend and Bever (Townsend & Bever, 2001; see also Bever, 1970), this simple pseudo-parsing preference is enormously powerful. Indeed, as they point out, this humble principle can account for quite a large amount of data in psycholinguistics, which makes clear that in addition to sophisticated parsing strategies the comprehension system must also rely on a set of fast and frugal heuristics (Gigerenzer et al., 1999).

5. General discussion

The central findings from the three experiments are summarized in Table 5. Perhaps the most obvious generalization is that sentences requiring the patient thematic role to be assigned before the agent thematic role are more difficult to understand than those that conform to the more common pattern. People identify the thematic roles in passives and object-clefts more slowly than in actives and subject-clefts. They also err more with the patient-before-agent utterances, and almost invariably their mistake is to name the other thematic role from the sentence. It appears, then, that sentences that require the comprehender to violate the NVN strategy are not as easy to understand as those that do not. These results support the claim that comprehension is based on heuristics, and in this case a particularly simple one.

It is remarkable that the difficulty of noncanonical structures held up for all sentence types, including the "nonreversible" sentences. Since the demise of the Derivational Theory of Complexity (see Garnham & Oakhill, 1987, for discussion) it has been argued (Slobin, 1966; Harley, 2001) that nonreversible sentences are easy to understand even in the passive form because the animacy contrast allows the comprehender to assign roles using the strong animacy cue (essentially by avoiding any type of structural analysis altogether, including even a heuristic like the NVN strategy). Yet nonreversible passives and nonreversible object-clefts were consistently processed more slowly than their active and subject-cleft counterparts. In addition, the effect of the semantic manipulation was not as clear as one would have expected based on the earlier studies examining nonreversible passives. In the first experiment, there was no interaction between structure and meaning for the nonreversible items,

Table 5
Main results from the three experiments

Experiment	Compared	Results
Experiment 1	Active vs. passive	• For every sentence type, passives comprehended less accurately • For every sentence type, passives comprehended more slowly • For biased, reversible sentences, implausible versions more difficult when in passive form
Experiment 2	Subject-clefts vs. passives	• For every sentence type, passives comprehended less accurately • For every sentence type, passives comprehended more slowly • For both biased, reversible and nonreversible sentences, implausible versions more difficult when in passive form
Experiment 1 vs. 2	Actives vs. subject-clefts	• Two forms are equally easy, despite the low frequency of the subject-cleft
Experiment 3	Subject-clefts vs. object-clefts	• For every sentence type, object-clefts comprehended less accurately • For every sentence type, object-clefts comprehended more slowly • For both biased, reversible and nonreversible sentences, implausible versions more difficult when in object-cleft form

in contrast with the results obtained for the biased, reversible items. That is, the implausibility (or rather, anomaly) cost for nonreversible sentences was the same for actives and passives, whereas for biased, reversible sentences it was larger for passives. This pattern is along the lines of what one would expect if the strong semantic constraint provided by animacy is sufficient to allow thematic roles to be assigned directly. But the second and third experiments yielded a different pattern: The cost for implausibility/anomaly was much larger when the sentence was in the passive form, and this difference held for both nonreversible and biased, reversible items.

What accounts for the unstable effects of the animacy contrast? Of course, at this point it is only possible to speculate, but two ideas suggest themselves. One is that any heuristic based on plausibility is simply weaker than the NVN strategy, and so its effects are not as easy to detect and are easily overwhelmed by people's heavier reliance on word order. The correlational analyses to be described in the next section (see Table 6) support this assumption that the NVN heuristic is weighted more heavily than is plausibility. Another possibility is that animacy should be viewed as a continuous rather than a binary variable, and as a fluid rather than a rigid concept. For example, cars, alarm clocks, and computers are inanimate objects, but they can be conceptualized as instigators of actions. Indeed, if one defines an agent as a system that can be in states and that can undertake actions (Minut, 2000; Sutton & Barto, 1998), then many objects that are traditionally classified as inanimate would have to be given a less humble status, at least in some sentential contexts (e.g., *the program removed the virus*). Moreover, with just a little imagination it is not difficult to imagine scenarios in which inanimate objects behave like agents. For example, one participant reported that when she got a sentence such as *the cheese ate the mouse* she imagined it came from the books and videos she encountered as a child, in which inanimate objects often have the ability to perform many human-like actions. It is in part for this reason that Dowty (1991) proposed the notion of a proto-agent. In the past, to deal with sentences such as *the alarm clock woke me up or the VCR ate my tape* linguists took pains to distinguish mere instruments from agents, arguing that in these examples the subjects were the former. But it is clear that it can be quite difficult to distinguish these two thematic roles from each other, and therefore a simpler approach is to argue, as Dowty has, that in any sentence describing an event there is one entity that is more agent-like than the other. On this sort of view, the varying effects of the animacy contrast in the current set of experiments is not necessarily surprising, because whether participants treated the nonreversible items as qualitatively different from the others would depend on their own individual predilections and how they chose to exercise their powers of imagination.

Another consistent result from the experiments is that comprehenders made fewer mistakes with thematic roles that they assigned first. Across the experiments, accuracy was generally higher for agents in actives and subject-clefts, and for patients in passives and object-clefts. One possible explanation for this tendency is that the first mentioned entity in the sentence has a privileged status in memory (Gernsbacher, 1990). Another comes from the Competition Model of language comprehension (MacWhinney & Bates, 1989; MacWhinney et al., 1984). A fundamental assumption of that approach is that the listener or reader understands a sentence from the perspective of the first mentioned entity, and the rest of the sentence is assimilated into that cognitive schema.

All the data presented here showing that passives and object-clefts are difficult to understand might lead one to speculate that the entire semantic representation for these structures is fragile and potentially up for misinterpretation. This idea is not consistent with the approach proposed here. Passives and other structures that are incompatible with the NVN strategy are difficult to understand because thematic role assignment is tricky. The NVN strategy yields one set of assignments, but algorithmic processing yields the opposite. For people to successfully understand such sentences, then, they must allow both the second stage algorithmic parse to run to completion, and they must reconcile the outputs of the pseudo- and the true parse. According to this approach, then, comprehenders should not be confused about aspects of the sentence's meaning that do not involve the thematic roles at issue in the NVN strategy—that is, the two arguments adjacent to the verb, namely the proto-agent and the proto-patient. To examine this prediction further, additional analyses of the data from the three experiments were performed. The 144 fillers included 36 active sentences and 36 passive sentences for which participants responded to either a color or a location probe. For example, *the blue book was on the shelf* is an active sentence for which a color or a location probe would be appropriate. The same is true for the passive filler *the pedestrian was hit by a yellow car on Fifth Avenue*. These fillers then make it possible to examine whether it is more difficult to identify the location or color concept in a passive sentence compared to an active. The results were unambiguous: For active sentences, accuracy at identifying the correct color was 99%, and accuracy was the same for locations. For passives, the values for color and location were 98 and 99%, respectively. These values did not differ, all F's < 1. Decision times did not differ either. Therefore, it is clear that what comprehenders have trouble keeping straight in the passive is the two roles immediately adjacent to the verb, which are precisely the two implicated in the NVN strategy. The representation for other thematic roles and semantic concepts seems to be as accurate as it is in active sentences.

The final result from the experiments that deserves comment is the intriguing finding that subject-clefts are no harder to understand than regular active sentences. It appears, then, that the base frequency of a global syntactic form is not what influences how accurately people process a sentence in a task like the one used in these experiments. What is critical is whether thematic roles are assigned in the most familiar way. A recent report concerning the way that experiencer verbs are processed provides further evidence for this view (Cupples, 2002). Participants were asked to read sentences with theme–experiencer verbs such as *amuse*. These verbs are unusual because in the active form the subject is a theme and the object is an experiencer (Ferreira, 1994). Therefore, sentences with these verbs should be rather difficult to process if people use an NVN heuristic during comprehension. The results were consistent with this prediction. People took longer to judge the plausibility of active sentences with theme–experiencer verbs than similar sentences with experiencer–theme verbs (e.g., *like*, which conforms to the canonical proto-agent before proto-theme template). Moreover, passives with theme–experiencer verbs were read faster than passives with other verb types. Similar results were obtained with other measures.

The Cupples (2002) study provides further evidence for people's reliance on an NVN heuristic. When someone encounters a sentence such as *the joke amused the host*, application of the NVN strategy will tempt the comprehender to make the subject a proto-agent (a term which is meant to encompass experiencer, as argued earlier) and the object a proto-patient (i.e., theme). That tendency must be overridden by plausibility and real syntactic analysis (the latter process will recover the syntactic information associated with the verb *amuse* which specifies the way thematic roles are assigned to grammatical positions). On the other hand, for the passive sentence *the host was amused by the joke*, the NVN strategy is consistent with the correct assignment of thematic roles in this case. As a result, this type of passive sentence is easier than one that includes a verb with a more standard thematic structure. It is important to note that in this study people did take longer to process passives than actives overall, even when the passives included a theme–experiencer verb. This result is not surprising, however, because heuristics such as the NVN strategy are not used instead of syntactic rules and principles; rather, the system somehow uses both heuristics and algorithms to arrive at an interpretation of a sentence.

6. Implications for models of language comprehension

The results of these experiments are inconsistent with any model of language comprehension which assumes that when a sentence is syntactically unambiguous, only the structure compatible with its syntactic form is generated (e.g., Frazier & Clifton, 1996; MacDonald et al., 1994; Trueswell et al., 1993). For example, according to these models, a sequence such as *Mary put the book on the table* would activate both the interpretation where *on the table* is a destination and the one where *on the table* is the modifier of book (*Mary put the book on the table onto the shelf*). Plausibility, contextual information, and other nonsyntactic sources of knowledge can influence the availability of one interpretation over the other. But what the sequence will not trigger (for instance) is the idea that the **table** is the thing that Mary moved, because no syntactic structure licenses that interpretation. (Indeed, this is the logic behind using unambiguous versions of garden-path sentences as the control conditions in experiments: The assumption is that the processor will be forced to a particular structure in the unambiguous case, and so performance when there is structural ambiguity can be compared to this baseline condition.) This premise is the fundamental idea behind most current models of sentence comprehension: An interpretation must be based on a syntactic frame, even if the activation level of that frame can be influenced by nonsyntactic sources of information.

On the other hand, the finding that people almost always construct the implausible interpretation of active and subject-cleft sentences rules out the complement of the syntax-only approach, which would be any model that attempted to get by on semantics alone. In "semantics-only" models, syntactic structures are rarely built. Instead, comprehenders adopt whatever interpretation is most plausible, violating rules of grammar along the way. Today this model is somewhat of a straw position; few would postulate that constraints as reliable and valid as those that come from syntax are ignored by the comprehension system. It is important to note, however, that this model has had its proponents, particularly among those who focus on the comprehension of multi-sentence discourses (Lehnert & Ringle, 1982; Osgood, 1980; Schank, 1972). Moreover, the general idea that people sometimes misinterpret sentences might lure one to an account that eschews syntax and allows lexical and discourse semantics to do all the work of language comprehension. The experiments presented here make clear that syntactic structures are computed during sentence comprehension: Not only do people almost always obtain the correct meaning of actives and active-clefts, they also take less time to process them than their passive and object-cleft counterparts, and they are much less likely to be misinterpreted.

The results from the present set of experiments suggest that the language comprehension system uses simple heuristics to process sentences, and somehow coordinates the output of those heuristics with the products of more rigorous syntactic algorithms. It remains for future research to determine how the heuristic and algorithmic systems are coordinated. As this article focuses on the importance of heuristics in sentence processing, the next question that will be addressed is how the heuristics that have been identified are coordinated and weighted. One heuristic is the NVN strategy, which says that the processor assumes the subject is an proto-agent and the object is a proto-patient/theme. The other is a plausibility strategy which states that the processor assumes the semantic analysis that is most consistent with world knowledge. A secondary issue related to plausibility is whether the heuristic system distinguishes plausibility and animacy—that is, are sentences like *the man was bitten by the dog* and *the cheese was eaten by the mouse* qualitatively different, or is there a single continuum of plausibility, with sentences involving anomalous events such as the cheese ate the mouse simply falling further to one end of that continuum?

To answer these questions, the heuristics and their different possible weightings were translated into five different simple models, and corresponding predictions were generated. These predictions were then correlated with the actual data. The conditions for which predictions were obtained were: (1) active biased reversible, plausible, (2) active biased reversible, implausible, (3) passive biased reversible, plausible, (4) passive biased reversible, implausible, (5) active nonreversible, plausible, (6) active nonreversible, implausible, (7) passive nonreversible, plausible, (8) passive nonreversible, implausible, (9) active symmetrical, and (10) passive symmetrical. (For Experiment 3, the labels "subject-cleft" and object-cleft" should be substituted for active and passive, respectively.) Thus, each correlation was based on 10 values. Rank-order correlations were then performed separately for each experiment; and within an experiment, correlations were computed separately for accuracy and for RTs, and separately for agent and for patient decisions. Therefore, for every model 12 rank-order correlations were computed, as shown in Table 6 (three experiments $\times$ accuracy data versus RTs $\times$ agent versus patient decisions). The mean correlations across the models were then compared.

The first model is labeled in the table as "NVN alone." The only heuristic that is used is the NVN strategy. This model would predict that all the active sentences

Table 6
Comparing correlations between experimental data and five simple models incorporating various heuristics

Experiment	Measure	Models				
		NVN alone	Only plaus and animacy conflated	Only plaus but animacy separate	NVN plus plaus with animacy conflated	NVN plus plausibility with animacy separate
1	ag accuracy	0.870	0.400	0.400	0.960	0.950
	pat accuracy	0.700	0.490	0.410	0.850	0.810
	ag RTs	0.730	0.490	0.520	0.880	0.890
	pat RTs	0.450	0.740	0.670	0.750	0.720
2	ag accuracy	0.630	0.390	0.470	0.740	0.780
	pat accuracy	0.770	0.330	0.260	0.840	0.800
	ag RTs	0.870	0.350	0.320	0.930	0.920
	pat RTs	0.730	0.450	0.440	0.860	0.850
3	ag accuracy	0.870	0.350	0.320	0.930	0.920
	pat accuracy	0.450	0.610	0.610	0.690	0.700
	ag RTs	0.800	0.300	0.270	0.850	0.830
	pat RTs	0.590	0.450	0.390	0.740	0.710
	Mean	0.705	0.446	0.423	0.835	0.823

should be equally difficult, all the passives should be equally difficult, and the passives overall should be harder. The correlations between this predicted ordering and the ordering obtained for the different measures in the three experiments range from .45 to .87, and the average is .71. Next, consider another equally simple model, one that uses just a single heuristic (the second column labeled "Only plausibility and animacy conflated" in Table 6). This model uses only plausibility to assign semantic roles, and animacy and plausibility are not distinguished. Because the NVN strategy is not used, the active and passive sentences should be equally easy to process. The easiest sentences should be the plausible ones (regardless of whether the sentence is reversible or nonreversible), the next hardest should be the symmetrical sentences, and the hardest should be the implausible sentences (again, regardless of reversibility). The correlations for this second model range from .30 to .74, and the average of all 12 correlations is .45. Clearly, this model does worse than the one that relies on NVN alone, as revealed by a t test comparing the two means (.71 vs. .45, $t(22) = 4.59$, $p < .01$).

The next model to evaluate is exactly the same as the second, but animacy and plausibility are distinguished. Again, because the NVN heuristic is not used, performance with all the actives should be the same as for the passives. The easiest sentences should be the plausible nonreversibles, then the plausible reversibles, then the implausible nonreversibles, followed by the implausible reversibles, and finally the symmetrical sentences (again, in all cases regardless of syntactic form). This model actually does slightly worse than the one that uses only the heuristic of plausibility and does not distinguish plausibility from animacy. The correlations for this third model range from .27 to .67, and the average of all 12 correlations is .42. This value of .42 does not differ from the value of .45 that was obtained for the second model, $t < 1$.

So far, it appears that a model that uses the NVN heuristic alone does better than one that considers only plausibility, and plausibility and animacy do not need to be distinguished. What if the syntactic and semantic heuristics were combined, with NVN weighted more heavily (i.e., with all passive sentences predicted to be more difficult than all the actives)? The correlations between predicted and obtained orderings of conditions for the combined models are shown in the fourth and fifth columns of Table 6 (labeled "NVN plus plausibility with animacy conflated" and "NVN plus plausibility with animacy separate"). The former model predicts that the easiest conditions should be the active plausible reversible condition and the active plausible nonreversible condition, followed by the implausible active nonreversible condition and the implausible active reversible condition, and the symmetricals should be the most difficult of all the actives. All the passives are predicted to be harder than all the actives, and the semantic conditions should pattern in the same way as for the actives. This fourth model yields correlations between .69 and .96, and the average correlation is .84. This overall correlation is significantly better than for the model using the NVN strategy alone (.84 vs. .71, $t(22) = 2.60$, $p < .02$). Finally, the fifth model uses the NVN heuristic as well as plausibility, and plausibility is distinguished from animacy. The correlations for this model range from .70 to .95, and the average of all 12 correlations is .82. This more complex model is no better than the one that is identical but does not distinguish animacy from plausibility (.82 vs. .84, $t(22) = 1.71$, $p > .70$).

These analyses show that a simple processing model does a remarkably good job of accounting for a large amount of data from three experiments. The model captures both the decision and reaction time data, for both agent and patient/theme decisions, and for actives, passives, and clefts. It is important to make explicit how simple this model really is. It assumes the use of just two fast and frugal heuristics: the NVN strategy, and a semantic association heuristic. The latter does not even need to distinguish categorically between animate and inanimate concepts. The difference between this approach and the one adopted in constraint-based models (which also emphasize probabilistic cues) is that the probabilistic heuristics described here are not used only to modulate activation levels for syntactically generated candidates; instead, the NVN and semantic association heuristics are used to directly generate hypotheses about the meanings of sentences. It appears, then, that for language processing as in other cognitive domains, people make use of a set of fast and frugal heuristics (Gigerenzer, 2000; Gigerenzer et al., 1999). When they engage in the sequential decision making task of comprehension, they sometimes generate a shallow representation of the input. Of course, it is important to acknowledge that the comprehension system uses syntactic algorithms as well. Indeed, given that overall people almost always did better than chance on the implausible passive and object-cleft sentences, it is clear that real syntactic knowledge is consulted and used when people comprehend sentences. At the same time, the errors that comprehenders do make reveal the use of simple heuristics as well.

7. Broader implications

In recent work, we (Christianson et al., 2001; Ferreira et al., 2002; Ferreira, Christianson, & Hollingworth, 2001; Ferreira & Henderson, 1999) have argued for what we term "good enough representations" in language processing (and other cognitive domains). The fundamental idea is to challenge the assumption that the language comprehension system always builds rich and complete representations for the utterances it encounters. The phenomena described in Section 1 are difficult to reconcile with the completeness assumption: the Moses illusion (Erickson & Mattson, 1981; Kamas et al., 1996), the earlier Fillenbaum (1971, 1974) work on normalization, and the Duffy et al. (1989) study on sentence-level facilitation of naming processes all reveal that people create incomplete and distorted representations. Recent work in visual cognition demonstrating so-called "change blindness" also undermines the completeness assumption (in another cognitive domain): Researchers have demonstrated that viewers of real-world visual scenes can be surprisingly insensitive to changes made to those scenes while they are being viewed (Henderson & Hollingworth, 1999; Simons & Levin, 1997). Thus, we argue that input systems are not designed to consistently deliver full and complete representations of the stimuli they encounter.

Christianson et al. (2001) demonstrated this point with garden-path sentences. They had participants read sentences such as *While Anna bathed the baby played in the crib*, and the participants' task was to answer comprehension questions after each one. Christianson et al. found that accuracy was high for a question such as *Did the baby play in the crib*, indicating that the noun phrase *the baby* had been successfully restructured as the subject of the second clause (from its initial position as object of the first). If Christianson et al. had stopped there, they might have concluded that people are able to understand these difficult sentences fully and completely. But by also asking questions such as *Did Anna bathe the baby*, to which participants replied "yes" the majority of the time, they were able to demonstrate that the ultimate semantic representation for the sentence was not complete and accurate. In addition, participants had a great deal of confidence in both their correct and incorrect answers. Furthermore, participants did not answer "yes" as often to the question about the subordinate clause when the clauses were inverted or when a comma separated the two clauses, because both of those modifications eliminated the garden-path (i.e., the ultimately incorrect syntactic structure that supported the misinterpretation). Christianson et al. concluded that the representations for sentences are "good enough" rather than faithful renditions of the actual information in the sentence.

Another study demonstrating the "good enough" nature of language comprehension was conducted recently by Brysbaert and Mitchell (2001). They used Dutch sentences to examine whether Dutch-speaking participants would use disambiguating gender cues to attach relative clauses in sentences. For example, even in English, the sentence *Mary liked the son of the actresses who are on the balconies* is unambiguous, because the form of *to be* makes clear that the relative clause attaches to actresses, not *son*. In a questionnaire study in which participants read sentences such as these at their own pace and then indicated where the relative clause attached (e.g., who was on the balcony), Brysbaert and Mitchell found that people were surprisingly insensitive to the morphological cues. Chance performance would have been 50%, and perfect performance (using all the cues consistently) would have been 100%. Their participants were accurate 79% of the time, suggesting that they sometimes used the information but they sometimes instead went with whatever interpretation seemed more semantically sensible to them. This result is important for a number of reasons, and one of them is methodological: Morphologically disambiguated sentences are often used as control conditions in psycholinguistic experiments, and these results suggest that this particular practice needs to be reevaluated. More importantly for current purposes, the results again suggest that the language comprehension system delivers representations that are "good enough" rather than rich and complete.

Finally, Gibson and Thomas (1999) studied readers' comprehension of center-embedded structures such as *The ancient manuscript that the graduate student who the new card catalog had confused a great deal was studying in the library was missing a page*. Their goal was to determine under what circumstances comprehenders overlook that a center-embedded sentence is missing a syntactically obligatory verb phrase. An example of such an ungrammatical sentence was *The ancient manuscript that the graduate student who the new card catalog had confused a great deal was studying in the library*. Gibson and Thomas found that these two items were rated as equivalently difficult to understand, even though the one missing a verb phrase should be rejected as uninterpretable because one of the noun phrase constituents is missing its required companion-its verb phrase. This study again illustrates that, particularly when a sentence becomes extremely difficult to parse and interpret, comprehenders adopt a good-enough strategy to trying to understand it. As a result, they even tolerate outright ungrammaticality. (Note that this analysis of the results is not the same as the one offered by Gibson and Thomas, although it is not incompatible with it.)

But what does "good enough" mean in the context of language understanding? It is important to note that all the results that have been mentioned or described thus far demonstrating less than perfect comprehension have been obtained under conditions that would seem to maximize the chances that people would obtain the right interpretations. In the language of LAST, the experiments would seem to give every opportunity for the second-stage, algorithmic parse to run to completion and to win out over the pseudo-parse. In psycholinguistic experiments such as the current set, sentences are presented without background noise, the input is perceptually clear, often the participant controls the pace at which the sentence is presented (in reading studies, not in the current experiments), and most critically, the goal of the participant is specifically to understand the sentence so as to perform well on a relatively public comprehension task. Contrast this situation with a more ecologically valid one such as two people having a conversation in a noisy restaurant, with music and other conversations competing with and occluding the input. The person's goal is probably not to construct faithful representations of her interlocutor's utterances, but rather to get enough out of them so that he or she can take her turn at appropriate times and keep the social exchange functioning properly. In other words, in a laboratory situation, the criterion for what will be considered "good enough" will likely be set quite high; during a casual conversation in a restaurant or bar, it might be set much lower. It appears, therefore, that even when a situation and a task put a premium on accurate comprehension, people often do not achieve it. Obviously, the system does not always need to establish even this level of accuracy, because some communications are conducted not for the purposes of exchanging information but instead are opportunities for social "grooming" (Dunbar, 1996).

Moreover, comprehenders likely take into account the fact that speakers often express themselves incorrectly (generating full word exchange errors, for instance; Garrett, 1975) and that their own memory systems are fallible, so that what they think they have heard does not always correspond to what was actually produced. In addition, listeners often fail to notice whole word exchange errors when speakers make them. This tendency suggests that not only are people often forgiving, but also that the mechanisms of the comprehension system itself are designed so that material can make it through even when it does not actually match the likely semantic intention of the speaker or the meaning constructed by the listener. Consider these examples of whole word exchanges (taken from the UCLA corpus):

(a) A small body of instruments written for these compositions
(b) I wonder when they're going to get the channel built across the tunnel.
(c) ...whether we're going to get a wife for his job
(d) When the paper hits the story
(e) ...used the door to open the key
(f) The theory raised in the specific hypothesis...

Notice how difficult it is to tell that an error has been made; the impression one gets is of having to focus intensely on the examples to detect the mistake. These cases are similar to the implausible and anomalous sentences tested in the current experiments (e.g., *the dog was bitten by the man*). Perhaps, then, it is not surprising that people misunderstand sentences of the sort tested in this study given that they seem to overlook word exchange errors in real spoken language. It is intriguing to speculate that the same mechanism might explain both effects.

The most important conclusion to draw from this study is that the language comprehension system uses a mixture of heuristics and syntactic algorithms. One of the most powerful heuristics is the NVN strategy, which was defined earlier as a tendency to assume that the first argument in a sentence is a proto-agent and the second is a proto-patient. Further investigations should be directed towards refining this heuristic further as well as the one that makes use of plausibility information. In addition, critical questions remain unanswered about the properties of the two systems. It is not yet known whether heuristics and algorithms are applied in parallel, or if one system is used only when the other fails. In addition, it will be important to determine how the outputs of the two systems are coordinated. The present study cannot answer these questions, but by demonstrating the importance of heuristics it sets up a new and possibly fruitful line of investigation for psycholinguistics, and one that connects the field to important trends in other areas of cognitive science.

Acknowledgments

This research was supported by the National Science Foundation (awards SBR-9319272 and BCS-9976584) and by a grant from Michigan State University. I thank Doug Davidson, Leah English, Mel Johnson, Robin Pulford, Janis Stacey, Ben Swets, Katie Tomak, and Angela Unitis for assistance with all aspects of data collection and analysis, Frederick L. Oswald for assistance with the correlational analyses, and John M. Henderson for comments on the manuscript. I also thank Simon Liversedge and the other three reviewers for their extremely helpful feedback.

Appendix A. Sentences used in the experiments

The sentences are presented below only in the active, plausible version. The implausible version can be created by reversing the nouns (for the symmetrical sentences, the two versions are arbitrary). The passive forms are created as follows: e.g., *the dog bit the man* → *the man was bitten by the dog*. The active-clefts are created as follows: e.g., *the dog bit the man* → *it was the dog that bit the man*.

Set 1: Biased, reversible sentences

1. The dog bit the man.
2. The cook ruined the food.
3. The bird ate the worm.
4. The cat chased the mouse.
5. The soldier protected the villager.
6. The lawyer sued the doctor.
7. The teacher quizzed the student.
8. The cop pursued the thief.
9. The waitress served the man.
10. The owner fed the cat.
11. The detective investigated the suspect.
12. The doctor treated the patient.
13. The politician deceived the voter.
14. The hiker killed the mosquito.
15. The horse threw the rider.
16. The golfer hit the ball.
17. The hunter shot the deer.
18. The frog ate the fly.
19. The ghost scared the boy.
20. The horse kicked the jockey.
21. The angler caught the fish.

22. The matador dodged the bull.
23. The officer arrested the citizen.
24. The prince slayed the dragon.

Set 2: Nonreversible sentences

1. The chef wore the apron.
2. The farmer planted the corn.
3. The mouse ate the cheese.
4. The dog buried the bone.
5. The editor reviewed the paper.
6. The comic told the joke.
7. The plumber fixed the drain.
8. The runner won the race.
9. The cow chewed the cud.
10. The DJ played the music.
11. The secretary typed the letter.
12. The artist painted the picture.
13. The termite chewed the wood.
14. The chicken laid the egg.
15. The ant built the hill.
16. The doctor took the X-ray.
17. The pirate buried the treasure.
18. The nurse gave the shot.
19. The dentist pulled the tooth.
20. The beaver gnawed the tree.
21. The child pulled the wagon.
22. The bulldozer pushed the dirt.
23. The tailor hemmed the skirt.
24. The pilot flew the plane.

Set 3: Symmetrical sentences

1. The boy kicked the girl.
2. The girlfriend kissed the boyfriend.
3. The sister hugged the brother.
4. The committee introduced the chairman.
5. The runner saw the driver.
6. The woman called the girl.
7. The man visited the woman.
8. The boy touched the man.
9. The bird heard the lady.
10. The producer recognized the director.
11. The priest approached the rabbi.
12. The child loved the puppy.
13. The butcher despised the baker.
14. The team chose the player.
15. The clerk thanked the customer.

16. The teacher greeted the parent.
17. The mother adored the son.
18. The realtor faxed the buyer.
19. The catcher signaled the pitcher.
20. The broker phoned the client.
21. The guest insulted the host.
22. The model met the photographer.
23. The prime minister embraced the pope.
24. The witch praised the wizard.

References

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*, 703–719.

Ball, C. N. (1994). The origins of the informative-presupposition It-Cleft. *Journal of Pragmatics, 22*, 603–628.

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory Cognition, 21*, 477–487.

Bates, E., Devescovi, A., & D'Amico, S. (1999). Processing complex sentences: A cross-linguistic study. *Language and Cognitive Processes, 14*, 69–123.

Bates, E., McNew, S., MacWhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence processing: A cross-linguistic study. *Cognition, 11*, 245–299.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York: Wiley.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.

Boland, J. E., & Boehm-Jernigan, H. (1998). Lexical constraints and prepositional phrase attachment. *Journal of Memory and Language, 39*, 684–719.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language, 33*, 251–283.

Brysbaert, M., & Mitchell, D. C. (2001). The failure to use gender information in parsing: A comment on van Berkum, Brown, and Haggort, 1999. Unpublished manuscript.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*, 368–407.

Crocker, M. W., & Brants, T. (2000). Wide-coverage probabilistic sentence processing. *Journal of Psycholinguistic Research, 29*, 647–669.

Cupples, L. (2002). The structural characteristics and on-line comprehension of experiencer-verb sentences. *Language and Cognitive Processes, 17*, 125–162.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition, 7*, 49–59.

Delahunty, G. P. (1984). The analysis of English cleft sentences. *Linguistic Analysis, 13*, 63–113.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20*, 611–629.

Dowty, D. (1991). Thematic proto-rules and argument selection. *Language, 67*, 547–619.

Duffy, S. A., Henderson, J. M., & Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology: Learning, Memory, Cognition, 15*, 791–801.

Dunbar, R. I. M. (1996). *Grooming, gossip, and the evolution of language*. Cambridge, MA: Harvard University Press.

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20*, 540–551.

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language, 33*, 715–736.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11–15.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research, 30*, 3–20.

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

Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 555–568.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 30*, 725–745.

Ferreira, F., & Henderson, J. M. (1999). Good enough representations in visual cognition and language. Paper presented at architectures and mechanisms of language processing conference, Edinburgh, Scotland.

Fillenbaum, S. (1971). Processing and recall of compatible and incompatible question and answer pairs. *Language and Speech, 14*, 256–265.

Fillenbaum, S. (1974). Pragmatic normalization: Further results for some conjunctive and disjunctive sentences. *Journal of Experimental Psychology, 102*, 574–578.

Fillmore, C. J. (1968). The case for case. In E. Bach, & R. T. Harms (Eds.), *Universals of linguistic theory* (pp. 1–88). New: Holt, Rinehart, and Winston.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210.

Garnham, A., & Oakhill, J. (1987). Interpreting elliptical verb phrases. *Quarterly Journal of Experimental Psychology, 39A*, 611–625.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation: vol. 9* (pp. 133–177). New York: Academic Press.

Garrett, M. F. (2000). Remarks on the architecture of language processing systems. In Y. Grodzinsky, & L. P. Shapiro (Eds.), *Language and the brain: Representation and processing. Foundations of neuropsychology series* (pp. 31–69). San Diego, CA: Academic Press.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes, 14*, 225–248.

Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world*. New York: Oxford University Press.

Gigerenzer, G., Todd, P.M., & ABC Research Group (1999). Simple heuristics that make us smart. New York: Oxford University Press.

Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.

Harley, T. (2001). *The psychology of language: From data to theory*. Hove, UK: Psychology Press.

Henderson, J. M., Falk, R., Minut, S., Dyer, F. C., & Mahadevan, S. (2001). Gaze control for face learning and recognition by humans and machines. In T. F. Shipley, & P. J. Kellman (Eds.), *From fragments to objects: Segmentation and grouping in vision* (pp. 463–481). Amsterdam: Elsevier.

Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science, 10*, 438–443.

Hoeks, J. C., Vonk, W., & Schriefers, H. (2002). Processing coordinated structures in context: The effect of topic-structure on ambiguity resolution. *Journal of Memory and Language, 46*, 99–119.

Hornby, P. A. (1974). Surface structure and presupposition. *Journal of Verbal Learning and Verbal Behavior, 13*, 530–538.

Houston, A., & McNamara, J. (1999). *Models of adaptive behaviour*. Cambridge, MA: Cambridge University Press.
Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*, 441–480.
Just, M. A., & Clark, H. H. (1973). Drawing inferences from the presuppositions and implications of affirmative and negative sentences. *Journal of Verbal Learning and Verbal Behavior, 12*, 21–31.
Kaelbling, L., Littman, M., & Cassandra, T. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence, 101*, 99–134.
Kamas, E. N., Reder, L. M., & Ayers, M. S. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. *Memory Cognition, 24*, 687–699.
King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580–602.
Lehnert, W., & Ringle, M. H. (1982). *Strategies for natural language processing*. Hillsdale, NJ: Erlbaum.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 9*, 157–201.
MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.
MacWhinney, B., & Bates, E. (Eds.). (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.
MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Memory and Language, 23*, 127–150.
Mahadevan, S., & Cornell, J. (1992). Automatic programming of behavior-based robots using reinforcement learning. *Artificial Intelligence*, 311–365.
Mahadevan, S., Theocharous, G., & Khaleeli, N. (1998). Rapid concept learning for mobile robots. *Autonomous robots, 5*, 239–251.
Minut, S. (2000). Selective visual attention in a search task: A reinforcement learning model. Unpublished master's thesis, Michigan State University.
Minut, S., & Mahadevan, S. (2001). A reinforcement learning model of selective visual attention. Paper presented at the 5th international conference on autonomous agents, Montreal, Canada.
Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 92–103.
Nelson, G. (1997). Cleft constructions in spoken and written English. *Journal of English Linguistics, 25*, 340–348.
Oakhill, J., & Garnham, A. (1996). The mental models theory of language comprehension. In B. Britton, & A. Graesser (Eds.), *Models of understanding text* (pp. 313–339). Hillsdale, NJ: Erlbaum.
Oakhill, J., Garnham, A., & Vonk, W. (1989). The on-line construction of discourse models. *Language and Cognitive Processes, 4*, SI263–SI286.
Osgood, C. E. (1980). *Lectures on language performance*. New York: Springer.
Puterman, M. (1994). Markov decision processes: Discrete stochastic dynamic programming. Wiley, New York.
Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
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.
Rochemont, M. (1986). *Focus in generative grammar*. Amsterdam: John Benjamins.
Rooth, M. (1992). A theory of focus interpretation. *Natural language semantics, 1*, 75–116.
Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33–58). Hillsdale, NJ: Erlbaum.
Sanford, A. J. (1999). Word meaning and discourse processing: A tutorial review. In S. Garrod, & M. Pickering (Eds.), *Language processing* (pp. 301–334). East Sussex, UK: Psychology Press.
Sanford, A. J. (2002). Context, attention, and depth of processing during interpretation. *Mind Language, 17*, 188–206.

Schank, R. C. (1972). Conceptual dependency: a theory of natural language understanding. *Cognitive Psychology, 3*, 552–631.

Schneider, D., & Philips, C. (2001). Grammatical search and reanalysis. *Journal of Memory and Language, 45*, 308–336.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Sheldon, A. (1974). On the role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior, 13*, 272–281.

Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review, 63*, 129–138.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Science, 1*, 261–267.

Slobin, D. I. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior, 5*, 219–227.

Speer, S. R., & Clifton, C. (1998). Plausibility and argument structure in sentence comprehension. *Memory Cognition, 26*, 965–978.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature, 2*, 191–196.

Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Traxler, M. J., Pickering, M. J., & Clifton, C., Jr. (1998). Adjunct attachment is not a form of ambiguity resolution. *Journal of Memory and Language, 35*, 454–475.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 528–553.

Tyler, L. K., & Warren, P. (1987). Local and global structure in spoken language comprehension. *Journal of Memory and Language, 26*, 638–657.

Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology, 11*, 92–107.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. London: B.T. Batsford.