Retrieval cues and syntactic ambiguity resolution

Retrieval cues and syntactic ambiguity resolution: speed-accuracy tradeoff evidence

Andrea E. Martin & Brian McElree

Submit your article to this journal ⬈

Article views: 237

View related articles ⬈

View Crossmark data ⬈

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Retrieval cues and syntactic ambiguity resolution: speed-accuracy tradeoff evidence

Andrea E. Martin [a,b,c] and Brian McElree[c]

[a]Department of Psychology, School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh, UK; [b]Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; [c]Cognition and Perception Program, Department of Psychology, New York University, New York, NY, USA

ABSTRACT

Language comprehension involves coping with ambiguity and recovering from misanalysis. Syntactic ambiguity resolution is associated with increased reading times, a classic finding that has shaped theories of sentence processing. However, reaction times conflate the time it takes a process to complete with the quality of the behavior-related information available to the system. We therefore used the speed-accuracy tradeoff procedure (SAT) to derive orthogonal estimates of processing time and interpretation accuracy, and tested whether stronger retrieval cues (via semantic relatedness: neighed->horse vs. fell->horse) aid interpretation during recovery. On average, ambiguous sentences took 250ms longer (SAT rate) to interpret than unambiguous controls, demonstrating veridical differences in processing time. Retrieval cues more strongly related to the true subject always increased accuracy, regardless of ambiguity. These findings are consistent with a language processing architecture where cue-driven operations give rise to interpretation, and wherein diagnostic cues aid retrieval, regardless of parsing difficulty or structural uncertainty.

Introduction

The defining characteristic of human language is an unbounded combinatorial capacity to form hierarchical sentence structures and compute meaning, even the words and phrases in those structures are separated in time and space, and by other linguistic representations. This capacity minimally requires that words and phrases can be interpreted together, whether they are adjacent or separated. For example, in (1):

(1) The toddler screamed.

In order to form a representation the sentence's meaning, a linguistic dependency between the noun *toddler* and the verb *screamed* must be established. Forming this dependency in (1) might be relatively computationally straightforward because *the toddler* is adjacent to *screamed,* and there are no other words in the sentence. But the integration of *the toddler* as the subject of the verb *screamed* becomes more complex in structure like (2):

(1) The toddler that the mother loved _endlessly _screamed.

In (2), there is an intervening reduced relative clause, *the mother loved endlessly*, the processing of which may displace the noun *the toddler* from the comprehender's focus of attention. At the verb *screamed*, a representation of the noun phrase *the toddler* must be retrieved in order to compute who did what to whom. Whenever comprehension requires retrieval of a previously processed representation from memory, constraints on memory may be a determinant of whether language comprehension is successful or not (see McElree, 2006, for a review).

Nonadjacent dependency formation, as in (2) above, cuts the computational problem of online language comprehension at its joints: how is incoming linguistic input recognised, structured, and interpreted accordingly? How might processing be shaped, biased, or even inhibited or facilitated, by recently processed representations?

In a growing body of literature, language processing is cast as a form of skilled memory retrieval (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; McElree, Foraker, & Dyer, 2003) where performance depends upon cues at retrieval, and their links to existing representations in memory, or cue diagnosticity (Martin,

---

CONTACT Andrea E. Martin ✉ andrea.martin@mpi.nl

2016; Martin & McElree, 2011; Martin, Nieuwland, & Carreiras, 2012, 2014; Nairne, 2002).

Under this conception of language processing, a particularly important memory constraint on comprehension is whether a required representation can be immediately retrieved from memory, or whether it must be retrieved via a search. Studies testing memory operations during dependency formation strongly suggest that dependent constituents are retrieved from memory via a direct-access operation (e.g. Foraker & McElree, 2007; Martin & McElree, 2008, 2009, 2011; McElree, 2000; McElree et al., 2003). Direct access occurs because retrieval is cue-based – retrieval cues are assumed to be a subset of the information available at the retrieval site (i.e. one part of the dependency, or the resolution of the open linguistic dependency via argument role filling using the retrieved constituent). Retrieval cues make contact with memory representations that have overlapping content via a unitary process, without recourse to a sequence of searches through irrelevant memories (e.g. Clark & Gronlund, 1996; Dosher & McElree, 2003; Kohonen, 1984). The strength to which the cues uniquely match the required target in memory determines whether retrieval is successful, and, consequently, whether the correct interpretation of a sentence is formed in comprehension. When more representations must be processed within or during the span of the two dependent constituents, the probability that retrieval is successful decreases (e.g. Foraker & McElree, 2007; Martin & McElree, 2008, 2009, 2011; Martin et al., 2012, 2014; McElree, 2000; McElree et al., 2003; Van Dyke & McElree, 2006, 2011). This is because the processing of additional representations increases retrieval interference, as the retrieval cues to a given item become less diagnostic or specific to a unique target in memory. If the interpolated representations are similar to the target in terms of their features, or any other representational aspect that may function as a retrieval cue, this state of affairs can result in *cue overload,* where the cues are insufficient for successful retrieval of a the target item (Nairne, 2002; Watkins & Watkins, 1975).

But the computational challenges of sentence comprehension do not always occur in perfect or uncertainty-free processing circumstances. In fact, part of the communicative power and representation parity of human language comes from the fact that it can be and often is ambiguous. Words and syntactic structures can often be at least temporarily ambiguous. There is a rich tradition in psycholinguistics to study the effects of ambiguity on sentence processing; the canonical finding being that sentences with ambiguous syntactic structure exhibit slower reaction times compared to sentences that are unambiguous in their structure. Competing theoretical accounts have offered a variety of

different explanations for slower response latency. Some of these accounts can be coarsely categorised into those claiming two sources of processing difficulty under ambiguity. The first category of account would claim that ambiguity is costly due to competing lexical or structural representations (so-called *representational difference* accounts, e.g. (1) Contextual accounts: Altmann, Garnham, & Dennis, 1992; Altmann, van Nice, Garnham, & Henstra, 1998; (2) Lexicalist or competition accounts: MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, Tanenhaus, & Garnsey, 1994; (3) A "race-based" account: van Gompel, Pickering, & Traxler, 2001; (4) A dynamical parsing account: Tabor, Galantucci, & Richardson, 2004, (5) A cue-based account: Van Dyke & Lewis, 2003). The second category seems to claim that difficulty arises due to the time it takes to reanalyse the ambiguous sentence structure (so-called *processing time difference* accounts, e.g (1). Minimal Attachment/ Late Closure: Frazier & Fodor, 1978; Frazier & Rayner, 1982; (2) Reanalysis and repair accounts: Ferreira & Henderson, 1991; Fodor & Ferreira, 1998; (3) A structural change vs. repair account: Sturt, 2007; Sturt, Pickering, & Crocker, 1999; Sturt, Scheepers, & Pickering, 2002; (4) A reprocessing account: Grodner, Gibson, Argaman, & Babyonyshev, 2003). However, the key evidence for all these theoretical explanations is slowed or delayed reaction times, usually from self-paced reading or eye-tracking paradigms.

Reaction time is a ubiquitous, useful and informative psychological measurement. But it is an inherent mixture of information about the internal representations and processes we are trying to measure (see Davidson & Martin, 2013; Wickelgren, 1977 for discussion). Participants (any agent, be it animal or device) can trade processing speed for processing accuracy, and can do so dynamically (Pachella & Pew, 1968; Schouten & Bekker, 1967; Wickelgren, 1977). Speed-accuracy tradeoff occurs because participants can prioritise a fast response over an accurate one simply by lowering their internal criterion, that is, by requiring the accumulation of less information (or lower representation strength or quality) before they reach criterion and make their response (see Macmillan & Creelman, 2005; Luce, 1986; for definitive guides on interpretation of response time). Participants can also do the opposite – they can prioritise the quality of the information on which they are basing their decision (i.e. their "decision" to move their eyes or press a button or speak) over the time it takes to make the response. Under these circumstances then, how can we discriminate between *representational difference* and *processing time difference accounts,* or even conclude that aspects of both accounts may be correct? From reaction time alone, it is entirely possible that any

and all of observed syntactic ambiguity resolution effects are due to differences in the quality or strength of the representations that are formed during sentence comprehension, or due to multiple parsing attempts (so-called *reanalysis*), or due to both. This last possibility, that ambiguity effects arise due to both representational differences *and* multiple attempts at parsing, is the prediction of a cue-driven account that is hybridised with tenants from both representational difference and processing time difference accounts.

If diagnostic cues determine successful retrieval, and in turn successful comprehension, then they should also play a role in the processing of syntactic ambiguity resolution in as much as that resolution depends upon forming the correct non-adjacent dependency. The processing of temporarily ambiguous structures has functioned as an important paradigm for testing hypotheses about the incremental parsing of incoming linguistic representations. Below is a classic example from Bever (1970):

(1) The horse raced past the barn fell.

Here, the verb *raced* is ambiguous between being the main verb of the clause *the horse raced past the barn* and being a past participle verb in a reduced relative clause *the horse raced past the barn*. Upon encountering the verb *fell*, the simpler or "default" main verb analysis (MV) of *raced* has adopted, it must be revised to the reduced relative analysis (RR). A large body of work has focused on the factors that affect the initial interpretation of structural ambiguity (e.g. Altmann & Steedman, 1988; Ferreira & Clifton, 1986; Frazier & Rayner, 1982; Lau & Ferreira, 2005; MacDonald et al., 1994; Sturt, 2007; Sturt et al., 2002; Trueswell et al., 1994). Less effort has been directed at examining factors that enable the recovery of a structural analysis in which the verb *fell* rather than *raced* serves as the main verb in a MV analysis. But it is an empirical question whether this recovery or reanalysis may be in part due to insufficient cues at retrieval. Specifically, is the cost of ambiguity in some sense due to whether the verb *fell* can provide sufficient cues to retrieve *the horse* as a potential subject?

Ambiguity, reanalysis, and cue diagnosticity

Given that a comprehender has interpreted *the horse raced past the barn* as a main clause in which *raced* is the MV, a reanalysis to the correct interpretation requires two things when encountering the verb *fell*: First, the comprehender must recognise that *raced* has a possible past participle analysis, which licenses a RR analysis of

raced past the barn, and, second, that *fell* can be bound to the subject *the horse* as the MV of the sentence. Recognising that *the horse* could be the subject for *fell* requires retrieving the subject when the verb is encountered. The success of that operation may depend on the diagnosticity of retrieval cues at *fell*, specifically, the ability to retrieve *the horse* rather than other potential noun phrases in memory.

Indeed, failure to recover from misanalysis of a syntactically ambiguous construction may be due to the lack of diagnosticity of the cues at the point where reanalysis is required. In the case of (3), *fell* may provide cues that resonate with *the barn* more than the distant noun phrase *the horse*, particularly given its recency to the verb. Reanalysis may be blocked by the persistence of a degenerate interpretation in which *fell* attaches to *the barn* as a modifier, equivalent to the syntactically well-formed sentence, *The horse raced past the barn that fell*.

We pursue the hypothesis that failure to recover from a garden path sometimes may be an extreme case of insufficient diagnosticity between the verb and the noun (see also Van Dyke & Lewis, 2003). As processing any subject-verb dependency, the cues available at retrieval may fail to elicit the dependent constituent some proportion of the time. Our focus is the contribution of diagnostic retrieval cues to recovery from misanalysis, both the probability that diagnostic cues can aid recovery and affect the timecourse of such a process via number of parsing attempts. While the notion of "reanalysis" stems from observed differences in reaction times (i.e. most commonly, differences in the amount of time a word or phrase is fixated during reading or in the latency of a button-press self-paced reading), as we outlined previously, such reaction time effects can also stem from differences in representation strength or quality, leading to a different internal criterion during the decision process to make the response (Davidson & Martin, 2013; Luce, 1986; Wickelgren, 1977). In this way, differences in representational aspects or states can engender differences that are sometimes interpreted as differences in processing timecourse. In order to compare the relative contributions of representational factors during syntactic ambiguity resolution, as touted by representational difference accounts, with processing time factors, as espoused by processing time difference accounts, we will employ the speed-accuracy tradeoff procedure (SAT). The SAT procedure models processing speed orthogonally from accuracy, thus enabling the detection of differences in representational quality while at the same time veridically estimating processing speed.

Though the nature of retrieval cues used during sentence comprehension is still an open question, some

evidence suggests that for subject-verb dependencies in English, the "semantic fit" of the verb to a subject affects retrieval success and parsing choices (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; McElree & Griffith, 1995; Tabor et al., 2004; Taraban & McClelland, 1988; Thornton & MacDonald, 2003; Trueswell et al., 1994; Van Dyke & McElree, 2011). For this reason, we ask whether the verb-semantics-based fit or match between a verb and its subject *versus* other nouns in the sentence, especially more recent nouns given the MV/RR ambiguity, affects the likelihood or incidence of reanalysis. For example, the verb *sparkle* is associated with the noun *jewelry*. Moreover, *to sparkle* is something that both actresses and jewelry can do. Consider the contrast between the sentences *The actress sent the jewelry sparkled* and *The actress sent the jewelry frowned* (see Table 1 and the Appendix). The semantic fit between the local noun *jewelry* and the verb *sparkled* is stronger than between *jewelry* and *frowned*. In order to quantify this difference in fit, semantic similarity values were calculated for the disambiguating verbs relative to each noun (the true subject and the local noun) using Latent Semantic Analysis (LSA, Landauer & Dumais, 1997; available on the Internet at http://lsa.colorado.edu). We attempted to parameterise the match between verb and target relative to local noun by selecting verb-noun combinations that yielded reliable mean differences between nouns by condition using the one-to-many comparison function- we entered each final verb and computed its relation to both nouns in the sentence and then took the difference between these values (see the Methods and Results sections and Table 2). Lastly, we included verbs with thematic requirements the recent noun cannot meet, such as *frowned* in the case of *jewelry*.

One possibility is that diagnostic cues decrease the likelihood that an ambiguous sentence must be reanalysed because on more trials, the correct representation is successfully elicited from memory. If this were the case, then more diagnostic cues might decrease the time it takes to resolve the dependency (NB: retrieval speed is assumed to be constant), as reanalysis occurs less frequently. A second possibility is that less diagnostic cues do not affect the timecourse or number of reanalysis attempts, but do affect the likelihood that the dependency will be successfully resolved at all (i.e. accuracy), therefore only affecting the success of interpretation, but not its timecourse. A third possibility is of course that cue diagnosticity affects both the time it takes to process the dependency *and* the likelihood that it will be successfully resolved. An analogue of these predictions could also come from the representational difference versus processing time difference categories of accounts, namely, since SAT will measure the degree to which both likelihood of interpretation and processing time are affected by ambiguity. On our interpretation of the literature, *Representational difference* accounts would claim that reaction time differences arise due to differences in the representations formed under ambiguity or from competition between structural analyses at the point of disambiguation. In SAT, such accounts would then predict that ambiguity should affect asymptotic accuracy, the estimate of the likelihood of successful retrieval and interpretation. Such accounts predict a difference in SAT asymptote parameters – that is, asymptotic differences should be observed with or without a dynamics parameters differences (i.e. estimates of processing speed). In contrast, so-called *processing time* accounts predict that ambiguous sentences should result in differences in the SAT function's dynamics parameters (intercept (δ) and rate (β)). However, neither category of account based on reaction time data can make definitive predictions about diagnosticity's effect either on the likelihood of successful interpretation or on SAT dynamics, because these are necessarily conflated in reaction time measures. Tabor et al. (2004) present a dynamical parsing account that arguably might predict that local match between noun and verb (the Local-match condition), a form of *local coherence*, would form a parse where *jewelry* is the subject of *sparkled*. It

Table 1. Example of materials.

Ambiguous, Local noun-related verb
1. The actress sent the jewelry sparkled
Ambiguous, Neutral verb
2. The actress sent the jewelry arrived
Ambiguous, Subject-related verb
3. The actress sent the jewelry frowned
Unambiguous, Local noun-related verb
4. The actress who was sent the jewelry sparkled
Unambiguous, Neutral verb
5. The actress who was sent the jewelry arrived
Unambiguous, Subject-related verb
6. The actress who was sent the jewelry frowned
Unacceptable Ambiguous
7. *The actress sent the jewelry clipped
Unacceptable Unambiguous
8. *The actress who was sent the jewelry clipped

Table 2. Mean LSA values (standard error) by condition and noun.

Condition	Subject noun	Local noun	Mean difference between Subject and Local noun	95% CI on Mean difference
Local noun-related verb	.11(.01)	.16(.02)	−.05	−.09 to −.01
Neutral verb	.16(.02)	.16(.02)	.004	−.04 to .05
Subject-related verb	.22(.04)	.10(.01)	.12	.05 to .20

is unclear whether such an interpretation would impact a participant's perception of acceptability or not, nor what the consequences for the timecourse of processing would be. Similarly, while cue-based accounts clearly predict that more diagnostic verbs should increase the likelihood of successful interpretation, they are not explicit about whether reanalysis effects in reaction time data stem from poorer representational quality or insufficient cue strength, or from additional parsing attempts at the point of disambiguation. It is an empirical question as to whether cue diagnosticity interacts with ambiguity resolution when disambiguation occurs. One possibility is that when there is uncertainty about syntactic structure or structural relationships, cue match between semantic features of nouns and verbs influences interpretation more such that diagnosticity effects are larger when ambiguity must be resolved. Another possibility is that ambiguity depresses any contribution of retrieval cues at the point of disambiguation (in this case, semantic features) because the syntactic cues are so uncertain or insufficient, a sort of syntactic bottleneck. In other words, the syntactic retrieval cues are not contributing to retrieval of the subject noun. Both of these outcomes would be expressed in an interaction between diagnosticity and ambiguity as factors.

To test these hypotheses, we employed the multiple response variant of SAT, where participants are probed for an acceptability judgment at multiple points in processing (Foraker & McElree, 2007; Martin & McElree, 2008, 2009, 2011; Van Dyke & McElree, 2011). The primary benefits of SAT are that speed and accuracy of processing are measured conjointly within a single task and that SAT forces participants to fully resolve ambiguity in order to make the judgment. SAT will allow us to discern whether longer reaction times found in the literature stem from true differences in timecourse or purely representational differences, or both. SAT accomplishes this by probing accuracy of interpretation at multiple times in relation to critical word (CW) onset, starting, crucially, just before (−300 msec) the CW onset. Then, a speed-accuracy function can be derived and fit to the points of empirical performance or accuracy at each lag latency (response tone time or lag + participant's response time to tone or latency). The crucial point is that the choice to respond is "taken away" from the participant, who must respond (within 200 msec) to the response tones, which have been set at lags that are determined by the experimenter. For each sampled time point (lag + latency), a d' measure of accuracy is constructed by scaling correct responses to grammatical expressions (hits) against incorrect responses to control expressions with ungrammatical interpretations (false alarms). This scaling provided a measure of the ability
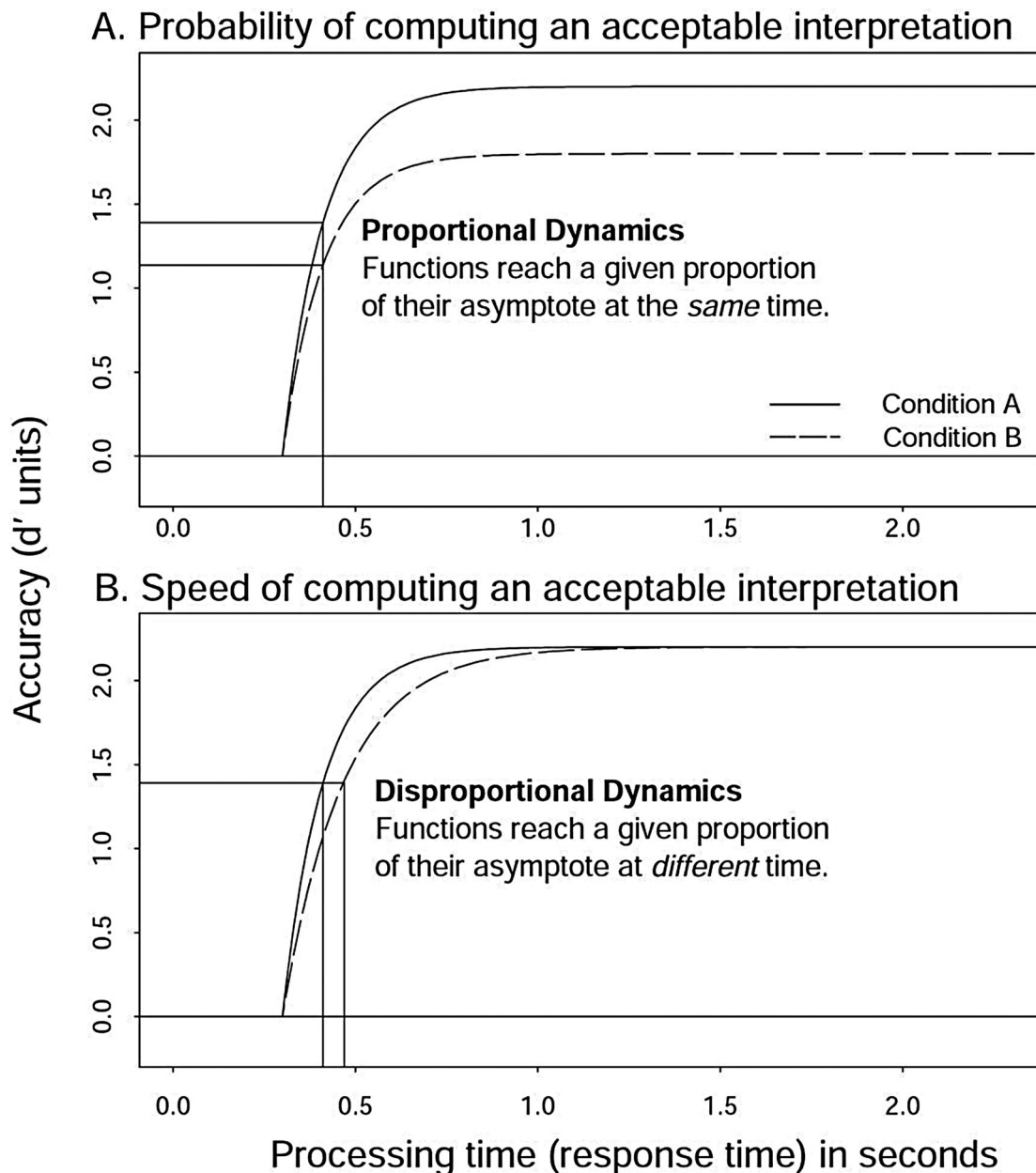
of participants to discriminate acceptable from unacceptable interpretations of temporarily ambiguous sentences. SAT functions typically show a period of chance performance ($d' = 0$), a period of increasing accuracy, and an asymptotic period where further processing does not improve performance (see Figure 1). Each participant's data were fit with an exponential function, quantifying how interpretation unfolded over time and allowing estimation of processing speed:

$$d' = \lambda(1 - e^{-\beta^{(t-\delta)}}) \text{ for } t > \delta; \text{ otherwise } t = 0. \quad (1)$$

The parameter λ, which estimates the asymptote of the function, measures the highest level of discrimination reached with maximal processing time, and hence yields a basic measure of processing accuracy. Here, asymptotes index how successful participants were at retrieving subject for the final verb, interpreting it, and converging upon a grammatical structural interpretation for the whole sentence. Increasing semantic fit with a recent noun may lower accuracy if it decreases diagnosticity between subject and verb in memory, making the subject less likely to be retrieved from memory.

The function's intercept (δ) and rate (β) provide joint measures of processing speed, indexing how quickly accuracy accrues to its asymptotic level. The parameter δ estimates the intercept, or the point where participants are first sensitive to the information necessary to discriminate acceptable from unacceptable (i.e. d' departs from 0, chance performance). The parameter β estimates the rate that accuracy grows from chance to asymptote. Decreasing how diagnostic a verb is to a unique subject in memory may slow the dynamics of the SAT function if, on some proportion of trials, the structure of the sentence must be revised or reanalysed more often, and diagnosticity in turn affects the number of reanalysis attempts needed to read an acceptable interpretation. Alternatively, ambiguity alone may slow the dynamics of the SAT function, as more parsing attempts result in increased processing time.

Because the speed-accuracy tradeoff procedure focuses on modelling individual speed-accuracy tradeoff functions and each participant's response bias via d', SAT studies require a high number of observations per participant, rather than a low number of observations across many participants as in more common in reaction time measurement studies (e.g. self-paced reading and eye-tracking). Each participant's data is analyzed separately, and the then parameter estimates across participants are evaluated in order to make inferences about falsification. The number of participants has ranged from 4 in single-response studies where tens of thousands trials per participant were collected (e.g. McElree

A. Probability of computing an acceptable interpretation

Proportional Dynamics
Functions reach a given proportion
of their asymptote at the *same* time.

Condition A
Condition B

B. Speed of computing an acceptable interpretation

Disproportional Dynamics
Functions reach a given proportion
of their asymptote at *different* time.

Processing time (response time) in seconds

Accuracy (d' units)

Figure 1. Hypothetical SAT functions illustrating two conditions differing in asymptote (Panel A) or rate (Panel B). The intersection of the horizontal and vertical lines shows the point in time (abscissa) when the functions reach two-thirds of their respective asymptote (ordinate). When dynamics are proportional (Panel A), the functions reach the two-thirds point at the same time.

& Dosher, 1989, 1993) to between 12 and 22 for multiple-response SAT where thousands of trials are collected per participant with multiple observations at different time points in a single trial (e.g. Foraker & McElree, 2007; Martin & McElree, 2008, 2009, 2011; McElree et al., 2003; Van Dyke & McElree, 2011).

Methods

Participants

Fifteen native speakers of American English (age range 18–26) from the New York University community were paid to participate in eight 60-minute sessions, and a 45-minute practice session.

Materials

Thirty-four sets of 8 sentences as illustrated in Table 1 featured sentences with the MV/RR ambiguity, unambiguous controls, and crossed with how "diagnostic" a verb is to its subject in relation to other nouns in memory, according to semantic similarity as reflected by LSA on the subject of the dependency. The Local noun-related verb conditions (see Table 1 for an example) contained a disambiguating verb that was

more similar to the local noun, e.g. *jewelry* than to the matrix subject noun, e.g. *actress*. This manipulation was meant to decrease the diagnostic match to the true subject *actress* compared to the local noun, and accomplished this by being both closer in the sentence and more similar to *jewelry*. The Neutral verb condition consisted of a verb that produced similar LSA values for both nouns (*actress* and *jewelry*), and that could combine with both nouns (see Table 2). A repeated measures analysis of variance (rmANOVA) on the LSA values for each noun by Condition showed an interaction between Condition and noun, F_2 (2, 66) = 12.39, $p < .001$). Pairwise comparisons revealed that within the Local noun-related verb condition, the LSA values for each noun given the verb differed reliably from each other such that the Local noun had a higher similarity value given the verb than the Subject noun did (see Table 2). Crucially, this pattern was reversed for the Subject-related verb condition, such that the Subject noun yielded a higher LSA value given the verb than the Local noun did, and in the Neutral verb condition, the nouns' LSA values did not differ from each other (see Table 2). The Subject-related verb condition used verbs that not only yielded higher LSA values with the true subject *actress*, but for which the recent noun *jewelry* would not be a plausible or grammatical subject. That is, actresses and jewelry can both sparkle and arrive, but only actresses can frown. For each triplet of Ambiguous and Unambiguous sentences, we created a matching unacceptable condition, (7) and (8) in Table 1, by making the sentence-final verb incompatible with the matrix subject, which created an unacceptable interpretation when the dependency is interpreted (e.g. *the actress clipped*). The unacceptable condition was designed to encourage participants to fully process the subject-verb dependency. To discriminate acceptable from unacceptable sentences, participants would have to process the dependency at least to the point where they had retrieved the subject and interpreted it with the local verb. Please see the Appendix for a full list of experimental stimuli and filler items.

Participants saw a total of 2576 sentence stimuli over 8 one-hour sessions, with 17 observations per trial for a total of 4,624 data points per participant. The sentence stimuli were as follows: There were 9 sub-experiments in the 9 sessions (8 experimental sessions with 322 stimuli, 34 from this study and 288 from other studies, plus a 45-minute practice session to familiarise the participants with the button-pressing part of the task). This resulted in ∼33% of the stimuli having a relative clause, where 16.4% was a full relative clause and 16.4% was reduced. There were a total of 1220 unacceptable stimuli, or 47.3% of the total stimuli. Of those violations,

41% stemmed from violation of the requirements of a verb in the sentence. Please see the Appendix for a more detailed description of the sub-experiments which composed the 10 session experiment. We note that our percentage of ∼33% relative clauses (RCs) is similar to or lower than the percentage often presented in the literature (e.g. Ferreira & Henderson, 1991 – 50% RCs; Grodner & Gibson, 2005– Exp 1: 35% RCs, Exp2: 50% RCs; Hemforth et al., 2015– 50% RCs; Traxler et al., 2002– Exps: ∼26.67% RCs).

Procedure

In eight experimental sessions, participants read 34 experimental sentences and 288 fillers for a total of 322 trials. There was one practice session to familiarise participants with the task. Participants saw every item in every condition, but at different points across the eight sessions. Participants completed no more than one session per day, on average completing all 9 sessions of the experiment in 10 days. Conditions were counterbalanced across sessions such that participants saw an equal number of items in each condition in each session, though the item used to represent that condition varied. Critical trials, including controls, constituted approximately 10% of each session, and were presented randomly among the remaining 90% fillers (please see Appendix). Sentences were presented phrase-by-phrase, 335 ms per word. Crucially, a 50 ms, 1000 Hz tone cued the first response 300 ms before the CW. This pre-CW response tone served to cue the participant to begin responding before they have perceived or fully processed the CW, in order to render their first response a guess and thus measure the full timecourse of processing. After CW onset, 16 more response signals occurred, 350 ms apart, while the CW remained on the screen, for a total of 17 response signals. In the practice session, participants were trained to synchronise their button presses to the response tone, responding maximally within 200 ms of each tone. They were instructed to simultaneously press both the "yes" and "no" keys as an initial (undecided) response, and then to select, and continue to press, only one of the two keys (with "yes" indicating a judgment that the sentence was acceptable, and "no" indicating that the sentence had been judged unacceptable) as soon as they decided whether they found the sentence acceptable or not.

Data analysis

Accuracy was calculated using a standard *d'* measure, *d'* = *z(hits)* – *z(false alarms)*, where a "hit" was an

"acceptable" response to an acceptable sentence and a "false alarm" was an "acceptable" response to an unacceptable sentence. The d' scores provide a measure of the participant's ability to discriminate acceptable from unacceptable, uncontaminated by response bias. We scaled the acceptable conditions against one unacceptable condition for Ambiguous, and another for Unambiguous.

A hierarchical model-testing scheme was used to determine whether conditions differed in asymptote (λ), rate (β), or intercept (δ) in Equation (1), constrained by reliable differences in empirical d'. Exponential model fits of the data ranged from a null model in which all functions were fit with a single asymptote, rate, and intercept parameter (a 1λ-1β-1δ fit) to a fully saturated (a 6λ-6β-6δ fit) model in which each condition was fit with a unique asymptote, rate, and intercept. For each participant and the averaged data, separate parameters were allotted to the different conditions if they systematically improved the fit of the SAT function to the observed d' data and the estimates were reliably different from one another over participants. Equation (1) was fit to the data with an iterative hill-climbing algorithm, which minimised the squared deviations of predicted values from observed data. Fit quality was assessed by an evaluation of the consistency of the parameter patterns across the individual participant fits. Model selection occurred via the performance inferential tests of significance computed over individual participants' d' data, used to constrain the model selected, and tests on the fitted parameter estimates for each of the candidate models, detailed in the Results section. In Table 4, we report 95% confidence intervals (CIs) around the mean difference for paired comparisons of interest. We then confirmed that the best fitting model from the above procedure also had the best adjusted-R^2 statistic – the proportion of variance accounted for by the fit, adjusted by the number of free parameters (Judd & McClelland, 1989).

Results

Figure 2 presents the average (across participants) d' values as a function of processing time, along with the best-fitting exponential model described below. Figure 3 presents the Ambiguous and Unambiguous conditions separately for visual illustrative clarity. Parameter values of the best-fitting models can be seen in Table 3. Inspection of Figures 2 and 3 illustrates that Ambiguous conditions were less accurately processed than Unambiguous ones, and that less diagnostic cues resulted in lower accuracy than more diagnostic cues.

Empirical d′ results

As an initial means of determining whether there were reliable differences in asymptotic performance as a function of Ambiguity and the relative relation between verb and nouns, or Diagnosticity, we averaged the d' values for each participant (and, for an item analysis, by each item) in each condition from 4.2 to 5 s post-initial response cue in order to derive an empirical estimate of asymptotic accuracy. An rmANOVA on these values for the conditions revealed a marginal main effect of Ambiguity, F_1 (1,14) = 2.96, p < .11, F_2 (1, 33) = 31.79, p < .001, and a reliable main effect of Diagnosticity, F_1 (2,28) = 23.4, p < .001, F_2 (2, 66) = 5.28, p < .01. There was a marginal interaction between Ambiguity and Diagnosticity, F_1 (2,28) = 2.96, p < .07, F_2 (2, 66) = .78, p < .47.[1]

Pairwise comparisons showed that all levels of the Diagnosticity factor differed reliably from each other for both Ambiguous and Unambiguous conditions (see Table 4). This pattern is consistent with the notion that how diagnostic a cue is shapes retrieval regardless of ambiguity; there is a subject-verb dependency that must be formed via retrieving the subject, in all conditions. Subject-related verbs resulted in higher accuracy than Neutral verbs, which in turn resulted in higher accuracy than Local noun-related verbs.

Model fits

Competitive fits of the exponential equation to the empirical data also yielded clear evidence that Ambiguous and Unambiguous conditions differed in processing speed, as well as in asymptotic accuracy: Models that did not allocate separate rate parameters asymptotes for the Ambiguous versus Unambiguous conditions produced poor fits to the empirical SAT data, and they left systematic residuals.

Models: null 1λ-1β-1δ to fully saturated asymptote model 6λ-1β-1δ

In fits of the average data, allocating separate asymptotes to each condition increased the adjusted-R^2 from .9722, observed with a null 1λ-1β-1δ model to .9917 with a 6λ-1β-1δ model. This model improved the quality of the fits of the individual participants' data, systematically increasing the adjusted-R^2 values over what was observed with a 1λ-1β-1δ model (ranging from .8316 to .9688 as compared to .7602 to .9557). In fits of the average data, allocating separate rates to Ambiguous and Unambiguous conditions
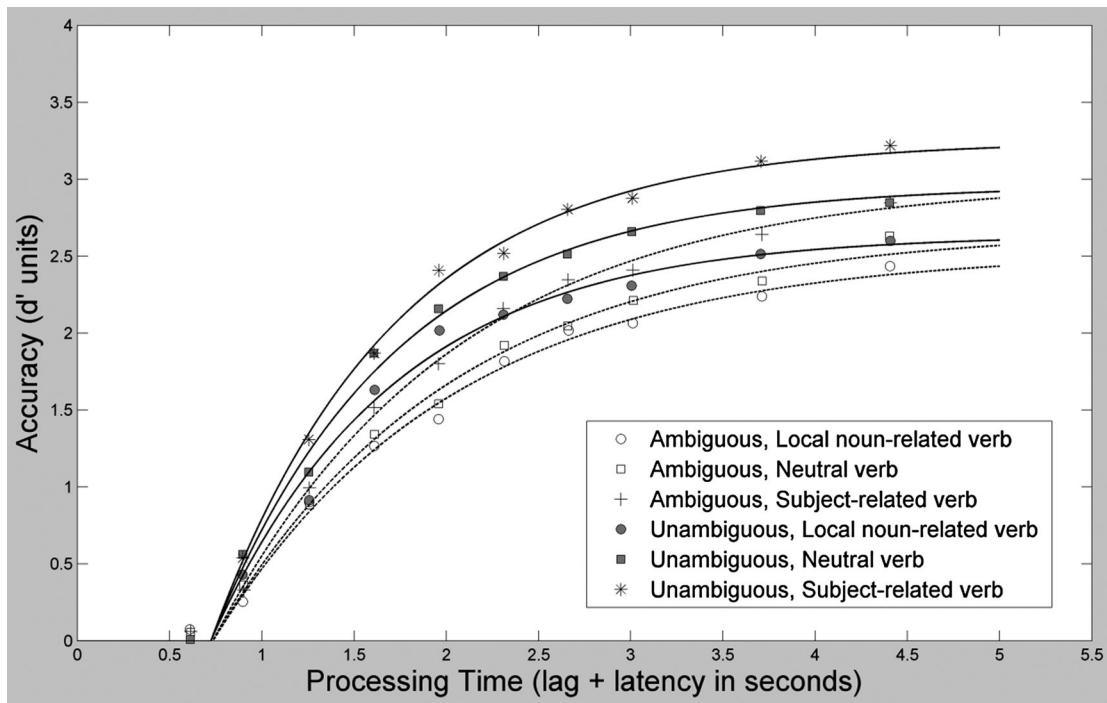
Figure 2. Average d' accuracy (symbols) as a function of processing time (lag of the interruption cue plus latency to response) for Ambiguous and Unambiguous conditions from Experiment 1. Smooth curves show the best-fitting exponential fit (see text).

increased the adjusted-R^2 from .9917, observed with a null 6λ-1β-1δ model to .9951 with a 6λ-2β-1δ model. This model improved the quality of the fits of the individual participants' data, systematically increasing the adjusted-R^2 values over what was observed with a 6λ-1β-1δ model (ranging from .667 to .961 as compared to .8316 to .9688).

### Separate model fits for ambiguous and unambiguous conditions

This pattern was also evident when the Ambiguous and Unambiguous conditions were fit separately, as evidenced by the estimates from the two 3λ-1β-1δ models (see Table 3). Per the results of the pairwise comparisons of empirical d', we constructed a 4λ-2β-1δ model where Ambiguous and Unambiguous conditions are fit together, but with four asymptotes approximating the reliable differences in empirical d'. In fits of the average data, the adjusted-R^2 increased to .9939 as compared to .9710 for a 4λ-1β-1δ model. Over participants, the adjusted-R^2 ranged from .8357 to .9658, increased from .7801 to .9590 in a 4λ-1β-1δ model, showing a systematic improvement of the fit over participants. We performed paired t-test on the asymptote parameter estimates for the four levels of d' to confirm that the estimates differed reliably from each other over participants.

Model fits with variable dynamics parameters

We implemented model of the asymptotes to construct a strong test of the rate difference posited by the 6λ-2β-1δ model. The rate parameters allotted to the Ambiguous and Unambiguous conditions remained reliably different from one another in both cases (paired t-tests on rate (β) estimates from 6λ-2β-1δ, $t = 2.65$, $p = .05$; 4λ-2β-1δ $t = 2.70$, $p = .05$; between separately fit by Ambiguity 3λ-1β-1δ models $t = 2.43$, $p = .05$; see Table 3). In sum, the models we evaluated all yielded better fits of the empirical data when separate rate parameters were allotted to Ambiguous and Unambiguous conditions (2β-1δ models). The most conservative estimate of the difference in processing speed observed is 250 msec from the 6λ-2β-1δ model (see Table 3 for a comparison of rate (β) over model fits), while the 4λ-2β-1δ model yielded an estimate of 455 msec. The 6λ-2β-1δ model is most conservative because it allows some of the variance to be captured by six different asymptotic parameters, rather than forcing that variance on the rate parameter as in 4λ-2β-1δ.

Discussion

There was a main effect of Diagnosticity on empirical d' such that lower diagnosticity resulted in lower likelihood of successful interpretation, regardless of Ambiguity. In the best fitting models (separately fit 3λ-1β-1δ's, 6λ-2β-
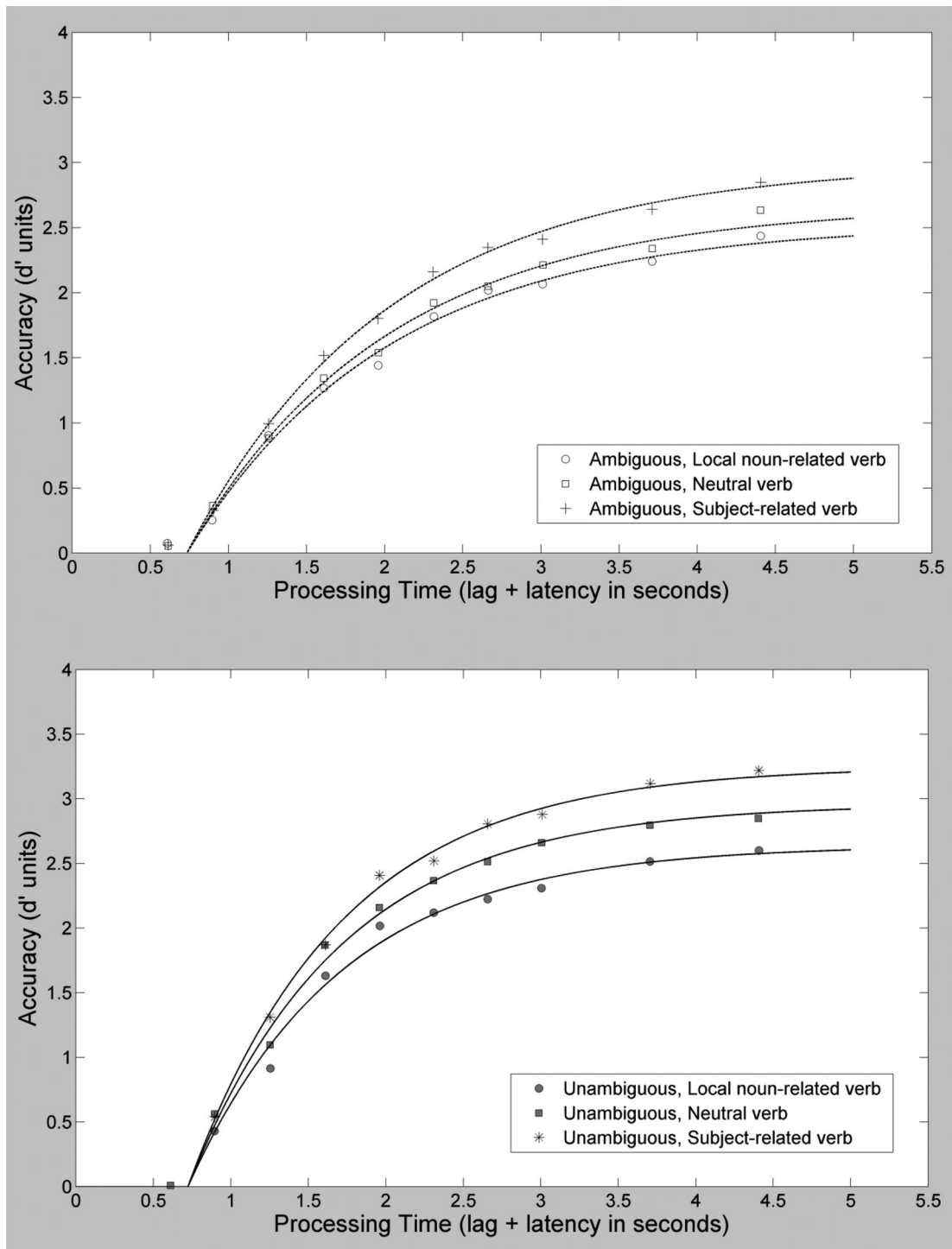
Figure 3. Average *d′* accuracy (symbols) as a function of processing time (lag of the interruption cue plus latency to response) for Ambiguous conditions (Top Panel) and Unambiguous conditions (Bottom Panel) from Experiment 1. Smooth curves show the best-fitting exponential fit (see text).

1δ, 4λ-2β-1δ), the differences in asymptote parameter estimates indicated that subjects of Local noun-related verbs and Neutral verbs were less likely to be successfully retrieved and interpreted than those of Subject-related verbs. Model fits revealed that Ambiguity affected the speed of processing the dependency, with difference in rate (β). Allocating separate rate or intercept parameters

to different conditions improved adjusted-R^2. Importantly, there were no consistent trends across participants intercept parameters when they were allowed to vary (e.g. models with dynamics parameters such as 1β-2δ, 2β-2δ, and various 1β-4δ and 2β-4δ models) and crucially t-tests on the intercept parameter (δ) estimates were not significant. Hence, there was evidence to

Table 3. Empirical d' and parameter estimates issuing from best-fitting models.

Condition	Average empirical d' across participants	6λ-2β-1δ average parameter estimates from fits of individual data	6λ-2β-1δ parameter estimates from fit of average empirical d'	4λ-2β-1δ average parameter estimates from fits of individual data	4λ-2β-1δ parameter estimates from fit of average empirical d'	3λ-1β-1δ average parameter estimates from fits of individual data, fit separately by Ambiguity	3λ-1β-1δ parameter estimates from fit of average empirical d', fit separately by Ambiguity
Ambiguous, Local noun-related verb	2.42	2.59	2.56	2.85 (λ_1)	2.63(λ_1)	2.80	2.53
Ambiguous, Neutral verb	2.63	2.72	2.70	3.09(λ_2)	2.87(λ_2)	2.95	2.67
Ambiguous, Subject-related verb	2.86	3.05	3.01	3.37(λ_3)	3.12(λ_3)	3.26	2.99
		Rate β: 1.0 β^{-1}: .998s	Rate β:.752 β^{-1}: 1.33s	Rate β:.749 β^{-1}: 1.34s	Rate β:.685 β^{-1}: 1.46s	Rate β:.831 β^{-1}: 1.20s Intercept δ: .855 Adjusted-R^2: .9322	Rate β: .771 β^{-1}: 1.30s Intercept δ: .734 Adjusted-R^2: .9953
Unambiguous, Local noun-related verb	2.57	2.77	2.64	3.09(λ_2)	2.87(λ_2)	2.79	2.64
Unambiguous, Neutral verb	2.85	3.01	2.95	3.37(λ_3)	3.12(λ_3)	3.08	2.96
Unambiguous, Subject-related verb	3.24	3.35	3.25	3.50(λ_4)	3.21(λ_4)	3.45	3.25
		Rate β:1.35 β^{-1}: .742s Intercept δ: .850 Adjusted-R^2: .9365	Rate β:1.01 β^{-1}: .990s Intercept δ: .726 Adjusted-R^2: .9951	Rate β:1.13 β^{-1}: .885s Intercept δ: .883 Adjusted-R^2: .9311	Rate β:1.05 β^{-1}: .956s Intercept δ: .721 Adjusted-R^2: .9939	Rate β:1.38 β^{-1}: .724s Intercept δ: .852 Adjusted-R^2: .9387	Rate β: 1.01 β^{-1}: .990s Intercept δ: .723 Adjusted-R^2: .9946
Diff. in estimates of Rate between Ambiguous and Unambiguous conditions in ms (β^{-1})		256	340	455	504	475	310

Table 4. Mean differences of interest with pairwise comparisons of empirical d'.

		Comparison	Mean difference (d' units)	95% confidence interval for difference
Ambiguity	Retrieval cue or diagnostic status			
A. Ambiguous	1. Local noun-related verb	1 vs. 2	−.21*	−.39 to −.02
	2. Neutral verb	1 vs. 3	−.43**	−.61 to −.25
	3. Subject-related verb	2 vs. 3	−.23*	−.44 to −.01
B. Unambiguous	4. Local noun-related verb	4 vs. 5	−.28*	−.54 to −.02
	5. Neutral verb	4 vs. 6	−.67**	−.87 to −.46
	6. Subject-related verb	5 vs. 6	−.39**	−.54 to −.23
		1 vs. 4	−.14	−.42 to .13
		2 vs. 5	−.22	−.55 to .12
		3 vs. 6	−.38*	−.76 to −.003

"*" indicates $p < .05$.
"**" indicates $p < $.

suggest that ambiguity affected processing speed and that this effect was expressed on the SAT rate parameter (β) or on the order of a 250–450 msec slowdown in processing time, but, there was no evidence that diagnostic cues affected the SAT function's dynamics parameters. In trends on empirical d', the marginal main effect of Ambiguity and the marginal interaction between Ambiguity and Diagnosticity suggested that ambiguity also tended to decrease likelihood of successful interpretation. We will not focus on interpreting these trends further as they are not reliable. We will also refrain from interpreting the lack of evidence for any interaction of variables.

General discussion

Diagnostic cues affected the likelihood that the dependent noun can be recovered from memory, regardless of ambiguity. Higher diagnosticity raised asymptotic accuracy, but did not affect the speed with which the dependent representation was accessed and interpreted. This suggests that lower diagnosticity results in lower likelihood of successful interpretation regardless of ambiguity, but there was no evidence to suggest that the number of reanalysis attempts or the time it takes to reanalyse is affected by diagnosticity. Although we do not interpret the lack of evidence of an effect as evidence of a lack of an effect, we note that a main effect of diagnosticity on accuracy, but not on processing speed, is incompatible with the strongest interpretation of *representational difference* accounts where representational factors (context, lexical features or statistics, parse frequency or probability) affect processing time. This finding is, however, consistent with the notion that representational differences partly underlie the differences observed in reaction time studies – namely, that cue strength or diagnosticity determines successful interpretation. This conclusion echoes those of Van Dyke and Lewis (2003), wherein cues initiate structural dependencies and any changes to them, and no

separate or additional repair or reanalysis mechanism needs to be postulated.

The timecourse profile is consistent with reaction time findings in that Ambiguous sentences took, as a conservative estimate, on average 250 msec longer to process then Unambiguous ones, with the less conservative estimate being 450 msec longer. By itself, this advantage for unambiguous sentences is not surprising – it confirms findings from a myriad of reaction time studies. However, that we find these effects in the rate of the SAT function does offer the first veridical timecourse evidence that reanalysis does indeed require additional processing time, and that the reaction time differences found in the literature do not simply result from differences in representation strength (Davidson & Martin, 2013; McElree, 2006; Wickelgren, 1977). This pattern of results is consistent with the predictions of *processing time difference* class of accounts, wherein longer latencies during syntactic ambiguity resolution are due to increased processing time (and not only due to representational factors). However, we do not interpret longer processing time as necessarily reflecting either repair to previously processed structures (so-called *repair-based reanalysis*) or a reanalysis situation with repair (e.g. additional retrieval attempts or starting processing over "from the beginning"). This is because the rate difference itself cannot distinguish between a repair account and other possibilities that would also slow processing time, for example, repeated or multiple cue-based parsing attempts. One advantage of the multiple retrieval attempts interpretation is that it can account for this pattern of results without the postulation of a specialised *reanalysis* mechanism or process (see Martin, 2016; Van Dyke & Lewis, 2003 for a unitary process model).

Although it is difficult to discriminate between a case where the retrieval and interpretation in ambiguous sentences takes longer (increase in process duration) and a case where it is attempted more times (increase in process attempts), differences in the rate parameter of the SAT function have been found when more than

one gap must be filled and when syntactic structure must be revised (Bornkessel, McElree, Schlesewsky, & Friederici, 2004; McElree et al., 2003), indicating that additional processes or instances of retrieval attempts affect of the rate of the SAT function. It also seems unlikely that the rate difference observed is due to differences in retrieval speed, given the growing body of evidence for a speed-invariant direct-access retrieval mechanism underlying sentence comprehension (see McElree, 2006). The interpretation of this rate difference as ambiguity requiring additional attempts to retrieve a subject and interpret it at the dependency, or *reanalysis*, is consistent with the processing situations which have led to the few SAT dynamics parameter results in the literature (Bornkessel et al., 2004; McElree et al., 2003; McElree, Pylkkänen, Pickering, & Traxler, 2006). This processing cost is expressed in the veridical estimates of timecourse available via the SAT procedure, and therefore does not seem to stem from representational differences alone, as could have been the case when reasoning from reaction times alone (Davidson & Martin, 2013). We interpret our results as offering evidence in support of different aspects of *representational difference* accounts, including cue-based accounts, as well as aspects of *processing time difference* accounts. That Diagnosticity raised the likelihood of successful interpretation regardless of ambiguity suggests that relative cue-match or diagnosticity, at least along the dimension of verb semantics, is always facilitatory to processing. But there was no evidence that it was more facilitatory when syntactic cues were ambiguous. The latter finding that veridical estimates of processing speed were slower under ambiguity confirms aspects of processing time difference accounts in as much as it can be taken as evidence for multiple parsing attempts. These hybrid conclusions, and the ability to detect evidence for both representational and processing differences, highlight the potential value of appealing to notions from cue-based retrieval theory, or at least, the value of supposing how constructs supported by convergent evidence from other areas of cognitive psychology might account for or illuminate classic sentence processing phenomena.

Acknowledgements

Disclosure Statement

Funding

Note

1. If the conditions are reduced to perform a 2 (Ambiguity: Ambiguous, Unambiguous) × 2(Diagnosticity: Local noun-related verb, Subject-related verb) rmANOVA, we find a marginal main effect of Ambiguity and a main effect of Diagnosticity, F_1 (1,14) = 3.18, $p < .10$, F_2 (1, 33) = 22.14, $p < .001$ for Ambiguity and F_1 (2,28) = 50.59, $p < .001$, F_2 (2, 66) = 12.28, $p < .01$ for Diagnosticity, respectively. The interaction between Ambiguity and Diagnosticity becomes significant by participants, F_1 (2,28) = 6.37, $p < .05$, F_2 (2, 66) = 1.20, $p < .28$.

ORCID

Andrea E. Martin http://orcid.org/0000-0002-3395-7234

References

Altmann, G. T., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language*, *31*(5), 685–712.

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.

Altmann, G. T., van Nice, K. Y., Garnham, A., & Henstra, J. A. (1998). Late closure in context. *Journal of Memory and Language*, *38*(4), 459–484.

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.

Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory and Language*, *51*, 495–522.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*, 37–60.

Davidson, D. J., & Martin, A. E. (2013). Modeling accuracy as a function of response time with the generalized linear mixed effects model. *Acta Psychologica*, 144(1), 83–96.

Dosher, B. A., & McElree, B. (2003). Memory search: Retrieval processes in short-term and long-term recognition. In J. H. Byrne (Ed.), *Learning & memory* (pp. 373–379). New York: Gale Group.

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

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30(6), 725–745.

Fodor, J. D., & Ferreira, F. (1998). *Sentence Reanalysis*. Dordrecht: Kluwer Academic.

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, 56, 357–383.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6(4), 291–325.

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.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, 29(2), 261–290.

Grodner, D., Gibson, E., Argaman, V., & Babyonyshev, M. (2003). Against repair-based reanalysis in sentence comprehension. *Journal of Psycholinguistic Research*, 32(2), 141–166.

Hemforth, B., Fernandez, S., Clifton, C., Frazier, L., Konieczny, L., & Walter, M. (2015). Relative clause attachment in German, English, Spanish and French: Effects of position and length. *Lingua*, 166, 43–64.

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model comparison approach*. San Diego: Harcourt Brace Jovanovich.

Kohonen, T. (1984). *Self-Organization and associative memory*. Berlin: Springer.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240.

Lau, E. F., & Ferreira, F. (2005). Lingering effects of disfluent material on the comprehension of garden path sentences. *Language and Cognitive Processes*, 20, 633–666.

Lewis, R. L., & Vasishth, S. (2005). An activation based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization (No. 8)*. New York: Oxford University Press.

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

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. New York: Psychology Press.

Martin, A. E. (2016). Language processing as cue integration: Grounding the psychology of language in perception and neurophysiology. *Frontiers in Psychology: Language Sciences*, 7, 120.

Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language*, 58, 879–906.

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1231–1239.

Martin, A. E., & McElree, B. (2011). Direct-access retrieval during sentence comprehension: Evidence from sluicing. *Journal of Memory and Language*, 64, 327–343.

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *NeuroImage*, 59, 1859–1869.

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2014). Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. *Brain and Language*, 135, 42–51.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, 29, 111–123.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46, pp. 155–200). San Diego: Academic Press.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, 118(4), 346–373.

McElree, B., & Dosher, B. A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General*, 122(3), 291.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48, 67–91.

McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 134–157.

McElree, B., Pylkkänen, L., Pickering, M. J., & Traxler, M. (2006). A time course analysis of enriched composition. *Psychonomic Bulletin & Review*, 13, 53–59.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, 10, 389–395.

Pachella, R. G., & Pew, R. W. (1968). Speed-Accuracy Tradeoff in Reaction Time: Effect of Discrete Criterion Times. *Journal of Experimental Psychology*, 76(1p1), 19.

Schouten, J. F., & Bekker, J. A. M. (1967). Reaction time and accuracy. *Acta Psychologica*, 27, 143–153.

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, 105, 477–488.

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*, 40(1), 136–150.

Sturt, P., Scheepers, C., & Pickering, M. (2002). Syntactic ambiguity resolution after initial misanalysis: The role of recency. *Journal of Memory and Language*, 46, 371–390.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50(4), 355–370.

Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language*, 27(6), 597–632.

Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language*, *48*(4), 740–759.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eyemovements. *Journal of Memory and Language*, *47*(1), 69–90.

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

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*, 247–263.

Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, *45*(2), 225–258.

Watkins, O. C., & Watkins, M. J. (1975). Build-up of proactive inhibition as a cue overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *1*, 442–452.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*, 67–85.

Appendices

Appendix 1.

Condition	Sentence
Item 1	
1	The horse raced past the old red barn fell
2	The horse raced past the old red barn stumbled
3	The horse raced past the old red barn neighed
4	The horse that was raced past the old red barn fell
5	The horse that was raced past the old red barn stumbled
6	The horse that was raced past the old red barn neighed
7	*The horse raced past the old red barn unfurled
8	*The horse that was raced past the old red barn unfurled
Item 2	
1	The dog walked near the young pretty woman sat
2	The dog walked near the young pretty woman tripped
3	The dog walked near the young pretty woman barked
4	The dog that was walked near the young pretty woman sat
5	The dog that was walked near the young pretty woman tripped
6	The dog that was walked near the young pretty woman barked
7	*The dog walked near the young pretty woman creased
8	*The dog that was walked near the young pretty woman creased
Item 3	
1	The ballerina sent the beautiful flowers blossomed
2	The ballerina sent the beautiful flowers paid
3	The ballerina sent the beautiful flowers danced
4	The ballerina who was sent the beautiful flowers blossomed
5	The ballerina who was sent the beautiful flowers paid
6	The ballerina who was sent the beautiful flowers danced
7	*The ballerina sent the beautiful flowers ripped
8	*The ballerina who was sent the beautiful flowers ripped
Item 4	
1	The boat floated down the wide rushing river turned
2	The boat floated down the wide rushing river drifted
3	The boat floated down the wide rushing river sank
4	The boat that was floated down the wide rushing river turned
5	The boat that was floated down the wide rushing river drifted
6	The boat that was floated down the wide rushing river sank
7	*The boat floated down the wide rushing river smiled
8	*The boat that was floated down the wide rushing river smiled
Item 5	
1	The car crashed into the tall oak tree burned
2	The car crashed into the tall oak tree exploded
3	The car crashed into the tall oak tree swerved
4	The car that was crashed into the tall oak tree burned
5	The car that was crashed into the tall oak tree exploded
6	The car that was crashed into the tall oak tree swerved
7	*The car crashed into the tall oak tree laughed
8	*The car that was crashed into the tall oak tree laughed
Item 6	
1	The child handed the cute brown puppy begged
2	The child handed the cute brown puppy squirmed
3	The child handed the cute brown puppy laughed
4	The child that was handed the cute brown puppy begged
5	The child that was handed the cute brown puppy squirmed
6	The child that was handed the cute brown puppy laughed
7	*The child handed the cute brown puppy populated
8	*The child that was handed the cute brown puppy populated
Item 7	
1	The woman sold the risky foreign stocks failed
2	The woman sold the risky foreign stocks profited
3	The woman sold the risky foreign stocks smiled
4	The woman who was sold the risky foreign stocks failed
5	The woman who was sold the risky foreign stocks profited
6	The woman who was sold the risky foreign stocks smiled
7	*The woman sold the risky foreign stocks clumped
8	*The woman who was sold the risky foreign stocks clumped
Item 8	
1	The well-known bank lent the foreign money disappeared
2	The well-known bank lent the foreign money closed
3	The well-known bank lent the foreign money collapsed
4	The well-known bank that was lent the foreign money disappeared

Appendix 1. Continued.

Condition	Sentence
5	The well-known bank that was lent the foreign money closed
6	The well-known bank that was lent the foreign money collapsed
7	*The well-known bank lent the foreign money curled
8	*The well-known bank that was lent the foreign money curled
Item 9	
1	The consulting company transferred the valuable assets merged
2	The consulting company transferred the valuable assets moved
3	The consulting company transferred the valuable assets advertised
4	The consulting company that was transferred the valuable assets merged
5	The consulting company that was transferred the valuable assets moved
6	The consulting company that was transferred the valuable assets advertised
7	*The consulting company transferred the valuable assets fluffed
8	*The consulting company that was transferred the valuable assets fluffed
Item 10	
1	The policeman dispatched the yearly crime report objected
2	The policeman dispatched the yearly crime report phoned
3	The policeman dispatched the yearly crime report yelled
4	The policeman who was dispatched the yearly crime report objected
5	The policeman who was dispatched the yearly crime report phoned
6	The policeman who was dispatched the yearly crime report yelled
7	*The policeman dispatched the yearly crime report accumulated
8	*The policeman who was dispatched the yearly crime report accumulated
Item 11	
1	The actress sent the expensive fine jewelry sparkled
2	The actress sent the expensive fine jewelry arrived
3	The actress sent the expensive fine jewelry frowned
4	The actress who was sent the expensive fine jewelry sparkled
5	The actress who was sent the expensive fine jewelry arrived
6	The actress who was sent the expensive fine jewelry frowned
7	*The actress sent the expensive fine jewelry clipped
8	*The actress who was sent the expensive fine jewelry clipped
Item 12	
1	The shopkeeper shipped the exotic goods vanished
2	The shopkeeper shipped the exotic goods appeared
3	The shopkeeper shipped the exotic goods called
4	The shopkeeper who was shipped the exotic goods vanished
5	The shopkeeper who was shipped the exotic goods appeared
6	The shopkeeper who was shipped the exotic goods called
7	*The shopkeeper shipped the exotic goods fizzled
8	*The shopkeeper who was shipped the exotic goods fizzled
Item 13	
1	The girl passed the folded note slipped
2	The girl passed the folded note whispered
3	The girl passed the folded note giggled
4	The girl who was passed the folded note slipped
5	The girl who was passed the folded note whispered
6	The girl who was passed the folded note giggled
7	*The girl passed the folded note snipped
8	*The girl who was passed the folded note snipped
Item 14	
1	The old man fed the warm crusty bread rose
2	The old man fed the warm crusty bread cleaned up
3	The old man fed the warm crusty bread burped
4	The old man who was fed the warm crusty bread rose
5	The old man who was fed the warm crusty bread cleaned up
6	The old man who was fed the warm crusty bread burped
7	*The old man fed the warm crusty bread rained
8	*The old man who was fed the warm crusty bread rained
Item 15	
1	The student passed the difficult math test began
2	The student passed the difficult math test celebrated
3	The student passed the difficult math test chortled
4	The student who was passed the difficult math test began
5	The student who was passed the difficult math test celebrated
6	The student who was passed the difficult math test chortled
7	*The student passed the difficult math test happened
8	*The student who was passed the difficult math test happened
Item 16	
1	The player passed the dirty football slipped
2	The player passed the dirty football scored
3	The player passed the dirty football shouted

Appendix 1. Continued.

Condition	Sentence
4	The player who was passed the dirty football slipped
5	The player who was passed the dirty football scored
6	The player who was passed the dirty football shouted
7	*The player passed the dirty football valued
8	*The player who was passed the dirty football valued
Item 17	
1	The leader awarded the prestigious honorary medal waved
2	The leader awarded the prestigious honorary medal clapped
3	The leader awarded the prestigious honorary medal smiled
4	The leader who was awarded the prestigious honorary medal waved
5	The leader who was awarded the prestigious honorary medal clapped
6	The leader who was awarded the prestigious honorary medal smiled
7	*The leader awarded the prestigious honorary medal cranked
8	*The leader who was awarded the prestigious honorary medal cranked
Item 18	
1	In the schoolyard, the child bounced the shiny red ball rolled
2	In the schoolyard, the child bounced the shiny red ball played
3	In the schoolyard, the child bounced the shiny red ball cried
4	In the schoolyard, the child who was bounced the shiny red ball rolled
5	In the schoolyard, the child who was bounced the shiny red ball played
6	In the schoolyard, the child who was bounced the shiny red ball cried
7	*In the schoolyard, the child bounced the shiny red ball crimped
8	*In the schoolyard, the child who was bounced the shiny red ball crimped
Item 19	
1	After a long day, the woman leased the two-door truck stalled
2	After a long day, the woman leased the two-door truck returned
3	After a long day, the woman leased the two-door truck frowned
4	After a long day, the woman who was leased the two-door truck stalled
5	After a long day, the woman who was leased the two-door truck returned
6	After a long day, the woman who was leased the two-door truck frowned
7	*After a long day, the woman leased the two-door truck sprinkled
8	*After a long day, the woman who was leased the two-door truck sprinkled
Item 20	
1	The teenager sold old beat-up car drove
2	The teenager sold old beat-up car signed
3	The teenager sold old beat-up car complained
4	The teenager who was sold old beat-up car drove
5	The teenager who was sold old beat-up car signed
6	The teenager who was sold old beat-up car complained
7	*The teenager sold old beat-up car drizzled
8	*The teenager who was sold old beat-up car drizzled
Item 21	
1	The driver rented the oversized tractor-trailer signalled
2	The driver rented the oversized tractor-trailer left
3	The driver rented the oversized tractor-trailer sighed
4	The driver who was rented the oversized tractor-trailer signalled
5	The driver who was rented the oversized tractor-trailer left
6	The driver who was rented the oversized tractor-trailer sighed
7	*The driver rented the oversized tractor-trailer doused
8	*The driver who was rented the oversized tractor-trailer doused
Item 22	
1	The writer advanced the lucrative book deal succeeded
2	The writer advanced the lucrative book deal signed
3	The writer advanced the lucrative book deal scoffed
4	The writer who was advanced the lucrative book deal succeeded
5	The writer who was advanced the lucrative book deal signed
6	The writer who was advanced the lucrative book deal scoffed
7	*The writer advanced the lucrative book deal wound
8	*The writer who was advanced the lucrative book deal wound
Item 23	
1	During the planning meeting, the contractor allotted the building materials arrived
2	During the planning meeting, the contractor allotted the building materials hovered
3	During the planning meeting, the contractor allotted the building materials spoke
4	During the planning meeting, the contractor who was allotted the building materials arrived
5	During the planning meeting, the contractor who was allotted the building materials hovered
6	During the planning meeting, the contractor who was allotted the building materials spoke
7	*During the planning meeting, the contractor allotted the building materials fizzed
8	*During the planning meeting, the contractor who was allotted the building materials fizzed
Item 24	
1	After the invasion, the regiment relayed the battle plan scattered
2	After the invasion, the regiment relayed the battle plan radioed in
3	After the invasion, the regiment relayed the battle plan surrendered

Appendix 1. Continued.

Condition	Sentence
4	After the invasion, the regiment who was relayed the battle plan scattered
5	After the invasion, the regiment who was relayed the battle plan radioed in
6	After the invasion, the regiment who was relayed the battle plan surrendered
7	*After the invasion, the regiment relayed the battle plan flicked
8	*After the invasion, the regiment who was relayed the battle plan flicked
Item 25	
1	At the fancy restaurant, the customer recommended the pricey entrée was served
2	At the fancy restaurant, the customer recommended the pricey entrée was well-known
3	At the fancy restaurant, the customer recommended the pricey entrée was talkative
4	At the fancy restaurant, the customer who was recommended the pricey entrée was served
5	At the fancy restaurant, the customer who was recommended the pricey entrée was well-known
6	At the fancy restaurant, the customer who was recommended the pricey entrée was talkative
7	*At the fancy restaurant, the customer who was recommended the pricey entrée was curled
8	*At the fancy restaurant, the customer who was recommended the pricey entrée was curled
Item 26	
1	At the debate, the student delegated the organisational task started
2	At the debate, the student delegated the organisational task conferred
3	At the debate, the student delegated the organisational task spoke
4	At the debate, the student who was delegated the organisational task started
5	At the debate, the student who was delegated the organisational task conferred
6	At the debate, the student who was delegated the organisational task spoke
7	*At the debate, the student delegated the organisational task proved
8	*At the debate, the student who was delegated the organisational task proved
Item 27	
1	Despite the economic downturn, the manager guaranteed the generous raise delivered
2	Despite the economic downturn, the manager guaranteed the generous raise celebrated
3	Despite the economic downturn, the manager guaranteed the generous raise smiled
4	Despite the economic downturn, the manager who was guaranteed the generous raise delivered
5	Despite the economic downturn, the manager who was guaranteed the generous raise celebrated
6	Despite the economic downturn, the manager who was guaranteed the generous raise smiled
7	*Despite the economic downturn, the manager guaranteed the generous raise smothered
8	*Despite the economic downturn, the manager who was guaranteed the generous raise smothered
Item 28	
1	At the hospital orientation, the trainee administered the blood test was negative
2	At the hospital orientation, the trainee administered the blood test was careful
3	At the hospital orientation, the trainee administered the blood test was nervous
4	At the hospital orientation, the trainee who was administered the blood test was negative
5	At the hospital orientation, the trainee who was administered the blood test was careful
6	At the hospital orientation, the trainee who was administered the blood test was nervous
7	*At the hospital orientation, the trainee administered the blood test was crinkled
8	*At the hospital orientation, the trainee who was administered the blood test was crinkled
Item 29	
1	At the administrative headquarters, the secretary remitted the reimbursement form filed
2	At the administrative headquarters, the secretary remitted the reimbursement form waited
3	At the administrative headquarters, the secretary remitted the reimbursement form chatted
4	At the administrative headquarters, the secretary who was remitted the reimbursement form filed
5	At the administrative headquarters, the secretary who was remitted the reimbursement form waited
6	At the administrative headquarters, the secretary who was remitted the reimbursement form chatted
7	*At the administrative headquarters, the secretary remitted the reimbursement form creaked
8	*At the administrative headquarters, the secretary who was remitted the reimbursement form creaked
Item 30	
1	In the busy press room, the celebrity granted the controversial interview prepared
2	In the busy press room, the celebrity granted the controversial interview posed
3	In the busy press room, the celebrity granted the controversial interview chatted
4	In the busy press room, the celebrity who was granted the controversial interview prepared
5	In the busy press room, the celebrity who was granted the controversial interview posed
6	In the busy press room, the celebrity who was granted the controversial interview chatted
7	*In the busy press room, the celebrity granted the controversial interview draped
8	*In the busy press room, the celebrity who was granted the controversial interview draped
Item 31	
1	Before the hearing, the prisoner read the criminal accusation deliberated
2	Before the hearing, the prisoner read the criminal accusation fidgeted
3	Before the hearing, the prisoner read the criminal accusation groaned
4	Before the hearing, the prisoner who was read the criminal accusation deliberated
5	Before the hearing, the prisoner who was read the criminal accusation fidgeted
6	Before the hearing, the prisoner who was read the criminal accusation groaned
7	*Before the hearing, the prisoner read the criminal accusation rumpled
8	*Before the hearing, the prisoner who was read the criminal accusation rumpled
Item 32	
1	At the crowded mall, the salesgirl designated the sales rack priced quickly
2	At the crowded mall, the salesgirl designated the sales rack worked efficiently

Appendix 1. Continued.

Condition	Sentence
3	At the crowded mall, the salesgirl designated the sales rack slacked off
4	At the crowded mall, the salesgirl who was designated the sales rack priced quickly
5	At the crowded mall, the salesgirl who was designated the sales rack worked efficiently
6	At the crowded mall, the salesgirl who was designated the sales rack slacked off
7	*At the crowded mall, the salesgirl designated the sales rack coiled
8	*At the crowded mall, the salesgirl who was designated the sales rack coiled
Item 33	
1	At the big play-off game, the pitcher lobbed the baseball slipped
2	At the big play-off game, the pitcher lobbed the baseball wound up
3	At the big play-off game, the pitcher lobbed the baseball shouted
4	At the big play-off game, the pitcher who was lobbed the baseball slipped
5	At the big play-off game, the pitcher who was lobbed the baseball wound up
6	At the big play-off game, the pitcher who was lobbed the baseball shouted
7	*At the big play-off game, the pitcher lobbed the baseball swarmed
8	*At the big play-off game, the pitcher who was lobbed the baseball swarmed
Item 34	
1	At the airport, the flight crew issued the free tickets dispersed
2	At the airport, the flight crew issued the free tickets departed
3	At the airport, the flight crew issued the free tickets celebrated
4	At the airport, the flight crew who were issued the free tickets dispersed
5	At the airport, the flight crew who were issued the free tickets departed
6	At the airport, the flight crew who were issued the free tickets celebrated
7	*At the airport, the flight crew issued the free tickets tore
8	*At the airport, the flight crew who were issued the free tickets tore

Appendix 2

Notes on Filler stimuli and experiments.

Participants saw a total of 2576 sentence stimuli over 8 one-hour sessions. The sentence stimuli were as follows: There were 9 sub-experiments in the 9 sessions (8 experimental sessions with 322 stimuli, 34 from this study and 288 from other studies, plus a 45-minute practice session to familiarise the participants with the button-pressing part of the task). This resulted in ~33% of the stimuli having a relative clause, where 16.4% was a full relative clause and 16.4% was reduced. With regard to the violations, all violations came from ungrammatical sentences. There were a total of 1220 ungrammatical stimuli, or 47.3% of the total stimuli. Of those violations, 41% stemmed from violation of the subcategorization requirements of a verb in the sentence. Of the total number of stimuli presented, that means that 19.4% had a sub-categorisation violation. Below I give a more detailed description of the sub-experiments which composed the 10 session experiment. Sub-experiments 1–3 featured relative clauses, and sub-experiments 1–4 used subcategorization violations to create ungrammaticality.

Sub-experiment (SE) 1. Garden-path study from this manuscript. 272 stimuli total, 34 items in 8 Conditions, 75% grammatical. Relative clauses: 50% full, 50% reduced. Source of violation: subcategorization restrictions.

SE2. Reduced vs. Unreduced Relative clauses study varying noun phrases. 288 stimuli total, 36 items in 8 conditions, 50% grammatical. Relative clauses: 50% full, 50% reduced. Source of violation: subcategorization restrictions.

SE3. Reduced vs. Unreduced Relative clauses study varying distance. 288 stimuli total, 36 items in 8 conditions, 50% grammatical. Relative clauses: 50% full, 50% reduced. Source of violation: subcategorization restrictions.

SE4. Sub-categorisation violation study. 288 stimuli total, 36 items in 8 conditions, 50% grammatical. No relative clauses. Source of violation: subcategorization restrictions.

SE5. Verb-phrase ellipsis study. 288 stimuli total, 36 items in 8 conditions, 50% grammatical. No relative clauses.

SE6. & 7. Sluicing study. 576 stimuli total, 36 items in 16 conditions, 50% grammatical. No relative clauses.

SE8. Pseudogapping study. 288 stimuli total, 36 items in 8 conditions, 50% grammatical. No relative clauses.

SE9. Gapping study. 288 stimuli total, 36 items in 8 conditions, 50% grammatical.

Given the diversity of the fillers, and the reliable presence of relative clauses (33%), we do not believe that either prediction or surprisal related to relative clause processing can explain our pattern of effects. Secondly, we asked participants to judge the acceptability of a sentence akin as they would in a perceptual task – is a natural meaning evidence without over thinking or reflection. Acceptability in this context is essentially a grammaticality judgment, though we do not refer to it in this way with participants. In our experience, if participants are asked to make a grammaticality judgment, they focus their attention on spelling, punctuation and other prescriptive phenomena rather than the perception of natural meaning/grammaticality that we intend. For this reason we refer to the judgment as an acceptability judgment.

The majority of the ungrammatical stimuli were ungrammatical due to a subcategorization violation. Thus we do not believe that participants were basing their decision on plausibility.

We see the reviewer's point – not all the items worked this way but some do use violation of transitivity to achieve a subcategorization violation leading to ungrammaticality. We would like to note that a period always occurred after the final verb so participants were aware that no more words or arguments were coming. We believe that the foil verbs are incompatible with both nouns in the sentence (except for Items 9 and 15 where the source of subcategorization violation is slightly different). We would also like to note that the verbs used to form violations were used grammatically in the filler experiments so that participants could not simply reject the sentence based on a verb class (if that were the case, it would predict no main effect of condition as we indeed observed).