JSLHR

Research Article

The Auditory Comprehension of *Wh*-Questions in Aphasia: Support for the Intervener Hypothesis

Shannon M. Sheppard,[a,b] Matthew Walenski,[a] Tracy Love,[a,b] and Lewis P. Shapiro[a,b]

Purpose: This study examines 3 hypotheses about the processing of *wh*-questions in both neurologically healthy adults and adults with Broca's aphasia.
Method: We used an eye tracking while listening method with 32 unimpaired participants (Experiment 1) and 8 participants with Broca's aphasia (Experiment 2). Accuracy, response time, and online gaze data were collected.
Results: In Experiment 1, we established a baseline for how unimpaired processing and comprehension of 4 types of *wh*-question (subject- and object-extracted *who*- and *which*-questions) manifest. There was no unambiguous support found for any of the 3 hypotheses in Experiment 1. In Experiment 2 with the Broca's participants, however, we found significantly lower accuracy, slower response times, and increased interference in our gaze data in the object-extracted *which*-questions relative to the other conditions.
Conclusions: Our results provide support for the intervener hypothesis, which states that sentence constructions that contain an intervener (a lexical noun phrase) between a displaced noun phrase and its gap site result in a significant processing disadvantage relative to other constructions. We argue that this hypothesis offers a compelling explanation for the comprehension deficits seen in some participants with Broca's aphasia.

We describe an investigation of the time course of processing *wh*-questions during sentence comprehension in a group of neurologically healthy adult participants and a group of participants with neurological impairment, specifically aphasia. To begin, consider this declarative sentence:

1. A fireman pushed the policeman.

Sentence 1 is in subject-verb-object (S-V-O), canonical word order. The verb *push* requires two arguments, one playing the thematic role of agent and the other playing the theme role. In Sentence 1, the agent role is assigned to the subject position occupied by the noun phrase (NP) *a fireman*, and the theme is assigned to the direct object position occupied by the NP *the policeman*.

In various linguistic and psycholinguistic accounts, *wh*-questions are derived from their S-V-O counterparts by extracting (and fronting) the questioned element (see Sentence 2 below). The similarities (and differences) across structures can be captured generally by how thematic roles are assigned by the verb or verb phrase to its argument positions. Consider the following:

2. Two firemen and a policeman got into a fight.
   a. Which fireman pushed the policeman?
   b. Which fireman did the policeman push <~~which fireman~~>?

Thematic role assignment in the subject-extracted *which*-question (Sentence 2a) is similar to that of the canonical declarative sentence in Sentence 1. In the object-extracted question (Sentence 2b), the *which*-phrase has been displaced from its underlying direct object position to a position before the verb, yielding noncanonical word order. Although the theme role is assigned to the direct object position as it is in Sentences 1 and 2a, that position is occupied by an unpronounced copy (or trace/gap) of the displaced *wh*-phrase; the thematic role is transferred to the displaced *wh*-phrase via a chain that connects the copy to its displaced element, forming a dependency relationship between the two positions (Chomsky, 1981, 1995).

Object-extracted questions, such as Sentence 2b, both intuitively and empirically, are more difficult to understand

[a]San Diego State University, CA
[b]University of California, San Diego, CA
Correspondence to Shannon M. Sheppard: SBM.Sheppard@gmail.com

Editor: Rhea Paul
Associate Editor: Jessica Richardson
Received April 9, 2014
Revision received July 8, 2014
Accepted January 10, 2015
DOI: 10.1044/2015_JSLHR-L-14-0099

Disclosure: The authors have declared that no competing interests existed at the time of publication.

than their subject-extracted counterparts (Sentence 2a). One simple explanation for this processing difference is a general word order hypothesis: In English, listeners expect and prefer sentences in S-V-O order, and thus, these structures yield a processing advantage over sentences that do not conform to S-V-O order, such as object-extracted structures (Bates, Friederici, & Wulfeck, 1987; O'Grady, Yamashita, & Lee, 2005). The word order hypothesis requires generalization to similar constructions (as would any hypothesis). Consider, then, subject- and object-extracted *who*-questions (Sentences 3a and 3b, respectively):

3a. Who pushed the policeman?

3b. Who did the policeman push <~~who~~>?

Similar structural and thematic role considerations apply. Thus, the word order hypothesis makes no distinction between these question types, that is, between *who*- and *which*-questions.

Yet there are linguistic and processing differences between these question types. One involves the syntax-discourse interface. *Which*-questions are discourse-linked (D-linked) because they must refer to an individual from a set of individuals in the discourse as in Sentence 2, and *who*-questions are not necessarily D-linked. It has been suggested that D-linked constructions are more difficult to process because interface conditions are more resource intensive (e.g., Avrutin, 2000, 2006; Burkhardt, 2005; Rothman & Slabakova, 2011; Shapiro, 2000). It is interesting that the distinction between these two question types has been observed even when a relevant discourse was presented for both *which*- and *who*-questions (e.g., Donkers & Stowe, 2006; Salis & Edwards, 2005; Shapiro, 2000). Thus, the D-linked hypothesis predicts a processing advantage for *who*-questions relative to *which*-questions. Note that D-linking should predict similar patterns for both subject- and object-extracted questions although Avrutin (2000) suggests that the combination of filling a gap and D-linking depletes the limited processing resources of people with Broca's aphasia, leading to difficulty with object-extracted *which*-questions. We return to this suggestion in the Discussion.

The intervener hypothesis is the primary focus of our study. Computing the dependency relationship between the displaced NP and its gap in both Sentences 2b and 3b requires crossing over another argument—the policeman—which we call the intervener. There is no such intervener in the subject-extracted examples in Sentences 2a and 3a. The intuition here is that the intervener interferes with computing the dependency relationship because it is a possible element in the dependency chain, rendering a processing disadvantage to such structures over those that don't contain an intervener. To be more formal, this hypothesis stems from Rizzi's relativized minimality account (1990; see also Grillo, 2005, 2009). Relationships among arguments in a sentence are constrained by a locality condition:

4. Given a structure: . . .X. . .Z. . .Y. . ., Y is in a local (minimal) configuration with X if and only if there is no Z that has the following properties:

   a. Z is of the same structural type as X, and

   b. Z intervenes between X and Y.

In other words, computing the dependency relationship between two elements becomes more difficult because the structurally similar intervener is a potential site for one of the elements in the dependency (e.g., see Friedmann, Belletti, & Rizzi, 2009). We suggest that some adults with a language disorder are particularly vulnerable to interveners during sentence processing perhaps because they are susceptible to interference among similarly structured NPs.

One issue for the intervener account is what *the same structural type* means. One possibility is that the intervener is restricted to a lexically specified NP (e.g., *the policeman* in Sentence 2b) that is similar in structure to the displaced phrase (e.g., *which fireman*), as opposed to, for example, a pronoun or proper name. This constraint suggests a distinction between *who*- and *which*-questions, in which the latter phrase (i.e., *which*-NP) has the structure of a fully specified NP (i.e., determiner-noun), and the former phrase (i.e., *who*) does not. The intervener hypothesis, then, suggests no distinction between subject- and object-extracted *who*-questions because neither involves an intervener. However, *which*-questions should reveal an asymmetry between subject- and object-extraction given that only the object-extracted *which*-question involves an intervener.

There is some evidence in the adult literature (healthy participants and those who have aphasia) that *who*- and *which*-questions yield different comprehension patterns (e.g., Donkers, Hoeks, & Stowe, 2013; Frazier & Clifton, 2002; Hickok & Avrutin, 1996). For example, Donkers and Stowe (2006) conducted a self-paced reading study with Dutch-speaking participants comparing standard *who*- and *which*-questions as well as a generic *which* condition (which person) that resembled *who*-questions because they were not set-restricted. They found that standard *which*-questions required longer processing times compared to *who*-questions in object-extracted but not subject-extracted constructions. They also discovered that standard *which*-questions required longer processing times compared to the generic *which*-questions. Donkers and Stowe suggested that the longer processing times often associated with *which*-questions may be due to the process of set-restriction that generally accompanies such questions (for similar accounts, see also De Vincenzi, 1991; Frazier & Clifton, 2002; Frazier, Plunkett, & Clifton, 1996).

Thus far we have kept to a description of an intervener in structural terms although this hypothesis can be extended to include other properties of the intervener that might affect sentence comprehension. Here we take our initial cue from the work of Gordon and colleagues, who have conducted several studies examining how similar NPs interfere with one another during adult sentence processing (e.g., Gordon, Hendrick, & Johnson, 2004; Gordon, Hendrick, Johnson, & Lee, 2006). This work suggests a similarity-based interference account of memory, according to which the demands on storage and retrieval during sentence comprehension are increased when there are NPs in a sentence that have similar representations. Using tasks such as eye tracking while reading, Gordon and colleagues found that reading times increased when two NPs (e.g., the displaced argument and the subject of a relative clause) were both descriptive (i.e., determiner-noun) relative to when the intervening NP was a proper name or a pronoun. This pattern is consistent with the intervener account (see also Van Dyke, 2007; Van Dyke & McElree, 2011, for a similar interference proposal).

More to the primary purpose of this article, differences between processing *who*- and *which*-questions have also been observed in those patients with Broca's aphasia who evince comprehension deficits. Broca's aphasia is characterized by nonfluent and halting speech and was originally thought to be a disorder of speech production (see Grodzinsky, 2000, for a history of Broca's aphasia). However, research that began in the 1970s has since revealed that a comprehension disorder may also be present although it is not surprising that there is considerable disagreement on the source(s) and generality of the disorder. One ubiquitous result is that these individuals have difficulty comprehending noncanonically ordered sentences in which an argument has been displaced, such as passives, object-extracted relative clauses, and *wh*-questions (e.g., Caramazza & Zurif, 1976; Drai & Grodzinsky, 2006; Grodzinsky, 1990).

Hickok and Avrutin (1996) investigated *who*- and *which*-question comprehension in two patients with Broca's aphasia using untimed sentence–picture matching tasks. Subject-extracted *which*-questions were comprehended significantly better than object-extracted *which*-questions, yet there was no difference in comprehension between subject- and object-extracted *who*-questions (see also Frazier & McNamara, 1995; Salis & Edwards, 2005). Thompson, Tait, Ballard, and Fix (1999) replicated Hickok and Avrutin's results in the comprehension of passivized *wh*-questions, using a figure-manipulation task and a picture-pointing task although in only one of four participants with agrammatic aphasia. Friedmann and Gvion (2012) tested the intervener hypothesis with four participants with agrammatic Broca's aphasia. They tested subject- and object-extracted relative clauses with each relative clause type having both an intervener and nonintervener condition (on the basis of Hebrew structure). They found that performance on nonintervener conditions was above chance, and performance on intervener conditions was no better than chance. These patterns suggest that comprehension success was not on the basis of sentence type (subject- vs. object-relatives) but instead was based on whether or not an NP intervened between the filler and gap (see also Friedmann & Shapiro, 2003, footnote 4). This is an important initial finding, but because offline sentence–picture matching tasks cannot measure how participants arrive at their final interpretation, the evidence that can be used to adjudicate different accounts is limited.

We end this section with a description of some online work that is relevant to our study. Dickey, Choy, and Thompson (2007) used an eye tracking while listening method to investigate processing of *wh*-questions. Participants listened to a story (as in Example 5 below) while their eye movements to pictures of the elements/characters in the story were recorded. The participants were then presented with critical comprehension probes that were either object *who*-questions (Sentence 5a), object clefts (Sentence 5b), or control yes/no questions (Sentence 5c):

5. This story is about a boy and a girl. One day they were at school. The girl was pretty, so the boy kissed the girl. They were both embarrassed after the kiss.
    a. Who did the boy kiss that day at school?
    b. It was the girl who the boy kissed that day at school.
    c. Did the boy kiss the girl that day at school?

Dickey et al. (2007) found that the participants in both the control and aphasia groups demonstrated eye movements indicative of successful online comprehension of the *who*-questions. However, participants with aphasia were significantly less accurate in their responses to the *who*-questions and the object cleft questions (Sentences 5a and 5b, respectively) compared to both the control yes/no questions (Sentence 5c) and the control group's responses to these types of questions. They concluded that agrammatic individuals' online processing of syntactic dependencies in *who*-questions is relatively unimpaired but that comprehension breaks down during the interpretation phase, possibly due to "weakened" syntactic representations.

In a follow up, Dickey and Thompson (2009) examined object-relative clauses and passives with an eye tracking while listening method. Consider their object-relative sentences (Sentence 6a):

6. One day a bride and groom were walking in a mall. The bride was feeling playful, so the bride tickled the groom. A clerk was amused.
    a. Point to who the bride was tickling ___ in the mall.

Once again, convincing evidence was observed for associating the filler to the gap for their participants with aphasia using eye-gaze measures although with a slightly delayed time course relative to control participants. However, the displaced NP was a bare *wh*-phrase. Thus, on the intervener account, there should be no interference among the NPs in the sentence, and normal patterns should emerge. Even so, contrary to the intervener account, the accuracy data clearly showed that the participants with aphasia did not understand these sentence structures; the intervener account predicts reasonably good performance on structures that do not contain interveners.[1]

[1]We note that the participants in the Dickey and Thompson (2009) study were diagnosed as agrammatic primarily on the basis of production measures. We take a different approach as described in our Method section. To be brief, here our participants were selected for their specific sentence-comprehension deficits because comprehension is the focus of our study.

In the current article, we report on two experiments using an eye tracking while listening method to investigate processing differences between four question types: subject- and object-extracted *who*- and *which*-questions. In spite of the offline evidence suggesting processing distinctions between *wh*-question types in unimpaired populations and populations with neurological impairment, no studies we are aware of have used an online method to examine processing differences between different types of *wh*-questions in patients with aphasia, and only two (Dickey et al., 2007; Dickey & Thompson, 2009) used an online method to examine the processing of any types of *wh*-questions in patients with agrammatic Broca's aphasia.

Three different hypotheses described in this introduction were investigated: word order, D-linking, and the intervener hypothesis. As shown in Table 1, investigating the four question types in this study can differentiate the predictions made by these accounts. If object-extracted questions were found to be more difficult than subject-extracted questions regardless of *wh*-type, the word order hypothesis would be supported. If *which*-questions were observed to be more difficult than *who*-questions regardless of extraction type, the D-linked hypothesis would be supported. Last, if the *which* object-extracted condition yielded distinct behavior from the other three question types, then the intervener hypothesis would be supported. We examined offline comprehension of these questions, and we also examined online gaze behavior for different segments in the sentences of interest, allowing us to understand if our participants' gaze behavior supported any of the hypotheses.

Experiment 1: *Wh*-Questions in College-Age Adults

We begin with an experiment testing our hypotheses in a group of neurologically healthy college-age participants. Although we do not expect these participants to evince offline comprehension difficulties with *wh*-questions, at least in terms of accuracy, we may be able to detail the underlying basis for their offline comprehension by charting their eye movements as they listen to sentences. Furthermore, this experiment serves as a baseline for our subsequent Experiment 2 that investigates *wh*-question comprehension in a group of individuals with aphasia, allowing us to determine if any of our hypotheses generalize to both neurologically intact participants and those with aphasia.

Method

Participants

We tested 32 neurologically unimpaired, right-handed, college-age students (24 women, eight men) who were monolingual speakers of American English. Their mean age was 20.2 years old (range 18–30). All had normal or corrected-to-normal self-reported visual and auditory acuity. As indicated by self-report, all participants were neurologically and physically stable at the time of testing with no history of alcohol or drug abuse, psychiatric illness, or other significant brain disorder or dysfunction.

Materials

We created 65 action pictures containing three figures interacting with each other (see Figure 1) to go with 65 discourse sentences describing the figures and the action in the scene. Forty-six of these pictures were experimental stimuli with four question types (see Table 2; *which*-subject, *which*-object, *who*-subject, *who*-object) matched to each experimental picture. The remaining 19 pictures served as fillers; each filler picture was matched with one filler *who*- and one filler *which*-question. In the pictures, the figure on the left was performing an action on the middle figure, who was, in turn, performing that same action on the figure on the right. The figure on the left was always the correct answer for the *who*-subject and *which*-subject questions, the figure on the right was always the correct answer for the *who*-object and *which*-object questions, and the middle figure was always the answer to the filler questions and was never a correct answer for any of the experimental questions. The discourse sentences and questions were recorded by a male speaker at a normal average speaking rate of 4.85 syllables/s.

Design

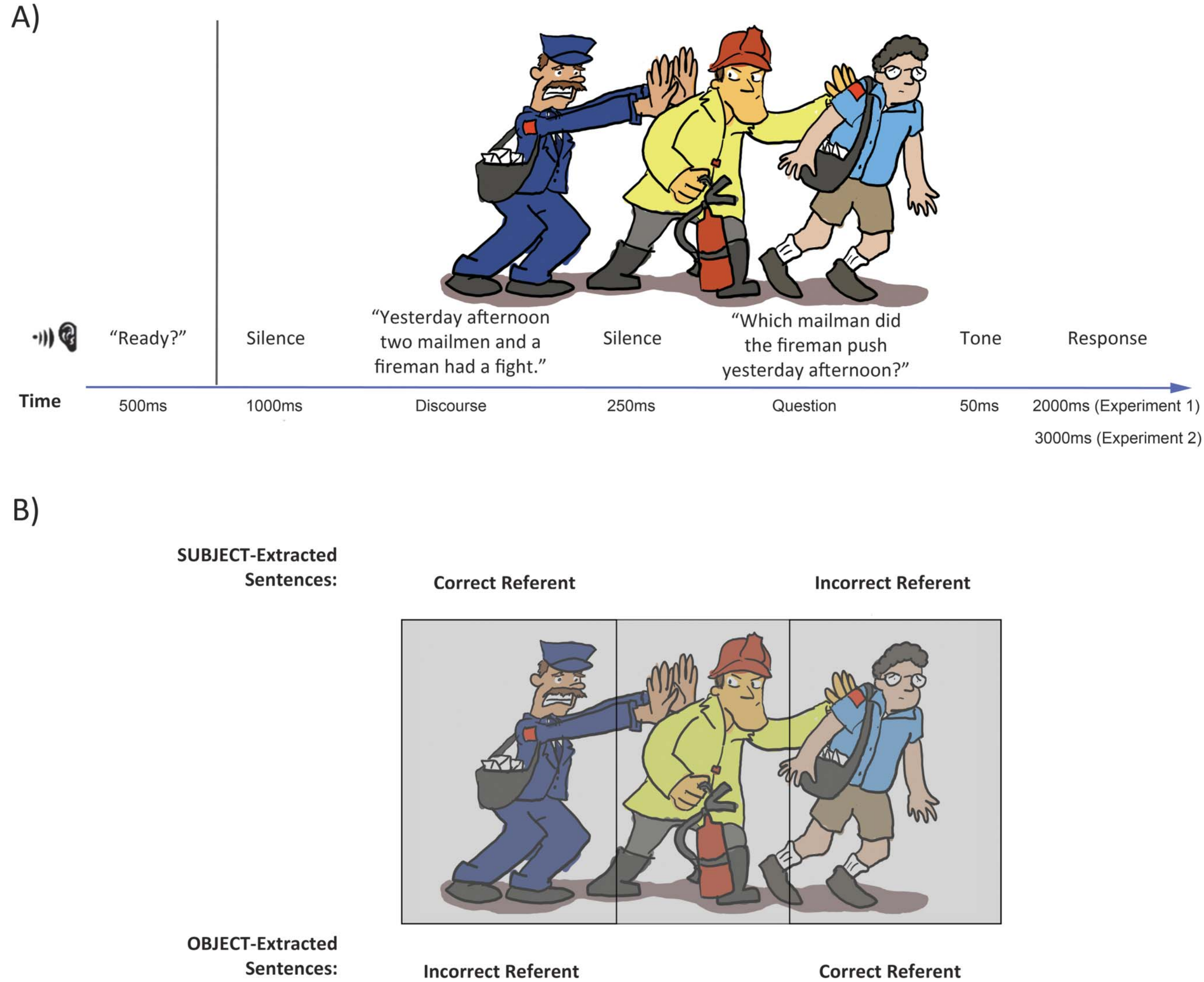
The 46 experimental sets consisting of four sentences each (a total of 184 experimental items) and 19 filler discourse sets consisting of two sentences each (a total of 38 items) were counterbalanced across four presentation lists such that each list contained one of the *wh*-question conditions for each experimental item (i.e., subject- and object-extracted *who*- and *which*-questions); the filler items were rotated through their *who*- and *which*-question versions twice. We used a partially within-subjects design, in which each participant completed two test sessions with one presentation list per session.

Table 1. Hypotheses tested.

Types of questions	Word order hypothesis	Discourse hypothesis	Intervener hypothesis
Who subject-extracted	+	+	+
Who object-extracted	–	+	+
Which subject-extracted	+	–	+
Which object-extracted	–	–	–

Note. + indicates a processing advantage; – indicates a processing disadvantage.

Figure 1. Examples of time course and regions of interest for one trial. (A) An example of the time course of one trial. Participants were presented with the word *Ready?* on the screen for 500 ms, after which the picture appeared and remained throughout the entire trial. Sentences were presented aurally at a normal speech rate, and eye movements were captured throughout the entire trial. The trial ended when the participant made a response, or in the event of no response, it ended after 2 s in Experiment 1 and after 3 s in Experiment 2. (B) Regions of interest for a sample picture. The correct (the figure on the left for the subject-extracted questions and the figure on the right for the object-extracted questions) and incorrect (the figure on the right for the subject-extracted questions and the figure on the left for the object-extracted questions) referents each comprised 40% of the picture with the remaining 20% for the middle referent. The words did not appear on screen but are shown here only to aid the reader. See text for additional details.



Experimental and filler items were presented in the same pseudorandom order for each presentation list and were intermixed such that the same question condition or discourse type (experimental, filler) never occurred more than twice in a row. Experimental sessions were separated by at least 1 week to minimize potential exposure effects.

Procedure

A Tobii eye-tracker with a sampling rate of 60 Hz was used to collect gaze data. The participants sat facing the eye-tracker with their eyes at a distance of 60 cm. The eye-tracker was calibrated for each participant at the beginning of each experimental session. The stimuli were presented with E-prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). The timing for each trial was as follows (see Figure 1A): A ready screen was presented for 500 ms, then the picture was presented for 1 s before the sentence discourse was presented aurally (the picture remained on the screen for the entire trial). At the end of the discourse sentence, there was a period of 250 ms of silence followed by the question probe; a 50-ms tone signaled the end of the question. Participants were instructed to answer the question using a three-button response box (recording response time [RT] and accuracy) as soon as they heard the tone. After a 2-s response period, the next trial automatically began. The response box had

Table 2. Example experimental sentences by type.

Experimental Sentences	Example
Discourse	Yesterday afternoon two mailmen and a fireman had a fight.
Who subject-extracted	Who ___ pushed the fireman yesterday afternoon?
Who object-extracted	Who did the fireman push ___ yesterday afternoon?
Which subject-extracted	Which mailman ___ pushed the fireman yesterday afternoon?
Which object-extracted	Which mailman did the fireman push ___ yesterday afternoon?
Filler Sentences	**Example**
Discourse	Two waitresses and a golfer met at target practice.
Who subject-extracted	Who ___ shot the waitress at target practice?
Which subject-extracted	Which person ___ shot the waitress at target practice?

three buttons labeled *left*, *middle*, and *right* corresponding to the three figures in each picture. Each participant was given a practice session at the start of each experimental session to insure that they understood the task and were acclimated to the experimental procedure. Eye-gaze location was recorded every 17 ms throughout the entire trial (beginning with the picture presentation and ending with their button-press response).

Results

We begin with our accuracy and the RT data. Analyses of these offline data were conducted using restricted maximum likelihood in mixed-effects regression models separately for each dependent variable. A logit-link function (for binary outcome data) was used for accuracy analyses (SAS 9.3 Proc Glimmix). Button presses corresponding to the wrong answer to the question and no response errors were scored as incorrect. RTs for correct responses were analyzed with SAS 9.3 Proc Mixed. In order to account for by-participant and by-item variance in a single statistical test (i.e., in lieu of separate by-subject and by-item tests), each model included crossed random effects on the intercept of participant and item. Each model also included fixed effects of extraction type (subject vs. object), *wh*-question type (*who* vs. *which*), and their interaction. The models were fit with an unstructured covariance matrix for each random effect. Type III *F* tests are reported for main effects and interactions. For a priori planned subcontrasts of our fixed effects (e.g., *who*-subject vs. *who*-object), we computed *t* tests of the differences of the least square means from the full model and report the regression coefficient B (with standard error in parentheses), *t* statistics, and 95% confidence intervals. All *p* values are reported two-tailed. Degrees of freedom were computed using the Satterthwaite approximation. Note that the degrees of freedom are large because, in these models, they are based on the number of data points, not the number of participants or items. For further discussion of these statistical methods, see Baayen (2004, 2008) and Littell, Milliken, Stroup, Wolfinger, and Schabenberger (2006). Note also that Barr, Levy, Scheepers, and Tily (2013) argue that random-intercepts-only models are anticonservative, at least when a model with a more maximal random effects structure converges. In our data, the more maximal models frequently failed to converge (even those with only one additional random effect). Thus, we report results for all models with random intercepts only.

For accuracy, there was a significant main effect of extraction type: subject-extracted questions (96.5%) were more accurate than object-extracted questions (93%), $F(1, 2{,}940) = 15.85$, $p < .0001$. A significant main effect of *wh*-question type was also observed, $F(1, 2{,}940) = 9.91$, $p = .002$, with *who*-questions (96%) yielding more accurate performance than *which*-questions (93.5%). The interaction between extraction site and question type did not reach significance, $F(1, 2{,}940) = 3.27$, $p = .07$. Even so, given our hypotheses, we examined if there were significant differences within and across question and extraction types. The object-extracted *which*-questions (92%) were not reliably different from the subject-extracted *which*-questions (95%), B = −0.45 (0.24), $t(2{,}940) = 1.90$, $p = .06$, 95% CI: (−0.91, 0.02). The object-extracted *who*-questions (94%) were answered less accurately than the subject-extracted *who*-questions (98%), B = −1.15 (0.32), $t(2{,}940) = 3.64$, $p = .0003$, 95% CI: (−1.77, −0.53). There was no difference in accuracy between the object-extracted *which*- (92%) and *who*-questions (94%), B = −0.26 (0.22), $t(2{,}940) = 1.19$, $p = .23$, 95% CI: (−0.69, 0.17), but a significant difference was found between the subject-extracted *which*- (95%), and *who*-questions (98%), B = −0.96 (0.32), $t(2{,}940) = 3.00$, $p = .003$, 95% CI: (−1.59, −0.33).

Analysis of the RT data revealed that, overall, there were significant main effects of extraction type, with object-extracted questions (709 ms) evincing slower RTs than subject-extracted questions (669 ms), $F(1, 2{,}683) = 29.69$, $p < .0001$, and *wh*-question type, with *which*-questions (697 ms) yielding longer RTs than *who*-questions (681 ms), $F(1, 2{,}683) = 4.04$, $p = .04$). The interaction between extraction type and *wh*-question type was not significant, $F(1, 2{,}684) = 2.36$, $p = .12$. As with the accuracy data, we analyzed RTs within and across question and extraction types. RTs for object-extracted *which*-questions (723 ms) were significantly slower than the RTs for subject-extracted *which*-questions (671 ms), B = 55 (11), $t(2{,}687) = 4.92$, $p < .0001$, 95% CI: (33, 77). Object-extracted *who*-questions also revealed slower RTs (695 ms) than subject-extracted *who*-questions (667 ms) B = 30 (11), $t(2{,}680) = 2.78$, $p = .006$, 95% CI: (9, 52). The RTs for object-extracted *which*-questions (723 ms) were significantly slower than the RTs for object-extracted *who*-questions (695 ms) B = 28 (11), $t(2{,}689) = 2.49$, $p = .01$, 95% CI: (6, 50). The RTs for subject-extracted *which*-questions did not differ from subject-extracted *who*-questions, B = 4 (11), $t(2,677) = 0.33$, $p = .74$, 95% CI: (−18, 25). Note that RT data were screened prior to analysis by removing outliers outside the inner fence of a box plot separately by condition (1.3% of data).

Gaze Analysis and Data

For analysis, each picture was divided into three regions of interest (see Figure 1B), corresponding to the left, middle, and right figures. For all items, we analyzed the mean proportion of gazes in each region of interest during the question portion and response period for each experimental item. We divided each item into multiple time windows, measuring the onset and offset of each window for each item individually, then adding 200 ms to both onset and offset to account for gaze delay (Allopenna, Magnuson, & Tanenhaus, 1998). The subject-extracted sentences were divided into four time windows (*wh*-phrase, verb, object, end of sentence, e.g., which mailman/who | pushed | the fireman | yesterday afternoon) plus the response period (i.e., the period from the end of the sentence until a button press response was made). The object-extracted sentences were divided into five time windows (*wh*-phrase, auxiliary, intervener, verb-gap, end of sentence, e.g., which mailman/who | did | the fireman | push | yesterday afternoon) plus the response period. A gaze was conservatively defined as seven consecutive looks to a particular region of interest (i.e., 102 ms or more gaze duration; see Manor & Gordon, 2003). For each participant and each item, we calculated the proportion of gazes to each region of interest (subject, middle, object) separately for each time window for each condition (*who*-subject, *who*-object, *which*-subject, *which*-object). This proportion (i.e., gazes to the region of interest out of all gazes during the time window) was our dependent variable. Note that the proportion of gazes was treated as a binary variable: Either the gaze was within the region or it wasn't.

We analyzed the gaze data's change over time for each condition separately using restricted maximum likelihood in mixed-effects regression models, using a logit-link function for binomial data. Each model included crossed random effects of participant and item on the intercept and a fixed effect of *wh*-type. Note that due to models frequently failing to converge, the random effect of item was not included in the results reported below (in models that did converge with this random effect, the results were essentially identical to the same model results without this factor). The models were fit with unstructured covariance matrices for each random effect. We report the coefficient (with standard error in parentheses), *t* statistics, and 95% confidence intervals. Alpha was set to 0.05 for all effects. Analyses were conducted using SAS version 9.3 Proc Glimmix (SAS Institute, Inc.). Given the differences in the linear positions of the time windows, object-extracted and subject-extracted sentences were not directly compared. We examined gazes in separate models for each extraction type in each time window for the correct referent (i.e., the subject in the subject-extracted sentences located to the left of the middle referent and the object in the object-extracted sentences located to the right of the middle referent) and the incorrect referent (i.e., the object in the subject-extracted sentences and the subject in the object-extracted sentences). Gazes to the middle region were not included in analysis as every trial began with a ready screen with the text centered on the screen where the middle figure was subsequently located, biasing gaze position. In addition, the left and right figures were always the same type of referent (i.e., the two mailmen in Figure 1A) whereas the middle figure was always a different type of referent (i.e., the fireman in Figure 1A).

Subject-Extracted Sentences

Gazes to the correct referent (see Figure 2A) revealed that *which*-questions yielded a larger proportion of correct gazes than *who*-questions at the verb time window (25% *which*; 16% *who*), B = 0.57 (0.14), $t(1,329) = 4.09$, $p < .0001$, 95% CI: (0.30, 0.85), and at the middle NP time window (31% *which;* 20% *who*), B = 0.63 (0.13), $t(1,351) = 4.89$, $p < .0001$, 95% CI: (0.38, 0.89), but not at the other time windows (*p*s > .30). Gazes to the incorrect referent were not different for *which*- and *who*-sentences at any time window (see Figure 2B; all *p*s > .08).

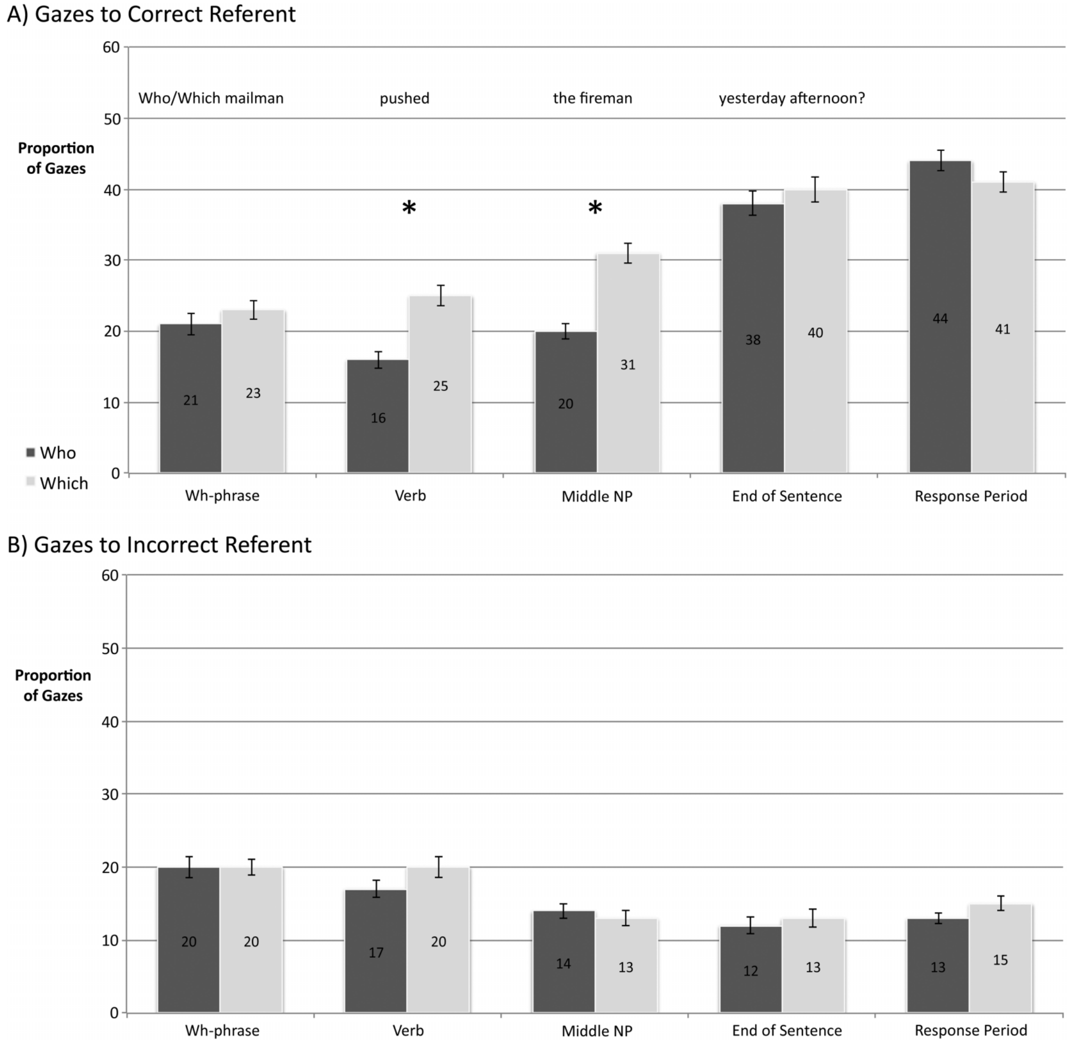
Object-Extracted Sentences

First examining gazes to the correct referent (i.e., the mailman on the right in Figure 1A), the results (see Figure 3A) indicate significant differences in gaze proportion between the *which*- and *who*-questions only for the intervener NP time window (25% *which*; 19% *who*), B = 0.32 (0.13), $t(1,352) = 2.36$, $p = .02$, 95% CI: (0.05, 0.58); all other *p*s > .17. For gazes to the incorrect referent (i.e., the mailman on the left in Figure 1A), the results (see Figure 3B) indicate significantly increased gaze proportions for *which*-questions relative to *who*-questions only at the auxiliary time window (25% *which*; 19% *who*), B = 0.41 (0.14), $t(1,266) = 2.93$, $p = .003$, 95% CI: (0.14, 0.69), and the intervener NP time window (18% *which*, 13% *who*), B = 0.39 (0.15), $t(1,352) = 2.53$, $p = .01$, 95% CI: (0.09, 0.69), all other *p*s > .43.

Discussion

The college-age participants responded with close-to-ceiling accuracy and well above chance in every condition (93%–98%). Responses were less accurate for object-extracted questions than subject-extracted questions (across question type), and less accurate for *which*-questions than *who*-questions (across extraction type). The former result suggests some support for the word order hypothesis (see Table 1) although we note that the effect of extraction was significant in pairwise contrasts only for *who*-questions, not *which*-questions. Likewise, the question-type result suggests support for the discourse hypothesis (see Table 1) although again we note that in pairwise comparisons the effect held only for subject-extracted sentences, not object-extracted sentences. Thus, the support is not strong for either account, and we urge caution

Figure 2. Subject-extracted sentences for Experiment 1 (unimpaired participants): Proportion of gazes for all responses (correctly and incorrectly answered by the participant) by time window to the correct referent (A) and incorrect referent (B) for subject-extracted *who*-questions (dark gray) and *which*-questions (light gray). Error bars represent standard error; * denotes a significant difference at the $p < .05$ level.

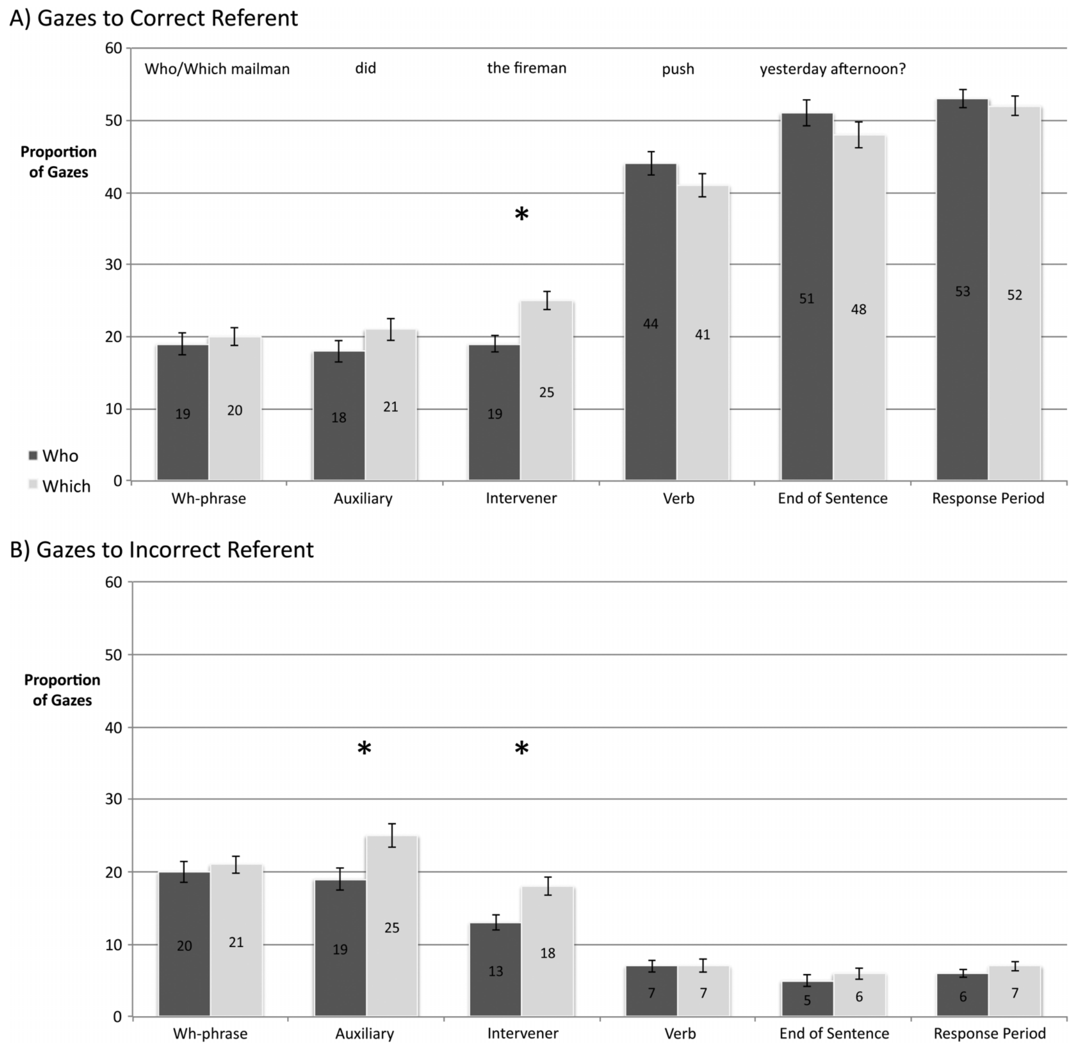


in interpreting these data. It may be that, as indicated by the high performance in all conditions, the task was too easy for our unimpaired participants and so does not discriminate the hypotheses well.

RTs are likely a more sensitive measure of processing when accuracy is at or close to ceiling. For RTs as well, there were main effects for extraction type (object-extracted slower than subject-extracted) and question type (*which*-questions slower than *who*-questions). The effect of question type suggests some support for the discourse hypothesis although this effect held only for the object-extracted sentences in pairwise comparisons, limiting support for this hypothesis. The extraction effect appears to more clearly support the word order hypothesis as here the effect of longer RTs for object-extracted questions held both for *which*-questions and for *who*-questions. Yet the object-extracted *which*-questions yielded significantly longer RTs than object-extracted *who*-questions. Only the intervener account predicts this pattern. However, the lack of a significant interaction weakens support for this hypothesis. It is also worth considering that the RT results—although perhaps more clear than the accuracy results—were also likely affected by the ease of the task for our college-age participants.

For our gaze data, we examined the proportion of gazes to the correct and incorrect referent during specified time windows as well as during the response period between the end of the sentence and the button-press response. For the subject-extracted questions, *which*-questions had increased gazes to the correct referent relative to *who*-questions at the verb and the middle NP. It's not clear what this signifies as the subject extracted *who*-questions actually had higher response accuracy than the subject-extracted *which*-questions. The two question types did not differ in gazes to the incorrect referent at any point in the sentence or response period.

Figure 3. Object-extracted sentences for Experiment 1 (unimpaired participants): Proportion of gazes for all responses (correctly and incorrectly answered by the participant) by time window to the correct referent (A) and incorrect referent (B) for object-extracted *who*-questions (dark gray) and *which*-questions (light gray). Error bars represent standard error; * denotes a significant difference at the $p < .05$ level.



For the object-extracted questions, gazes to the correct referent were greater for *which*-questions than *who*-questions only in the intervening NP time window, in which there were also more gazes to the incorrect referent for *which*-questions. This latter finding suggests greater difficulty for the *which*-questions. However, this effect starts before the intervening NP, at the auxiliary time window, although given the small number of data points contributing to this (very short) time window, we think this result should not be given much weight. However, there were no effects observed at the gap and beyond, where the intervener hypothesis would predict difficulties in computing the dependency relationship.

Thus, in this experiment, none of the three hypotheses were unambiguously supported by our data although RT patterns suggest some support for the word order hypothesis. We now move to our second experiment, in which we examine *wh*-question comprehension in a group of participants with aphasia.

Experiment 2: *Wh*-Questions in Aphasia

Our second experiment tested our hypotheses on participants with Broca's aphasia who have sentence-comprehension deficits. Here we used the same eye tracking while listening method as Experiment 1 and examined accuracy, RT, and gaze data.

Method

Participants

Eight adults with aphasia participated in the study (see Table 3). All participants experienced a single unilateral left

Table 3. Aphasia participant information.

Participant	Group	Sex	BDAE	Years poststroke	Lesion location	Age at testing (years)	Education level	SOAP: Canonical	SOAP: Noncanonical
LHD009	Broca	M	3	12	Large L lesion involving inferior frontal gyrus (BA44, 45)	52	1 year grad school	75%	55%
LHD101	Broca	M	2	6	Large L lesion involving posterior inferior frontal gyrus (BA44) with posterior extension	63	Ph.D.	95%	35%
LHD130	Broca	M	4	5	L IPL with posterior extension sparing STG	60	4 years college	75%	55%
LHD132	Broca	M	4	8	Large L lesion involving inferior frontal regions with extension to the anterior two thirds of STG & MTG	49	4 years college	85%	55%
LHD140	Broca	F	2	13	L MCA infarct secondary to occlusion of L proximal CA	38	4 years college	80%	30%
LHD138	Broca	M	2	14	L MCA infarct	35	Some college	70%	25%
LHD158	Broca	F	2	4	L CVA	56	4 years college	65%	25%
LHD159	Broca	F	3	3	L MCA infarct	60	College	100%	70%

Note. BDAE = Boston Diagnostic Aphasia Examination; L = left; BA = Brodmann area; IPL = inferior parietal lobule; STG = superior temporal gyrus; MTG = middle temporal gyrus; MCA = middle cerebral artery; CA = cerebral artery; CVA = cerebrovascular accident.

hemisphere stroke, were monolingual native speakers of English, and had normal or corrected-to-normal visual and auditory acuity. At the time of testing, all participants were neurologically and physically stable (i.e., at least 6 months postonset) with no reported history of alcohol or drug abuse, psychiatric illness, or other significant brain disorder or dysfunction.

Participants were diagnosed as having Broca's aphasia with specific sentence-comprehension deficits. Diagnosis of aphasia was based on the convergence of clinical consensus and the results of the Boston Diagnostic Aphasia Examination (version 3; Goodglass Kaplan, & Barresi, 2000). Sentence-comprehension deficits were defined as at- or below-chance performance on sentences not conforming to S-V-O order (e.g., passives and object-extracted relative clauses) along with above-chance performance on sentences with S-V-O word order (e.g., actives and subject-extracted relatives) via the SOAP Test of Sentence Comprehension (Love & Oster, 2002). Each participant was tested in four 1-hr sessions at least 1 week apart at the Language and Neuroscience Group Laboratory located at San Diego State University and was compensated $15 per session.

Design and Procedure

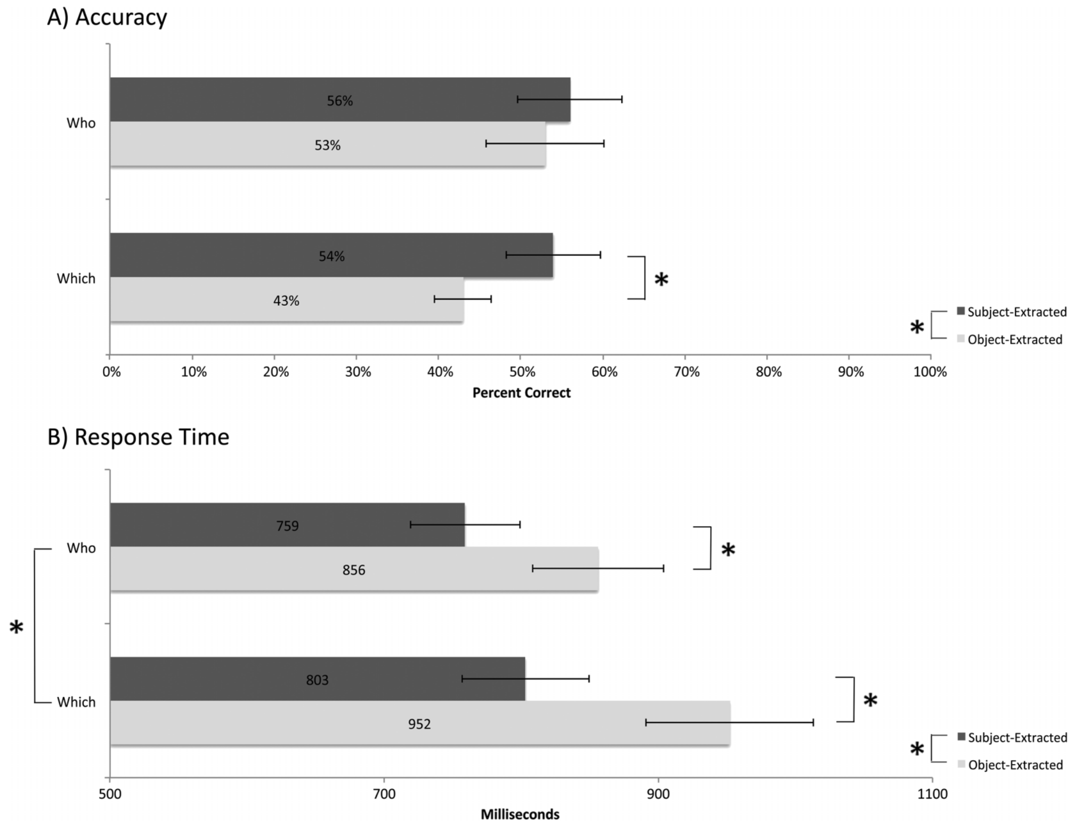
The design and procedure for Experiment 2 were nearly identical to those of Experiment 1 except that this design was fully within-subjects with participants completing all four lists across four sessions. In order to control for referent order effects, we created a duplicate set of pictures with the direction of action reversed with the action moving from right to left. In these reversed right-to-left pictures, the figure on the right was the correct choice for object-extracted questions and the figure on the left the correct choice for subject-extracted questions. In two sessions, participants saw pictures with the action moving left to right and in two sessions saw the pictures depicting the action moving right to left. In this way, each participant received every condition for each picture over the four sessions. In addition, Experiment 2 used fewer items (40 blocks each consisting of four experimental sentences for a total of 160 experimental items and 15 blocks each consisting of two filler sentences for a total of 30 filler items) and a longer response period (3 s).

Results

Offline Analysis and Data

We used the same data analysis procedures described in Experiment 1. Turning to the results from our accuracy data first (see Figure 4A), overall there was a significant main effect of extraction type: subject-extracted questions (55%) yielded more accurate performance than object-extracted questions (48%), $F(1, 1{,}196) = 7.79$, $p = .005$. A nearly significant main effect of *wh*-question type was observed: *which*-questions (48%) yielded poorer performance than *who*-questions (54.5%), $F(1, 1{,}196) = 3.67$, $p = .06$. The interaction did not reach significance, $F(1, 1{,}196) = 2.51$, $p = .11$. On further analysis, we found that accuracy for the object-extracted *which*-questions (43%) was lower than that for the object-extracted *who*-questions (53%), B = −0.43 (0.18), $t(1{,}196) = 2.46$, $p = .01$, 95% CI: (−0.78, −0.09), lower than subject-extracted *which*-questions (54%), B = −0.54 (0.18), $t(1{,}196) = 3.08$, $p =.002$, 95% CI: (−0.89, −0.20), and lower than subject-extracted *who*-questions (56%), B = −0.58 (0.18), $t(1{,}196) = 3.31$, $p = .0009$, 95% CI: (−0.93, −0.24). The latter three conditions did not differ from one another (all *p*s > .39). It is important that only object-extracted *which*-questions (43%) did not differ from chance performance (33%): $t(7) = 1.73$, $p = .13$. Note that participants pressed the middle button 12.5% of the time on average across all conditions, suggesting that they were in fact using all three buttons to respond, and so it is therefore appropriate to set chance at 33%.

Figure 4. Experiment 2 (participants with aphasia): Mean accuracy (A) and response time (B) across the four experimental conditions. Error bars represent standard error; * denotes a significant difference at the $p < .05$ level.



Consistent with this accuracy pattern, RTs (see Figure 4B) for the object-extracted *which*-questions (952 ms) were slower than for object-extracted *who*-questions (856 ms), B = 118 (62), $t(609) = 1.92$, $p = .06$, 95% CI: (−3, 239), subject-extracted *which*-questions (803 ms), B = 201 (61), $t(603) = 3.30$, $p = 0.001$, 95% CI: (81, 320), and for subject-extracted *who*-questions (759 ms), B = 249 (62), $t(601) = 4.04$, $p < .0001$, 95% CI: (128, 369). Among the latter three conditions, only object-extracted *who*-questions differed from subject-extracted *who*-questions, B = 130 (59), $t(609) = 2.22$, $p = .03$, 95% CI: (15, 246); the others did not differ from one another (ps > .15). Overall, the interaction between extraction type and *wh*-question type was not significant, $F(1, 609) = 0.69$, $p = .41$, but there were significant main effects of extraction type (object-extracted had slower RTs than subject-extracted), $F(1,602) = 15.31$, $p = .0001$, and *wh*-question type (*which*-questions had slower RTs than *who*-questions), $F(1, 599) = 3.88$, $p = .05$. RTs were screened as in Experiment 1 except using the outer fence of the box plot, removing 3.3% of data.

Gaze Analysis and Data

Gaze data were analyzed as described for Experiment 1 with the exception that all models converged when both random effects for participant and item were included, so results are reported with both random effects included.

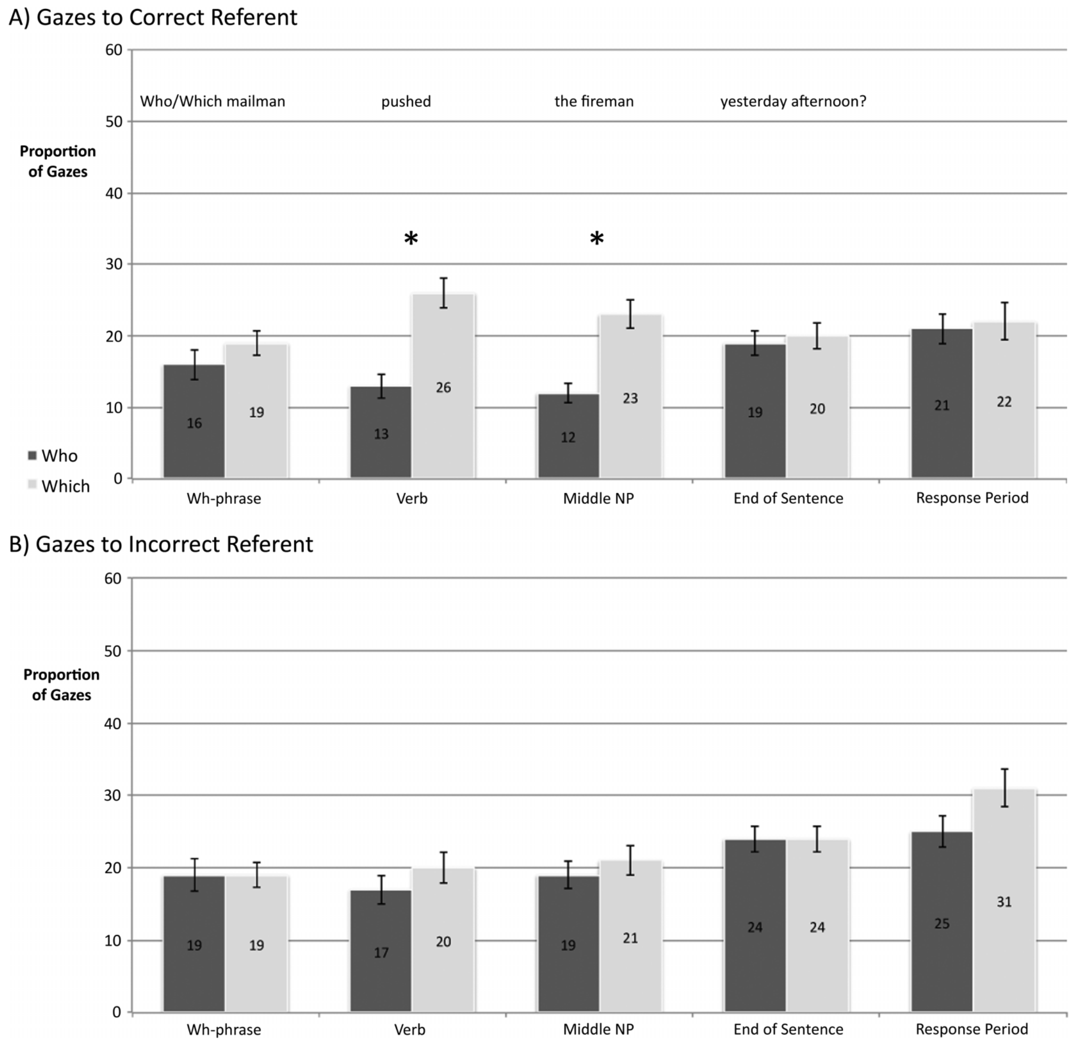
Subject-Extracted Sentences

For the subject-extracted sentences, gazes to the correct referent (see Figure 5A) revealed that *which*-questions (26%) produced a larger proportion of correct gazes than *who*-questions (13%) at the verb, B = 0.90 (0.23), $t(549) = 3.87$, $p = .0001$, 95% CI: (0.44, 1.36) and the middle NP time windows (*which*: 23%; *who*: 12%), B = 0.82 (0.24), $t(562) = 3.46$, $p = .0006$, 95% CI: (0.35, 1.28), but not at the other time windows (ps > .40). Gazes to the incorrect referent were not different for *which*- and *who*-questions at any time window (see Figure 5B; all ps > .12).

Object-Extracted Sentences

Gazes to the correct referent (see Figure 6A) revealed no significant differences in gaze proportion between the *which*- and *who*-questions for any time window (all ps > .07). However, gazes to the incorrect referent revealed an intriguing pattern (see Figure 6B). Beginning at the verb-gap region, *which*-questions (17%) yielded a higher proportion of gazes to the incorrect referent than *who*-questions (9%), B = 0.68 (0.26), $t(548) = 2.59$, $p = .01$, 95% CI: (0.16, 1.19). This difference continued through the end of the sentence (*which*: 16%; *who*: 9%), B = 0.68 (0.27), $t(570) = 2.54$, $p = .01$, 95% CI: (0.15, 1.21), and through the response period (*which*: 20%;

Figure 5. Subject-extracted sentences for Experiment 2 (participants with aphasia): Proportion of gazes for all responses (correctly and incorrectly answered by the participant) by time window for the correct referent (A) and incorrect referent (B) for subject-extracted *who*-questions (dark gray) and *which*-questions (light gray). Error bars represent standard error; * denotes a significant difference at the $p < .05$ level.

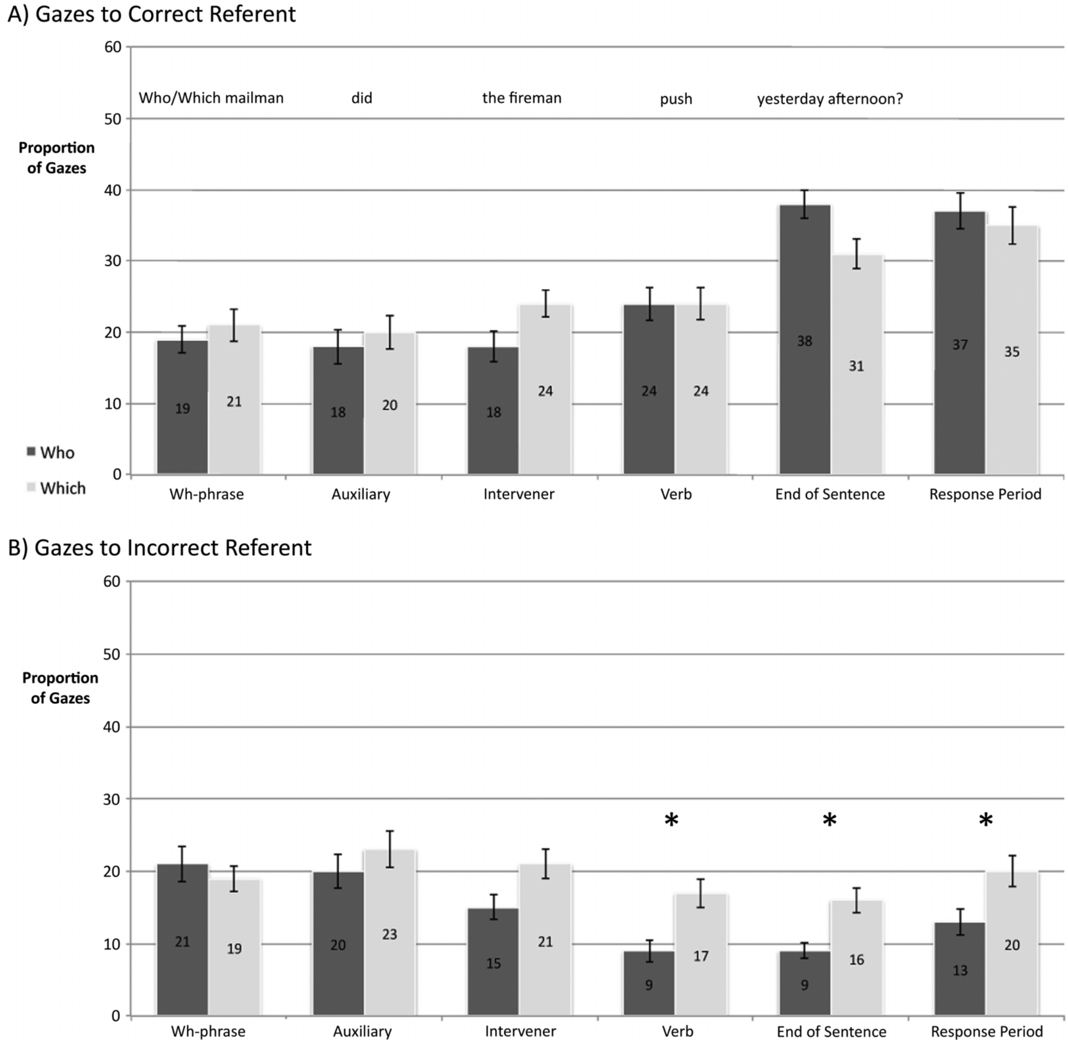


who: 13%), B = 0.53 (0.24), $t(509) = 2.17$, $p = .03$, 95% CI: (0.05, 1.01). Prior to the verb-gap region, there were no significant differences in gazes to the incorrect referent (all ps > .08).

We examined this apparent interference in more detail prospectively by analyzing if the proportion of gazes to either referent for an item predicted response accuracy for that item at any point in the sentence (see Table 4). That is, we examined if gazes to the incorrect referent (e.g., the agent-mailman) at any point in the sentence predicted a decreased likelihood of a correct response and, correspondingly, if the proportion of gazes to the correct referent (e.g., the theme-mailman) predicted an increased likelihood of a correct response. We used similar regression analyses as above with a logit link function for binary responses and crossed random effects of participant and item except that accuracy (correct vs. incorrect) was the dependent variable and proportion of gazes in the region of interest (object or subject in separate analyses) was the (continuous) fixed effect.

Two patterns are apparent (see Table 4). First, interference was evident for the *which*-sentences (black shading with white text): Increased gazes to the incorrect referent corresponded to an increased likelihood of an incorrect response. Of note, this pattern held from the verb to the end of the sentence and through the response period, exactly those time windows in which the comparison against the *who*-sentences also suggested interference. For the *who*-sentences, no such interference was found until the response period. Second, increased gazes to the correct referent (over all items) corresponded to an increased likelihood of a correct response (gray shading in Table 4), starting at the verb in the *which*-sentences (and continuing through the response period) and somewhat later for the *who*-sentences, starting

Figure 6. Object-extracted sentences for Experiment 2 (participants with aphasia): Proportion of gazes for all responses (correctly and incorrectly answered by the participant) by time window for the correct referent (A) and incorrect referent (B) for object-extracted *who*-questions (dark gray) and *which*-questions (light gray). Error bars represent standard error; * denotes a significant difference at the $p < .05$ level.



at the end of the sentence window and continuing for the response period.

Discussion

The participants with aphasia performed well on the task with above-chance accuracy for three of the four conditions. Indeed, the object-extracted *which*-question condition (43%) was the only condition that was not significantly different from chance (33%) and was significantly worse than performance on each of the other three types of questions. Accuracy in the other three conditions was not different from each other. Despite the lack of a significant interaction, this pattern is consistent with the predictions of the intervener hypothesis and is not predicted by the other two hypotheses we examined.

With respect to the other hypotheses, there was a main effect of extraction type with performance for object-extracted sentences worse than for subject-extracted sentences. This is consistent with the predictions of the word order hypothesis although in pairwise contrasts the effect was found only for *which*-questions, not for *who*-questions, limiting support for this hypothesis as no differences between question types are expected on this view. We also observed a nearly significant effect of question type with poorer performance for *which*-questions than *who*-questions, consistent with the discourse hypothesis. However, this effect was found only for the object-extracted questions in pairwise contrasts. Note that it has been suggested that object-extracted *which*-questions might be particularly problematic on the D-linking hypothesis (Avrutin, 2000, 2006). The reasoning is that more processing resources are required to fill a gap with an antecedent that is D-linked and that some

Table 4. Experiment 2 object-extracted sentences, the regression coefficient (B) with significance (*t*, *p*) for the analysis examining if gazes to the correct or incorrect referent predicted button-press response accuracy.

	Looks to correct referent			Looks to incorrect referent		
Time window	**B**	***t***	***p***	**B**	***t***	***p***
Which-sentences						
Wh-phrase	0.33	0.82	.41	0.24	0.52	.61
Auxiliary	0.34	0.98	.33	0.06	0.17	.87
Intervener	0.74	1.94	.05	–0.59	1.37	0.17
Verb	**0.92**	**2.61**	**.01**	–1.60	3.40	.0008
End of sentence	**1.21**	**3.13**	**.002**	–1.98	3.43	.0007
Response	**1.53**	**4.50**	**< .0001**	–1.81	3.94	.0001
Who-sentences						
Wh-phrase	–0.66	1.73	0.08	–0.48	1.29	.20
Auxiliary	–0.38	1.05	0.29	–0.20	0.57	.57
Intervener	–0.67	1.55	0.12	0.25	0.53	.60
Verb	–0.69	1.71	0.09	0.32	0.62	.54
End of sentence	**0.94**	**2.35**	**0.02**	–1.37	1.72	.09
Response	**1.36**	**3.62**	**0.0004**	–1.57	2.87	.004

Note. A significant positive coefficient indicates that an increased proportion of gazes predicts greater accuracy (bold); a significant negative coefficient indicates that an increased proportion of gazes predicts reduced accuracy (i.e., interference; underline).

people with aphasia have depleted resources to compute an accurate representation of such structures. Ignoring the thorny issue of what is meant by *processing resources* in this account, there is considerable linguistic and processing evidence that even subject-extracted relative clauses and *wh*-questions contain a copy/trace of a displaced (subject) NP (see, for example, Zurif, Swinney, Prather, Solomon, & Bushell, 1993) and thus there is no theoretical reason for the D-linking hypothesis to distinguish subject- from object-extraction.

In terms of our RT data, the object-extracted *which*-questions (952 ms) were significantly slower than the RTs for the subject-extracted *who*- (759 ms) and *which*- (803 ms) questions and the object-extracted *who*-questions (856 ms). Thus, the RT patterns basically conformed to the accuracy data described above with RTs for object-extracted *which*-questions significantly slower than for the other three conditions. Again, this pattern is only predicted by the intervener hypothesis and not the other accounts, albeit again without a significant interaction. There was a significant main effect of extraction type (object-extracted slower than subject-extracted), which held in pairwise contrasts both for the *which*-questions and for the *who*-questions, consistent with the word order hypothesis. Yet the worse performance for object-extracted *which*-questions relative to the object-extracted *who*-questions is not expected on this view. We also found a main effect of *wh*-type (*which*-questions slower than *who*-questions), consistent with the discourse hypothesis although pairwise comparisons revealed that this pattern only held for the object-extracted questions. Therefore, although there may be some support for the word order and discourse hypotheses, the observation that the object-extracted *which*-questions clearly stand out from the other three question types provides stronger support for the intervener hypothesis.

For the online gaze data, we compared the proportion of gazes to the correct and incorrect referents within extraction type; the two subject-extracted conditions were compared to each other, and the two object-extracted conditions were compared to each other. For the subject-extracted sentences, the patients showed precisely the same pattern as the unimpaired controls did in Experiment 1. There was an increase in gazes to the correct referent for *which*-questions relative to *who*-questions during the verb and middle NP time windows. This pattern suggests that the participants with aphasia are processing these sentences similarly to the control participants even if it may not be entirely clear precisely what kind of process this signifies (see discussion to Experiment 1). Moreover, as with the control participants, there were no differences between *which*-questions and *who*-questions with respect to gazes to the incorrect referent.

For the object-extracted sentences, a clearer pattern emerges. There were no differences between *which*-questions and *who*-questions in gazes to the correct referent, but unlike what was observed for our control participants in Experiment 1, here we observed consistently more gazes to the incorrect referent for *which*-questions from the verb-gap region through the end of the sentence and the response period. Moreover, the gaze behavior through these regions predicted response accuracy for the *which*-questions but not for the *who*-questions, which showed only a more restricted relationship between gaze location and accuracy.

This pattern in the gaze data suggests that only the object-extracted *which*-questions were problematic for the participants with aphasia to process, consistent with the predictions of the intervener hypothesis but not the other hypotheses. Our results are also consistent with those of prior offline studies. Hickok and Avrutin (1996) found a subject-object asymmetry only for *which*- and not for *who*-questions. Friedmann and Gvion (2012) report results consistent with an intervener effect for object-extracted relative clauses. In addition, Dickey et al. (2007) found on-time gazes at the gap for object-extracted *who*-questions, suggesting that their participants with Broca's aphasia were able to process these sentences. The present study corroborates this finding. Yet we have gone further and observed that not all *wh*-questions are treated similarly. The patients with Broca's aphasia who also have comprehension deficits in the present study had significantly more offline and online difficulty with questions containing an intervener (object-extracted *which*-questions) compared to those questions that did not (object-extracted *who*-questions, subject-extracted *wh*-questions).

General Discussion

We argue that the results from our participants with aphasia support the intervener hypothesis. In terms of the *wh*-questions we investigated, an intervener is a fully specified NP that occurs between a gap and its displaced *wh*-phrase.

The intervener interferes with computing the dependency relationship because it is a possible element in the dependency chain, rendering a processing disadvantage to such structures over those that don't contain an intervener. We suggest that some adults with a language disorder are particularly vulnerable to interveners during sentence processing, perhaps because they are susceptible to interference among similarly structured NPs.

To be sure, there are several unanswered questions that are raised by our findings. One question is whether the intervener hypothesis is specific to individuals with aphasia. Our results suggest this may be the case as only Experiment 2 found relatively strong support for the account in terms of accuracy, RTs, and gaze data. Yet the theoretical basis for the account comes not only from the linguistic literature (i.e., relativized minimality), but also from the processing literature using neurologically intact adult participants. For example, Friedmann and Gvion (2012) specifically couched their intervener results in terms of a linguistic deficit. In their view, participants with agrammatic Broca's aphasia cannot construct fully realized syntactic trees, and thus, locality comes to the rescue if the syntax cannot provide a structure to connect a verb to its (displaced) arguments. In terms of processing, Gordon and colleagues (Gordon, Hendrick, & Johnson, 2004; Gordon, Hendrick, Johnson, et al., 2006) have suggested a similarity-based interference account of normal memory, in which the demands on storage and retrieval during sentence comprehension are increased when there are NPs that have similar representations. The idea is that interference is mediated by a direct-access retrieval mechanism that is sensitive to different cues, including semantic, pragmatic, and syntactic ones. Because we found intervener effects only for our participants with aphasia, perhaps their linguistic deficit stems from an increased sensitivity to similarity because of memory processing limitations.

An additional account, the trace-deletion hypothesis (TDH; Grodzinsky, 1995; Grodzinsky & Finkel, 1998), bears mention here. The TDH claims that individuals with Broca's aphasia delete traces in syntactic representations, leaving displaced arguments without a grammatically specified thematic role. A nonlinguistic agent-first heuristic (assume that the first NP in a sentence is the agent) is used to interpret an argument that is left without a grammatically specified thematic role. Skipping details, this heuristic leads to chance performance on offline sentence–picture matching tasks with sentences containing displaced direct object arguments. However, because the TDH suggests that only referential NPs (those that refer to an individual from a set of individuals) are input to the agent-first heuristic, only questions headed by a *which*-phrase should be affected. The TDH therefore makes the same predictions as the intervener hypothesis for offline measures, namely that object-extracted *which*-questions should have the lowest accuracy. Even so, the TDH requires both the deletion of traces and the use of a nonlinguistic heuristic to explain offline patterns, and the intervener account suggests a single mechanism: interference from similarly structured NPs. Furthermore, the TDH has relied solely on offline measures; online predictions from eye tracking for the TDH are not clear. Thus our view is that the intervener hypothesis may be a simpler account of sentence comprehension performance in Broca's aphasia relative to the TDH, and this is also buttressed by support from our online eye gaze measurements. Further experimentation can resolve this issue by examining other constructions that do (and do not) contain interveners and on which the two accounts make distinct predictions.

Last, there is the possibility of a general language-processing impairment, such as that proposed by a delayed lexical access account (Ferrill, Love, Walenski, & Shapiro, 2012; Love, Swinney, Walenski, & Zurif, 2008) or a delayed syntactic processing account (Avrutin, 2006; Burkhardt, Avrutin, Piñango, Ruigendijk, 2008; Piñango, 2000). The delayed lexical access hypothesis predicts that patients with Broca's aphasia have delayed gap filling due to delayed access to lexical information. The delayed syntactic processing account likewise predicts slower-than-normal gap filling due to slowed syntactic processing. However, because we observed clear differences between the object-extracted *who*- and *which*-questions, both requiring gap filling, we did not find support for either of these hypotheses.

We end our discussion with a note on variability. As is well known in the literature, variability in behavior across participants with aphasia is a continuing concern and often interferes with interpreting the results from different studies. In particular, this issue has often targeted the syndrome of Broca's aphasia (see, for example, Drai & Grodzinsky, 2006). Our tactic here and in other studies is to select our participants with aphasia on the basis of specific theoretical issues. The issue that we addressed in this study is about the underlying deficit in comprehension in aphasia. In the present work, then, we selected our participants on the basis of their comprehension profiles, including only participants with Broca's aphasia on standard testing who also revealed at- or below-chance performance on noncanonically ordered sentences (e.g., passives and object-extracted relatives) relative to good performance on S-V-O ordered sentences (e.g., actives and subject-extracted relatives). It remains to be seen whether or not other types of aphasia—those that also involve a comprehension deficit—would reveal similar results.

To conclude, the present study found strong evidence to support the intervener hypothesis of sentence comprehension in aphasia. If this work is confirmed and extended (e.g., to other intervener-type constructions), treatment programs could be developed that focus on the similarity of NPs in sentences that yield good versus poor comprehension. Thus, this line of research could have far-reaching implications and benefits for patients with aphasia.

Acknowledgments

The work reported in this article was supported by National Institute on Deafness and Other Communication Disorders Grants R01DC009272 (awarded to Tracy Love and Lewis P. Shapiro), R01DC000494 (awarded to Lewis P. Shapiro), and T32DC007361 (awarded to Lewis P. Shapiro). We thank our research assistants and our participants and their families for their time.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

Avrutin, S. (2000). Comprehension of discourse-linked and non-discourse-linked questions by children and Broca's aphasics. In Y. Grodzinsky, L. P. Shapiro, & D. Swinney (Eds.), *Language and the brain: Representation and processing* (pp. 295–313), San Diego, CA: Academic Press.

Avrutin, S. (2006). Weak syntax. In Y. Grodzinsky & K. Amunts (Eds.), *Broca's region* (pp. 49–62). Oxford, UK: Oxford University Press.

Baayen, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers, 1*(1), 1–47.

Baayen, R. H. (2008). *Analyzing linguistic data* (Vol. 505). Cambridge, UK: Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278.

Bates, E., Friederici, A., & Wulfeck, B. (1987). Comprehension in aphasia: A cross-linguistic study. *Brain and Language, 32,* 19–67.

Burkhardt, P. (2005). *The syntax-discourse interface: Representing and interpreting dependency* (Vol. 80). Philadelphia, PA: John Benjamins.

Burkhardt, P., Avrutin, S., Piñango, M. M., & Ruigendijk, E. (2008). Slower-than-normal syntactic processing in agrammatic Broca's aphasia: Evidence from Dutch. *Journal of Neurolinguistics, 21,* 120–137.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language, 3,* 572–582.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, the Netherlands: Foris.

Chomsky, N. (1995). *The minimalist program* (Vol. 28). Cambridge, MA: MIT Press.

De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian: The minimal chain principle* (Vol. 12). Dordrecht, the Netherlands: Springer.

Dickey, M., Choy, J. W., & Thompson, C. K. (2007). Real-time comprehension of *wh*-movement in aphasia: Evidence from eye tracking while listening. *Brain and Language, 100,* 1–22.

Dickey, M., & Thompson, C. (2009). Automatic processing of *wh*- and NP-movement in agrammatic aphasia: Evidence from eye tracking. *Journal of Neurolinguistics, 22,* 563–583.

Donkers, J., Hoeks, J., & Stowe, L. (2013). D-linking or set-restriction? Processing *which*-questions in Dutch. *Language and Cognitive Processes, 28,* 9–28.

Donkers, J., & Stowe, L. (2006). *Wh-questions and the nature of D-linking: A processing perspective*. Presented at the International Conference on Linguistic Evidence, University of Tübingen, Germany.

Drai, D., & Grodzinsky, Y. (2006). A new empirical angle on the variability debate: Quantitative neurosyntactic analyses of a large data set from Broca's Aphasia. *Brain and Language, 96,* 117–128.

Ferrill, M., Love, T., Walenski, M., & Shapiro, L. P. (2012). The time-course of lexical activation during sentence comprehension in people with aphasia. *American Journal of Speech-Language Pathology, 21,* S179–S189.

Frazier, L., & Clifton, C. (2002). Processing "d-linked" phrases. *Journal of Psycholinguistic Research, 31,* 633–659.

Frazier, L., & McNamara, P. (1995). Favor referential representation. *Brain and Language, 49,* 224–240.

Frazier, L., Plunkett, B., & Clifton, C. (1996). Reconstruction and scope. *University of Massachusetts Occasional Papers in Linguistics, 19,* 239–260.

Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua, 119,* 67–88.

Friedmann, N., & Gvion, A. (2012). Intervention and locality in agrammatic aphasia. In *Enjoy Linguistics! Papers Offered to Luigi Rizzi on the Occasion of His 60th Birthday* (pp. 1–13). Siena, Italy: CISCLWorking Papers.

Friedmann, N., & Shapiro, L. P. (2003). Agrammatic comprehension of simple active sentences with moved constituents: Hebrew OSV and OVS structures. *Journal of Speech, Language, and Hearing Research, 46,* 288–297.

Goodglass, H., Kaplan, E., & Barresi, B. (2000). *Boston Diagnostic Aphasia Examination (3rd Ed.)*. Philadelphia, PA: Taylor & Francis.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51,* 97–114.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1304–1321.

Grillo, N. (2005). Minimality effects in agrammatic comprehension. In S. Blaho, L. Vicente, & E. Schoorlemmer (Eds.), *Proceedings of ConSOLE XIII* (pp. 106–120). Leiden, the Netherlands: Leiden University.

Grillo, N. (2009). Generalized minimality: Feature impoverishment and comprehension deficits in agrammatism. *Lingua, 119,* 1426–1443.

Grodzinsky, Y. (1990). *Theoretical perspectives on language deficits*. Cambridge, MA: The MIT Press.

Grodzinsky, Y. (1995). A restrictive theory of agrammatic comprehension. *Brain and Language, 50,* 27–51.

Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences, 23,* 1–71.

Grodzinsky, Y., & Finkel, L. (1998). The neurology of empty categories: Aphasic's failure to detect ungrammaticality. *Journal of Cognitive Neuroscience, 10,* 281–292.

Hickok, G., & Avrutin, S. (1996). Comprehension of *wh*-questions in two Broca's aphasics. *Brain and Language, 52,* 314–327.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS for linear mixed models* (2nd ed.) Cary, NC: SAS Institute.

Love, T., & Oster, E. (2002). On the categorization of aphasic typologies: The SOAP (a test of syntactic complexity). *Journal of Psycholinguistic Research, 31,* 503–529.

Love, T., Swinney, D., Walenski, M., & Zurif, E. (2008). How left inferior frontal cortex participates in syntactic processing: Evidence from aphasia. *Brain and Language. 107,* 203–219.

Manor, B. R., & Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of Neuroscience Methods, 128,* 85–93.

O'Grady, W., Yamashita, Y., & Lee, S. Y. (2005). A note on canonical word order. *Applied Linguistics, 26,* 453–458.

Piñango, M. M. (2000). On the proper generalization for Broca's aphasia comprehension pattern: Why argument movement may not be at the source of the Broca's deficit. *Behavioral and Brain Sciences, 23,* 48–49.

Rizzi, L. (1990). *Relativized minimality*. Cambridge, MA: MIT Press.

Rothman, J., & Slabakova, R. (2011). The mind-context divide: On acquisition at the linguistic interfaces. *Lingua, 121,* 568–576.
Salis, C., & Edwards, S. (2005). Discourse linking, canonicity, and comprehension of *wh*-questions in agrammatism. *Brain and Language, 96,* 86–87.
Shapiro, L. P. (2000). Some recent investigations of gap filling in normal listeners: Implications for normal and disordered language processing. In Y. Grodzinsky, L. P. Shapiro, & D. Swinney (Eds.), *Language and the brain: Representation and processing* (pp. 357–376). San Diego, CA: Academic Press.
Thompson, C. K., Tait, M. E., Ballard, K. J., & Fix, S. C. (1999). Agrammatic aphasic subjects' comprehension of subject and object extracted *wh*-questions. *Brain and Language, 67,* 169–187.
Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 407–430.
Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language, 65,* 247–263.
Zurif, E., Swinney, D., Prather, P., Solomon, J., & Bushell, C. (1993). An on-line analysis of syntactic processing in Broca's and Wernicke's aphasia. *Brain and Language, 45,* 448–464.