JSLHR

Research Article

Processing Binding Relations in Specific Language Impairment

Richard G. Schwartz,[a] Arild Hestvik,[b] Liat Seiger-Gardner,[a,c] and Diana Almodovar[c]

Purpose: This sentence processing experiment examined the abilities of children with specific language impairment (SLI) and children with typical language development (TD) to establish relations between pronouns or reflexives and their antecedents in real time.
Method: Twenty-two children with SLI and 24 age-matched children with TD (7;3–10;11 [years;months]) participated in a cross-modal picture priming experiment to determine whether they selectively activated the correct referent at the pronoun or reflexive in sentences. Triplets of auditory sentences, identical except for the presence of a pronoun, a reflexive, or a noun phrase along with a picture probe were used.
Results: The children with TD were slightly more accurate in their animacy judgments of pictures, but the groups exhibited the same reaction time (RT) pattern. Both groups were slower for sentences with pronouns than with reflexives or noun phrases. The children with SLI had longer RTs than their peers with TD.
Conclusions: Children with SLI activated only the appropriate antecedent at the pronoun or reflexive, reflecting intact core knowledge of binding as was true for their TD peers. The overall slower RT for children with SLI suggests that any deficit may be the result of processing deficits, perhaps attributable to interference effects.

There is a continuing controversy as to whether specific language impairment (SLI) is the result of a deficit in underlying linguistic representation or operations, a deficit in language-specific processing, or a general processing deficit (e.g., see reviews in Bishop, 1997; Leonard, 2014; Schwartz, 2009). The study of antecedent assignment to pronouns, referred to as *binding*, in typically developing (TD) children and in children with SLI, illustrates this controversy.

Binding theory, first formally described within Government and Binding Theory (N. Chomsky, 1993), is the set of syntactic constraints on the reference of pronouns, reflexives, and nouns/names that entails three principles or conditions. The *syntactic constraints* are specified in terms of the structural properties and relations in the syntactic tree structure. Principle A states that an *anaphor* (i.e., a reflexive or reciprocal) must be locally bound within its clause. For example, in the sentence, *The mother says the girl splashed herself*, the reflexive must refer to *the girl*. Principle B states that a pronoun must be *locally free* (i.e., it cannot refer to the noun or noun phrase [NP] subject in the same clause). In the sentence, *The mother says that the girl splashed her*, the pronoun must refer to *mother*, but in the sentence *Max splashed him*, the pronoun cannot refer to *Max*. Principle C states that *R-expressions* (names such as *Brianna*, and NPs such as *the dog*) must be free. For example, in *She asked the mother to splash the girl*, the antecedent of *she* cannot be *the girl* and in sentences such as *She saw Brianna in the pool,* the antecedent of *she* cannot be *Brianna*.

These binding principles are considered to be part of the innate knowledge of language (Chien & Wexler, 1990; C. Chomsky, 1969; Lust, 1986; McDaniel, Cairns, & Hsu, 1990). Studies using a variety of methods including children's acting out tasks, truth-value judgments of auditory sentences (in relation to acting out by adults or to pictures), or grammaticality judgments suggest that most young children reliably follow Principles A (reflexives) and C (names and nouns), but appear to violate Principle B (pronouns) with referential NPs (*The girl splashed her*), but not with quantified NPs (e.g., *Every girl splashed her*) until they achieve full mastery somewhat later in development. This has been referred to as *quantificational asymmetry* (Elbourne, 2005).

The findings and the explanations offered for quantificational asymmetry and the acquisition of binding theory

[a]The Graduate Center, City University of New York
[b]Department of Linguistics and Cognitive Science, University of Delaware, Newark
[c]Speech-Language-Hearing Science, Lehman College, Bronx, NY
Correspondence to Richard G. Schwartz: rschwartz@gc.cuny.edu

Editor: Rhea Paul
Associate Editor: Lizbeth Finestack
Received March 18, 2015
Revision received August 27, 2015
Accepted February 5, 2016
DOI: 10.1044/2016_JSLHR-L-15-0107

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

in general have been and continue to be quite mixed. The explanations have included a pragmatic principle (Principle P, which precludes coreference) that young children have not mastered, the *lexical learning hypothesis* (e.g., Chien & Wexler, 1990; Reinhart, 1983, 1986), a combination of plausibility and relative character salience in a story (Elbourn, 2005), and a processing limitation account (Solan, 1987). Nevertheless, Principles A, B, and C remain a useful framework for characterizing issues in development.

This literature is characterized by numerous methodological issues (see extensive discussions in Conroy, Takahasi, Lidz, & Phillips, 2009; Elbourne, 2005; Grimshaw & Rosen, 1990). In *truth-value judgment tasks*, the experimenter acts out a scene, a puppet or the experimenter makes a statement, and the child judges the statement as true or false in relation to the acted out scene. These tasks may only reveal preferences. These preferences are simply children's tendencies to accept the truth of the sentence, especially if the ungrammatical reading is made plausible by the context. The bases for grammaticality or truth-value judgments provided by young children are often undetermined even when investigators assume that a context is unambiguous. In some cases, the stimuli themselves may have led children to incorrect anaphoric interpretations. Grimshaw and Rosen (1990) noted that in two-picture studies (e.g., the child chooses one of two pictures in response to a sentence), children's errors on sentences involving possessives (*Lindsay's sister kissed her/herself*) might reflect a misunderstanding of the possessive rather than binding errors. Four-picture studies (i.e., the child makes a choice from four pictures in response to an auditory sentence) can distinguish binding and possessive interpretation errors. In their own study, Grimshaw and Rosen concluded that children do know the rules, but the equivalency of performance on reflexives and pronouns remains in question and subject to the challenges of specific experiments. In a more detailed critique of truth-value judgment tasks, which are the basis for most studies of binding, Elbourn (2005) argued that it is possible to create truth-value judgment tasks that are not subject to many of these biases and limitations. With better experimental materials, children perform accurately with pronouns regardless of whether the antecedents are referential or quantified and, when stimulus conditions from previous studies are introduced, the referential/quantified NP differences appear (Conroy et al., 2009). Additional issues include extensive evidence that young children do obey Principle B and similarly extensive evidence that they do not (Elbourne, 2005). Children's responses also vary with the specific pronoun, but do not conform to verb frequency predictions, supporting neither nativist nor usage-based accounts and leading back to processing explanations (Matthews, Lieven, Theakston, & Tomasello, 2009). We have provided only a brief and superficial overview of binding theory and its acquisition, especially of the theoretical and methodological issues controversy surrounding the quantificational asymmetry, as the current study focuses on binding theory more narrowly. More extensive and detailed discussions can be found in the references cited.

Online Studies of Binding

Although researchers have made inferences about processing from postcomprehension tasks, few studies have used sentence processing paradigms to examine binding. Love and Swinney (1996) used a cross-modal lexical decision task to determine whether reflexives and pronouns differentially reactivated potential antecedents in sentences such as:

1. a. *The boxer told the skier that the doctor for the team$_1$ would blame himself$_2$ for the injury.*
1. b. *The boxer told the skier that the doctor for the team$_1$ would blame him$_2$ for the injury.*

While participants listened to the sentence, a word or nonword appeared on the screen at either Probe Point 1 or Probe Point 2. They had to press one of two buttons to indicate whether they saw a word or nonword. The words were related or unrelated to one of the potential antecedents. The presence or absence of priming effects (reduced RT compared to an unrelated probe) for one of the two related probes at the pronoun or reflexive indicated what antecedent was activated at the reflexive or pronoun. At *himself* only the antecedent allowed by binding theory (i.e., *doctor*) was reactivated, whereas at *him*, only *the boxer* and *the skier* were reactivated. These findings indicated that binding relations are established virtually instantaneously and automatically upon encountering the anaphor or pronoun, and that the parser directly incorporates the information provided by binding theory to regulate sentence processing.

With pictures instead of orthographic probes, a similar study (McKee, Nicol, & McDaniel, 1993) examined binding in adults and in TD children (4;1–6;4). The stimuli were triplets of sentences that differed only in whether the embedded object was a pronoun, a reflexive, or a referential NP (e.g., *The alligator knows that the leopard with the green eyes is patting him/himself/the boy on the head with a soft pillow*). The picture for that triplet was LEOPARD. Each sentence in the triplet was presented once with the same picture each time, depicting the embedded clause subject (*a leopard*) at the offset of the embedded object (reflexive, pronoun, or noun). The participant's task was to decide whether the picture was animate or not. The prediction was that a reflexive should trigger reactivation of the embedded clause subject and prime the picture (leading to a faster reaction time [RT]), whereas the picture judgment at the pronoun would be slower because there is no reactivation of the antecedent. The noun (*the boy*) condition served as a baseline.

For adults, the mean RT at reflexives was faster than the baseline, the mean RT at pronouns was not faster than the baseline, and reflexives were faster than pronouns. In children, picture judgments in the reflexive condition were significantly faster than in the noun condition, whereas picture judgments in the pronoun condition were not faster than in the noun condition. There was no difference between the reflexive and pronoun condition. A subsequent truth-value judgment task (on the basis of Chien & Wexler, 1990), with all its inherent limitations, allowed them to divide their participants into two groups—one that exhibited adult-like performance on the judgment task and one that performed more poorly. In a reexamination of the RT data, the first subgroup exhibited the expected difference between pronouns and reflexives, whereas the other group did not exhibit the RT differential. Combined with the offline data from this latter group, McKee et al. (1993) concluded that these children exhibited delay of Principle B.

Binding in SLI

Franks and Connell (1996) examined reflexive comprehension in a descriptive study of 20 adults, 13 TD children (3;7–8;6), and 11 SLI children (3;9–7;8). Participants viewed video clips and were asked yes/no questions about them. For example, Mickey Mouse pours a glass of juice and says it is for Bugs Bunny. Bugs doesn't hear and says to Ernie, *Was that juice poured for me?* Ernie then says, *Yes, it's for you* and Bugs drinks the juice. The experimenter then asked questions such as *Did Bugs ask Ernie if Mickey poured juice for himself?* The expected answer was *no* and the adults provided that response consistently. The study is beset by methodological issues and lacked statistical analysis. Nevertheless, SLI children showed a preference for the nearest possible antecedent (i.e., the subject or object in the same clause as the reflexive), whereas the TD children were more variable in allowing long-distance binding interpretations (ungrammatical in English, but acceptable in languages such as Chinese, Icelandic, and Russian). Children with SLI used a strategy that minimized the distance between the reflexive and the antecedent when interpreting sentences. This could be consistent with a nonstructural process for interpreting sentences with binding relations—a lack of binding principles—or with a lack of structural analysis.

The other study of binding in SLI (van der Lely & Stollwerck, 1997) used a picture verification task to examine whether the interpretation of sentences with reflexives and pronouns by children with SLI followed binding principles. The participants were 12 children with grammatical SLI (9;3–12;10) and three control groups (12 children in each) of TD children—one group (5;5–6;4) matched on the Test of Reception of Grammar (TROG; Bishop, 1983) and the Grammatical Closure subtest of the Illinois Test of Psycholinguistic Abilities (Kirk, McArthy, & Kirk, 1968), one group (6;5–7;4) matched on expressive and receptive vocabulary scores, and one group (7;5–8;9) matched only for expressive vocabulary scores. The receptive vocabulary scores actually overlapped in these latter two groups. The children with SLI exhibited severe language impairments (with two exceptions who scored within 1.5 *SD*s of the mean); they scored at least 1.5 *SD*s below the mean on the TROG, up to 2.5 *SD*s below the mean on TROG, and as many as 5.5 *SD*s below the mean on the Grammatical Closure subtest.

In the first experiment, the experimenter showed the child a picture (e.g., Mowgli tickling Baloo Bear) along with a context sentence followed by a question (e.g., *This is Mowgli; this is Baloo Bear. Is Mowgli tickling himself?*). The antecedent was a quantifier (e.g., *every*) or a referential NP (e.g., *the bear*), and either a pronoun or a reflexive. In the match condition (picture matches the correct sentence interpretation), all four groups performed at ceiling on the name-pronoun, quantifier-pronoun conditions. In the name-pronoun condition, the children with SLI dropped down to slightly over 90% correct, whereas the controls were all at or near 100%. Any significant difference here was most certainly an artifact of the ceiling effects. In the quantifier-reflexive condition, the children with SLI dropped to approximately 75% correct (well above chance), whereas the control groups ranged from 90% to 100% correct. Even though their average percent correct was lower, the average score was 4.5 out of 6 items with a *SD* of 1.45. In the mismatch conditions, typically thought to be more difficult because they involve recognizing a violation, children with SLI did drop to 65% correct (still above chance; mean score 3.83 out of 6 with a *SD* of 1.70) on name-pronouns, and to 55% (chance; mean score 3.25 with a *SD* of 1.76) on quantifier-reflexives, whereas the control groups were between 80% and 90% correct and between 70% and 90% correct, respectively. All groups exhibited ceiling effects on quantifier-pronoun and name-reflexive items. The fact that the maximum score was 6 in each match and mismatch condition, with high standard deviations in many conditions, further complicates the interpretation of the findings. The use of percentages with such small numbers of trials distorts the findings in their figures. Although van der Lely and Stollwerck (1997) concluded that the performance of SLI children was due to their failure to interpret sentences with quantifiers correctly, this is not supported by the data.

In their second experiment, the sentences were embedded in a main clause to provide an additional sentence-internal potential antecedent (e.g., *Mowgli says Baloo Bear is tickling himself*). The children with SLI again showed the poorest performance (M = 2.75, SD = 1.96) on sentences with reflexives that had quantified antecedents in the mismatch condition (*Mowgli says every monkey is tickling himself*, matched to either a picture of every monkey tickling Mowgli, or every monkey performing a self-oriented action). They also performed more poorly than one or more of the control groups on name-reflexive syntax (M = 3.00, SD = 1.86) and quantifier-reflexive syntax (M = 3.75, SD = 1.86). These findings are subject to the same issues as their Experiment 1 findings. The findings and conclusions are mitigated by the same ceiling effects, the high standard deviations, the distortion introduced by using percentages, the task issues, the lower scores for children with SLI that remained well above chance, the separate scoring of matches and mismatches, and the fact that the differences were only in mismatches. The overlap in the control groups remains the same as in their Experiment 1. Both experiments are subject to the issues described in the previous section.

Despite these equivocal findings, van der Lely and Stollwerck (1997) interpreted their results as showing that "SLI children… performed at chance when syntactic information was required to rule out inappropriate coreference" (p. 275). It's not clear that the other conditions could be interpreted correctly with only semantic information. Furthermore, they did not include match trials for these conditions. Given the small number of items and relatively large standard deviations, many of the children with SLI performed above chance. It is also difficult to see how their data led to their conclusion that SLI children "do not have the syntactic knowledge characterized by BT" (binding theory), and that "their syntactic representation appears to be 'underspecified' with respect to coindexation between constituents" (p. 276). Joanisse and Seidenberg (2003) also disputed this interpretation, pointing out that the SLI children's performance was above chance (but still below the TD group's performance). In conclusion, it does not seem that the pattern of differences from the TD groups fit into a simple *having or not having knowledge of binding theory* explanation. In particular, there is no simple coherent theoretical way to model the SLI children's problems with quantifier-reflexives sentences, but perfect performance on quantifier-pronoun sentences.

In summary, these two judgment task experiments do not converge on a coherent picture of SLI children's knowledge of binding conditions. At best, they might indicate that SLI children rely less on structural representations when they interpret sentences with pronouns and reflexives. If so, SLI children would be expected to differ from TD children in their online processing of sentences.

Current Study

We used a cross-modal picture priming paradigm (McKee et al., 1993) to test whether SLI children exhibit knowledge of binding theory during online processing and to compare their performance to that of their age-matched TD peers. As in McKee at al., we used triplets of auditory sentences differing only in the presence of a pronoun, a reflexive, or a NP (e.g., *The alligator knows that the leopard with the green eyes is patting him/himself/the girl on the head with a soft pillow*). For each triplet, the same picture probe was used for all three sentences (e.g., LEOPARD) at the point of the pronoun, reflexive, or NP. For TD children, we expected that their RT for picture probe animacy judgments in sentences with reflexives would be faster than their RT for sentences with pronouns because the probe matches the antecedent activated for reflexives but not for pronouns. The NP RT was intended as a baseline.

Method

Participants

A total of 58 children were recruited for the study from New York City area schools, clinics, and private practice speech-language pathologists (SLPs). Nine participants failed to complete the full protocol, either by missing one of the three experimental sessions or failing to complete the full standardized testing. One participant was excluded because of experimenter error. Of the remaining 48 children, 26 were TD and 23 had been previously diagnosed with SLI (7;3–10;11). After the completion of the experiment, one SLI participant was excluded because he had only 9% valid responses (within the time limit established) and because most of the remaining RTs were greater than 2 s; such slow responses were not typical of other SLI participants. One TD participant was also excluded because of chance performance on probe animacy decisions. Thus, 46 children were included in the final analysis: 22 children with SLI and 24 children with TD (see Table 1). The TD group included 13 girls and 11 boys; the SLI group included nine girls and 13 boys. The children in both groups were drawn from a variety of socioeconomic status (SES) backgrounds (Hollingshead, 1975). Among the TD children, seven were from upper SES families, nine from upper middle-class families, seven from middle-class families, and one from lower middle-class families (two declined to specify SES). In the SLI group, one child was from an upper SES family, seven from upper middle SES families, five from middle SES families, four from lower middle SES families, and four from lower SES (one declined to specify SES). Thus, the SES distribution was slightly skewed toward upper SES for the children with TD group and slightly toward lower SES for the children with SLI. The racial and ethnic composition of the groups was as follows. The SLI group included 12 children who were White and non-Hispanic, four children who were White and Hispanic, three children who were Black and non-Hispanic, two children who were Black and Hispanic, and one child who was Native American/Alaskan and Hispanic. The TD group included 17 children who were White and non-Hispanic, four children who were Black and non-Hispanic, two children who were White and non-Hispanic, and one child who was Black and Hispanic. Three of the TD children were left-handed and four of the SLI children were left-handed.

All children passed a standard hearing screening at 20 db (500, 1000, 2000, and 4000 Hz), and all had normal articulation and phonological skills in conversational speech, as judged by an SLP. The children included did not

Table 1. Participant descriptors and scores.

Descriptors	TD (*n* = 24)	SLI (*n* = 22)
Age	9;4 (0;9/7;6–10;8)	8;7 (0;9/7;11–10;8)
CELF-R	110 (13/88–37)	82 (11/63–103)*
CELF-E	114 (12/95–140)	75 (13/53–102)*
CELF	112 (11/94–136)	75 (11/50–93)*
PPVT-IV	108 (15/86–150)	87 (11/72–106)*
TONI	108 (16/87–142)	99 (12/83–123)**

Note. All descriptors except age are mean standard scores (distribution with 100 as a mean and 15 as a standard deviation) for groups with (*SD*/range). TD = children with typical language development; SLI = children with specific language impairment; CELF-R = Clinical Evaluation of Language Fundamentals–Fourth Edition, Receptive Language Score; CELF-E = Clinical Evaluation of Language Fundamentals–Fourth Edition, Expressive Language Score; CELF = Clinical Evaluation of Language Fundamentals–Fourth Edition, Core Language Score; PPVT-IV = Peabody Picture Vocabulary Test–Fourth Edition; TONI = Test of Nonverbal Intelligence–Third Edition.

* = groups were significantly different, $p < .0001$. ** $p < .05$.

have reported neurological impairments, seizure disorders, motor deficits, psychological or emotional disorders, or neurodevelopmental disorders (including attention-deficit/hyperactivity disorder). All came from homes in which English was the primary language spoken. For children who were speakers of African American English as identified by the Diagnostic Evaluation of Language Variation Screener (Seymour, Roeper, & deVilliers, 2003), we adapted our scoring of the language tests described below.

We administered a battery of standardized tests: the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2004); the Peabody Picture Vocabulary Test–Revised (PPVT-III; Dunn & Dunn, 2007), and the Test of Nonverbal Intelligence–Third Edition (TONI-3; Brown, Sherbenou, & Johnsen, 1997). We originally adopted a criterion of 1 *SD* below the mean on at least one of the composite scores of the CELF (Receptive or Expressive Language), but four children fell just short of this criterion. Nevertheless, these children scored more than 1 *SD* below the mean on at least two subtests. One participant scored 2 *SD*s below the mean on Recalling Sentences and on Concepts and Directions; one participant scored more than 2 *SD*s below the mean on Recalling Sentences and Formulated Sentences and 1 *SD* below the mean on Sentence Structure; one participant scored more than 3 *SD*s below the mean on Formulated Sentences and 2.5 *SD*s below the mean on Sentence Structure; and one participant scored at least 2 *SD*s below the mean on Structured Sentences, Concepts and Directions, Formulated Sentences, and Recalling Sentences. These children were receiving therapy and had been diagnosed by an SLP as having SLI. All of the remaining children with SLI also had a history of or were receiving intervention. The groups differed significantly on all standardized test scores, CELF-R: $t(44) = 6.8$, $p < .0001$, $d = 2.33$; CELF-E: $t(44) = 9.9$, $p < .0001$, $d = 3.12$; CELF: $t(44) = 11$, $p < .0001$, $d = 3.36$; PPVT: $t(44) = 4.8$, $p < .0001$, $d = 1.62$; TONI: $t(44) = 2.1$, $p < .05$, $d = .64$. The age distributions were not significantly different, $t(44) = 1.7$, $p = 0.10$. Although all the children had nonverbal IQs within normal limits, the children with TD had higher nonverbal IQs than the SLI group children, though with a much smaller effect size. Hierarchical linear modeling allowed us to use TONI and CELF composite scores as subject-level predictors to test whether these scores predicted experimental effects.

Stimuli

We used the 30 sentence triplets listed in the appendix of McKee et al. (1993), with only minor modifications. The sentences in each triplet differed from the others only by the word *him*, *himself*, or a NP as the embedded object, as follows:

- Pronoun condition: *The alligator knows that the leopard with the green eyes is patting <u>him</u> on the head with a soft pillow.*
- Reflexive condition: *The alligator knows that the leopard with the green eyes is patting <u>himself</u> on the head with a soft pillow.*
- Noun condition: *The alligator knows that the leopard with the green eyes is patting <u>the girl</u> on the head with a soft pillow.*

Thus, there were a total of 90 experimental sentences. In addition, there were six practice sentences and 10 pseudo-experimental sentences that were identical in structure to the experimental sentences (e.g., *The guinea pig thinks that the seal is buying himself a beautiful toy to play*), as well as 14 filler sentences with varied syntactic structures (e.g., *The pencils are for sale in the elephant's favorite store, so he's going to go there early tomorrow morning*).

Instead of the black and white pictures (Snodgrass & Vanderwart, 1980) used by McKee et al. (1993), we used a color and texture modified version (Rossion & Pourtois, 2001), with two changes: BUFFALO was replaced with COW, and TURKEY was replaced with PENGUIN, and we changed those corresponding sentence triplets.

The same picture was presented with each experimental sentence in a given triplet across each of the presentation conditions. The picture always corresponded to the embedded subject (i.e., the antecedent of the reflexive). For example, with the aforementioned sentence triplet, a picture of a LEOPARD was presented at the offset of the underlined pronoun, reflexive, or noun. Thus, the 90 experimental sentences were paired with animate probes. For the 14 filler sentences and10 pseudoexperimental sentences, the picture probe was an inanimate corresponding to a word that appeared in the sentence (e.g., *The guinea pig thinks that the seal with the balloon is buying himself a beautiful toy to play with*. Probe = BALLOON). For the six practice sentences, half had animate probes and half had inanimate probes.

The 30 sets of sentence triplets were distributed across three sessions such that in each session a single member of the triplet was presented in one of the three conditions. The practice sentences and their picture probes were presented at the beginning of each session. The order of sentence presentation following the six practice sentences in a single session was randomized for each participant so that no participant heard the sentences within each list in the same order. This randomization also ensured that there was no fixed presentation sequence of conditions in a single session. By the end of the three sessions, each participant had heard all of the 30 triplets and had seen their corresponding pictures. The 10 pseudoexperimental sentences and the 14 filler sentences along with their probes were repeated in each of the three sessions. Their presentation was randomized along with the experimental sentences in each session. Thus, there were a total of 50 sentences and their probes in each of the three sessions. In a given session, there were 33 animate probes (experimental sentences and three practice sentences) and 27 inanimate probes.

Procedure

The experiment was divided into three sessions, with approximately a week between each session. Participants were seated in a comfortable chair in front of a computer screen and a two-button response box. At the beginning of each session, six practice trials were presented, followed by a brief break. There were 30 experimental trials interspersed with 24 filler trials.

A single trial included a blank white screen presented for 1,500 ms, followed by the number of remaining trials presented in parentheses at the center of the screen, which also served as a fixation point. Then, while the number was still on the screen, the sentence was auditorily presented to the participant via headphones. At the offset of the pronoun, reflexive, or name, a color picture was presented centrally on the screen for a maximum duration of 1,000 ms. The participant's task was to decide, as quickly as possible, whether the picture represented an animate or inanimate object, by pressing one of two buttons (one with a picture of a lion and one with a picture of an airplane, corresponding to animate and inanimate decision, respectively). If the participant pressed the button within 1,000 ms, the picture disappeared at the button press. If the participant did not respond within 1,000 ms, the picture would disappear, but the participant still had an additional 1,000 ms to respond, giving a maximum response window of 2,000 ms. If a participant did not respond within 2,000 ms, the trial was coded as a missing response. Each response was coded for accuracy.

Ten randomly selected trials of the 30 experimental trials in each block were followed by a tone and a pause, and the participants was then asked a paraphrase the sentence they had just heard. This response was not designed to be used as a dependent measure, as it served primarily to maintain the participant's attention to the stimulus sentences.

During the pause between each trial, the experimenter could temporarily halt the flow of the experiment by pressing a button on a PST Serial Response box (Psychology Software Tools Inc., Sharpsburg, PA) connected to the PC. This would be done if the child required an unplanned break, or if the experimenter determined that the participant was inattentive and needed a break with an encouragement to pay attention. Each session lasted approximately 20 min.

Apparatus

The experiment was programmed using E-Prime (Schneider, Eschman, & Zuccolotto, 2002a, 2002b). Two response buttons were placed between the child and the computer screen. Participants listened to the sentences over headphones and watched a computer screen.

Timing accuracy was examined by comparing E-Prime measured timing to an external measurement. A light-detecting diode on the computer screen was connected to a tone generator that produced one tone when a picture appeared, and another tone when the picture disappeared. We then used a sound-editing program (CoolEdit Pro; now Adobe Audition; Adobe Systems Inc., San Jose, CA) to calculate each tone-on/tone-off interval on the digital recording. These measures were compared with the E-Prime reported interval. For a single measurement session with 60 trials, the mean difference score was 0.15 ms (*SD* = 7.32 ms). The two measurements did not differ, $t(58) = 0.16$ $p = .87$. Thus, E-Prime was accurate and did not differ from an external measurement of the same event sequence. The standard deviation was the critical measure, because it indicates the divergence from the intended presentation point. With the exception of one trial with a 28-ms discrepancy (an apparatus error), the maximum discrepancy was ±14 ms, which is consistent with the refresh rate of the CRT screen. This is typically controlled for by synchronizing stimulus presentation with the onset of the vertical CRT screen blank, but because our sentences had a different sentence onset and picture onset latencies, this was not possible. Instead, we accepted that 68% of the trials would be within 7 ms of the desired presentation time.

Results

Coding and Outlier Treatment

Overall, 11% of the responses were either incorrect or late and were coded as errors. In addition, 18 trials with RTs below 300 ms were recoded as errors, as this latency is well below the threshold of possible responses.

Multiple regression analyses for each participant were performed to identify slow and fast outliers. We first performed a multiple regression analysis of condition effects within participants, and then removed outliers within condition for each participant. However, for some participants, variance differed among conditions: what would count as an outlier for one condition would be within 2 *SD*s for another condition. This caused the data to be unsmoothed for those participants (i.e., the data comb was uneven). We therefore chose a broader approach by using group (TD/SLI) as a single predictor in a regression analysis (this presumed that participant variance is consistent within group). This removed RTs above or below a more stable threshold. Another 180 trials were excluded, and the outliers were not replaced by condition or participant means. The total number of trials that were either incorrect, too slow, or were outliers according to this procedure was 497 (12% of the total).

Statistical Analysis

Probe Accuracy

Although three SLI participants had low probe accuracy scores (i.e., correctly judging whether a picture was animate or inanimate) ranging from 58% to 73% out of the 90 experimental trials, these scores were all significantly better than chance (binomial probability $p < .03$). All other participants except one (84%) had accuracy scores exceeding 90%. The mean correct probe decisions for children with SLI was 89% (*SD* = 10%), and the mean for children with TD was 95% (*SD* = 4%). A one-way analysis of variance (ANOVA) suggested that TD children were significantly better at the probe decision task than the SLI children, $F(1, 44) = 6.9$, $p < .05$, $d = -0.86$. Because of the ceiling effect and the very small effect size, this difference may be a statistical artifact. The children with TD were accurate for only five items more (out of 90) than their peers with SLI.

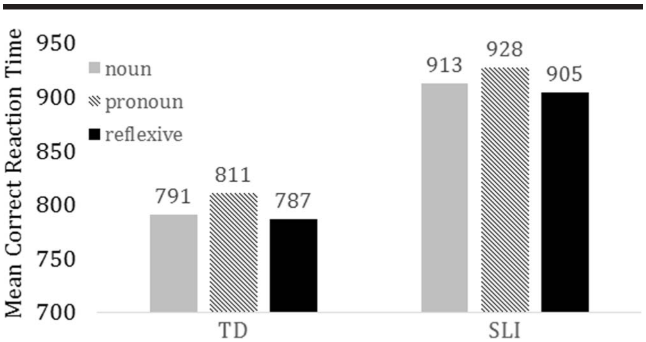
RT Data

The dependent measure was the RT to make a correct animacy judgment of the picture probe. The hypotheses were based on the assumption that only the grammatical antecedent is reactivated at the pronoun or reflexive. The picture probe across the items of a triplet consistently represented the antecedent of the reflexive. So, in the reflexive condition, the reactivated grammatical antecedent (Principle A) corresponds to the picture probe, and thus, there should be priming (i.e., faster RT for the animacy decision). In the pronoun condition, the reactivated grammatical antecedent (Principle B) does not correspond to the picture probe, and thus, there should be no priming (relatively slower RT for the animacy decision). In the noun condition, only the noun should be activated at that position, and because it does not correspond to picture probe, there should also be no priming (relatively slower RT similar to the pronoun condition). This is the pattern we expected if knowledge of Principles A and B is intact. If the children in either group lack this knowledge of binding theory, the RTs for the reflexive condition should be faster than for pronoun sentences. Figure 1 shows the descriptive statistics of the RTs by group and condition.

The pronoun condition was the slowest condition (M = 873 ms, SD = 194 ms), with both reflexives (M = 847 ms, SD = 188 ms) and nouns being faster (M = 854 ms, SD = 180 ms). To be specific, reflexives were only 7 ms faster than the baseline noun control condition, whereas pronouns were 19 ms slower than the baseline noun control condition. As is evident in Figure 1, this pattern was observed for both groups. The noun condition did not function as a baseline, contradicting expectations (McKee et al., 1993).

The data were statistically analyzed with hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002). This analysis treated the repeated measures of RT for animacy judgments across sentence types as nested within individual participants and participants as nested within groups for a three-level model. This allowed us to simultaneously test participants and items as random factors and to conduct tests of the relationships between participant-level properties such as CELF/TONI scores and the priming effect of reflexives.

Figure 1. Observed outcome by group and condition. Both groups exhibit the same pattern; reflexives and nouns have similar RTs, and pronouns have a slower RT.



In this statistical approach, the first analysis was at the level of individual participants. Within each participant, sentence types can be thought of as nested within the pictures (for which the child makes animacy decisions) as follows. Participant n: Picture 1: Sentence 1 (noun), Sentence 2 (pronoun), Sentence 3 (reflexive); Picture 2: Sentence 1 (noun), Sentence 2 (pronoun), Sentence 3 (reflexive); …Picture 30: Sentence 1 (noun), Sentence 2 (pronoun), Sentence 3 (reflexive). Each participant received three presentations of each picture, with sentence type (three members of a triplet) varying across those picture presentations. To describe the data at this level, the HLM constructed, for each participant, a set of microregression equations, one for each picture. A complete dataset for a single participant was 30 such equations, one for each sentence triplet/picture j. The RT to a given picture j (on three occasions for a single participant i) was described by this equation: $RT_{ij} = \pi_{0j} + \pi_{1j}$ (PRONOUN) + π_{2j} (REFLEXIVE). Because this exhausted the stimuli and trials, there was no error term. The coefficients in these equations expressed how much the RT was altered relative to the baseline (which was set to be the noun condition). Having knowledge of binding theory predicts that the reflexive coefficients should generally result in a significant reduction of RT relative to the intercept, whereas the pronoun coefficients should not significantly change the RT. This constituted Level 1 in our three-level HLM model.

From those 30 equations, three new regression equations were constructed for the 30 intercepts, the 30 pronoun coefficients, and the 30 reflexive coefficients. Without any predictors at this level, this yielded the following set of Level 2 equations: $\pi_{0j} = \beta_{00j} + r_{0j}$; $\pi_{1j} = \beta_{10j} + r_{1j}$; and $\pi_{2j} = \beta_{20j} + r_{2j}$. β_{00} represented the mean baseline (noun) RTs for all noun intercepts and r_{0j} the variance around this mean; β_{10} represented the mean of the pronoun coefficients, and β_{20} represented the mean of the reflexive coefficients. Level 2 represented how each single sentence was nested inside a stimulus triplet (30 triplets), defined by a single picture probe, and exhaustively models a single participant. Because there were no factors dividing the picture stimuli into different categories, Level 2 has no coefficients, only intercepts.

Once each participant had an equation similar to this, the entire set of participants was modeled by constructing Level 3 regression equations for the β-intercepts and β-coefficients: $\beta_{00j} = \gamma_{000j} + \gamma_{001}(SLI)_{+}u_{00j}$; $\beta_{10j} = \gamma_{100j} + \gamma_{101}(SLI)_{+}u_{10j}$; $\beta_{20j} = \gamma_{200j} + \gamma_{201}(SLI)_{+}u_{20j}$. At this level, categorical variables such as group (SLI or TD) and continuous variables (e.g., TONI scores) were added to test for interactions between between-participants and within-participants variables. For example, did being an SLI child significantly change the effect of the reflexive coefficient? Did the TONI score make a difference for the strength of the reflexive effect? The Level 3 model tests the effect that being an SLI child has on the intercept, pronoun coefficient, and reflexive coefficient, respectively. The coefficients at Level 3 correspond to the interaction terms in a repeated measures ANOVA. If both groups have knowledge of binding theory, the reflexive coefficient π_2 should be significant, but the interaction term γ_{201} should not be (i.e., having SLI should not significantly modulate the RT change for reflexives). In a similar manner, the pronoun coefficient π_1 and the interaction term γ_{101} should not be significant. If children with SLI don't have knowledge of binding theory, the interaction term γ_{201} should be significant, because of a significant increase in RT relative to TD children in this condition (i.e., reflexive RT should increase and be equivalent to pronoun RT).

We started by testing the model without any group-level interaction term added at Level 3, as if the data came from one single homogeneous group of participants, testing for the main effect of condition only. For the final estimation of fixed effects, we used robust standard errors in all tests. In this model, the mean intercept was 853 ms (SE = 28 ms). The reflexive coefficient was −7 ms, a nonsignificant change, $t(45) = -0.61$, $p = .5$. However, the pronoun coefficient represented an increase of 20 ms, which was significant, $t(45) = 1.98$, $p = .05$, $d = 0.58$. Changing from a noun condition to a pronoun condition led to a slower RT. To test whether there was an interaction between condition and group, we added SLI as a coefficient at Level 3 for each of the components at Level 1: intercept (noun), pronoun, and reflexive coefficient. This resulted in a significant coefficient for SLI on the intercept (i.e., a main effect of group in an ANOVA); children with SLI were 135 ms slower in the noun condition, $t(44) = 2.05$, $p = .046$, $d = 0.61$. However, adding SLI to the reflexive and pronoun slopes did not result in a significant change in RT relative to the increase in intercept. SLI children's reflexive condition was 10 ms faster than nouns for reflexives, and pronouns were 15 ms slower; both changes were not significant, $t(44) = -0.335$ and −0.403, respectively, $p > .05$. Both groups exhibited the same pattern of effects.

We also analyzed the difference between reflexives and pronouns. This was done by using the pronoun condition as the intercept. The rationale was that priming for reflexives and nonpriming for pronouns should yield faster RTs for the reflexives compared with the pronouns. Reflecting the use of the pronoun condition as intercept, we changed the model to this Level 1 equation: $RT_{ij} = \pi_{0j} + \pi_{1j}$ (NOUN) + π_{2j} (REFLEXIVE). The intercept (mean RT for pronouns) was 872 ms. The reflexive coefficient represented a 26-ms decrease in RT from pronoun, which was significant, $t(45) = -2.77$, $p = .008$, $d = 0.82$. The noun coefficient was −19 ms (i.e., nouns were 19 ms faster than pronouns in this model); this was not significant, $t(45) = -1.75$, $p = .086$.

We then added SLI as a predictor at all three levels. As in the previous model, the SLI coefficient changed the intercept by adding 126 ms; however, this was not significant in the current model, $t(44) = 1.14$, $p = .25$, with pronoun as intercept. The SLI interaction term for the reflexive was not significant, differing only by 1 ms, $t(44) = 0.04$, $p = .9$. SLI children had the same facilitation effect as the TD children for reflexives when compared with pronouns. The noun coefficient in this model was 23 ms faster than the pronoun condition, which was not significant, $t(44) = -1.57$, $p = .12$. Adding SLI to this term was also not significant; the SLI noun condition was 15 ms faster than pronouns, $t(44) = 0.31$, $p = .75$. These results are summarized in Table 2.

This replicates the pattern observed by McKee et al. (1993), in which pronouns and nouns did not differ significantly. Reflexives were processed faster than pronouns, which can be interpreted as a priming effect compared with the pronoun condition.

We next tested whether participants' scores on the standardized tests were predictors of the priming effect for reflexives (Baayen, Davidson, & Bates, 2008). For example, we tested whether TONI score interacted with the coefficients of interest, by adding TONI scores as a predictor at the level of participants (Level 3) in the model with reflexive and noun coefficients at Level 1. The reflexive slope was significant, $t(44) = -2.67$, $p = .011$, $d = 0.79$, but the TONI scores showed no interaction with the intercept or the coefficients (the increment being less than 1 ms per TONI unit). Thus, nonverbal intelligence had no effect on processing differences between reflexives and pronouns. The same was true for the PPVT scores and compound CELF scores. Replacing the SLI variable with the overall CELF scores had no significant effects on intercepts and no effect on the reflexive and noun slopes. In a similar manner, replacing the SLI group variable with the continuous predictors CELF-expressive, the CELF-receptive, or the PPVT scores resulted in no significant effects, Last, we also tested the effect of SES, coded as 0 to 4 (for *lower*, *lower-middle*, *middle*, *upper-middle*, and *upper class*). Again, SES had no significant effects on intercept, noun, or reflexive slopes.

Discussion

The present study examined a core aspect of binding knowledge, Principles A and B, in children with SLI using an online cross-modal picture priming paradigm. Although children with SLI were generally slower in categorizing picture probes (*alive/not alive*) than their TD peers, they exhibited the same pattern of priming as their TD peers. To be specific, RTs were significantly faster in the reflexive condition, in which the picture probe matched the grammatical (Principle A) reactivated antecedent, than in the pronoun condition, in which the picture probe did not match the grammatical reactivated antecedent. This suggests that their knowledge of relations between pronouns or reflexives and their antecedents are intact, at least in the basic structural constraints that govern reflexives and pronouns. It seems to place the locus of this and perhaps other sentence comprehension deficits in the realm of processing rather than in lack of grammatical knowledge.

There are several caveats regarding this conclusion. First, we did not directly test sentence comprehension offline or online. An offline task might have just led us to a correlation between the offline and online tasks, but still would have added information to our findings. The potential biases in truth-value judgments tasks or in picture point tasks might not actually reflect comprehension and perhaps should not be viewed as a gold standard in this case. Although cross modal priming tasks do permit an examination of the

Table 2. Results of hierarchical linear modeling with pronoun as baseline (with robust standard errors).

Fixed effect	Coefficient	Standard error	*t* ratio	Approx. *df*	*p* value
For INTRCPT1, π_0 (pronoun condition)					
For INTRCPT2, β_{00} (mean of pronoun intercepts)					
INTRCPT3, γ_{000} (TD mean RT pronouns)	812	100	8.119	44	<.001
SLI, γ_{001} (relative change from TD group)	126	110	1.143	44	.259
For NOUN slope, π_1					
For INTRCPT2, β_{10} (mean change in noun RT)					
INTRCPT3, γ_{100} (TD mean noun effect)	−23	14	−1.578	44	.122
SLI, γ_{101} (relative change from TD group)	8	27	0.313	44	.756
For REFLEXIV slope, π_2					
For INTRCPT2, β_{20} (mean change in reflexive RT)					
INTRCPT3, γ_{200} (TD mean reflexive effect)	−26	10	−2.623	44	.012
SLI, γ_{201} (relative change from TD group)	0.67	17	0.039	44	.969

Note. The numbers after the model terms (e.g., the 1 in INTRCPT1) refer to the level of the hierarchical linear model. TD = children with typical development; RT = reaction time; SLI = children with specific language impairment.

information that is active in real time, they can examine only discrete points in time. A more informative approach on both counts would be to use an eye tracking experiment with pictures representing possible interpretations, which would permit tracking the child's consideration of different interpretations as the sentence unfolds as well as the child's endpoint response in selecting a picture.

The noun condition RT was not different from the reflexive RT, and both were faster that the pronoun in contrast to our predictions on the basis of McKee et al. (1993). In those children (4;1–5;6) and in adults, animacy judgments in the baseline noun and pronoun conditions did not differ and were slower than the reflexive condition. Thus, the noun condition did not function as a baseline in our experiment. We cannot readily account for the different results for the noun condition across the two studies, but the NP may not have been the logically correct control condition.

Noun processing, as was required in this study, may differ from the sentence internal reference relations of pronouns and reflexives, and there would be no specific prediction concerning the relative speed of NP processing (in which a picture probe appears several words after the corresponding word is heard in the sentence) versus reflexive or pronoun processing (in which an automatic process is assumed to reactivate the grammatical antecedent). On the other hand, it is possible that because reflexive and pronoun processing require the establishment of syntactic reference chains (i.e., syntactically represented binding relations), whereas NP processing does not, nouns might be processed faster because fewer operations are required. This would explain our findings but not those reported by McKee et al. (1993). One additional issue is that two RTs may reflect different processes even within the same task. For example, the noun condition may have yielded a faster RT simply because it did not entail accessing an antecedent, whereas the reflexives may have yielded comparably fast RTs simply because the probe matched the antecedent. In standard sentence processing experiments with adults using cross-modal priming (e.g., Love & Swinney, 1996), at each probe point there is a probe related to the assumed antecedent to be activated and a completely unrelated probe, and priming occurs when the related probe decision is faster than the unrelated probe (one that did not correspond to a word in the sentence). Using large numbers of participants, these experiments are typically done cross-sectionally for probe types and multiple probe points. The approach we adapted was a substantially shortened within-participants design that was more appropriate for children. Thus, an ideal set of probes was not possible. The critical comparison in our study, though, for the purposes of investigating binding knowledge was between reflexives and pronouns. And we did find the probes in the reflexive condition had faster RTs than the same probe in the pronoun condition, indicating that the appropriate reference was reactivated in the former but not in the latter. This provides some counter evidence to the findings and conclusions of van der Lely and Stollwerck (1997) regarding the inability of children with SLI to correctly pair antecedents with pronouns and reflexives without semantic cues.

One final issue concerning the findings are the overall slower responses of the children with SLI compared with their TD peers. This certainly is consistent with the observation that children with SLI often exhibit slower RTs than their TD peers (e.g., Miller, Kail, Leonard, & Tomblin, 2001). In the current study, a number of specific factors may have contributed to these slower responses besides perhaps being slow to make animacy decisions about picture probes.

Recent proposals concerning sentence comprehension posit that underlying processes such as interference may be involved in comprehending sentences with pronouns or reflexives as well as other structures (Gordon, Hendrick, Johnson, & Lee, 2006; MacKenzie, Walenski, Love, & Shapiro, 2015). Along with interference, binding of stimuli (features, objects) or stimuli and responses and release from binding also play a role in recent characterizations of working memory, executive functions, and the microstructure of phenomena such as remembering and forgetting (Hasher, Lustig, & Zacks, 2007; Lewis, Vasishth, & Van Dyke, 2006; Oberauer & Lange, 2009; Van Dyke & Johns, 2012). The term *binding* in this context does not have the same meaning as in linguistic theory; here it refers to how features of an object, a stimulus and its location, or even parts of sentences (chunking) are linked in working memory. In working memory, binding may actually include the automatic and unconscious processes of the parser (Baddeley, Hitch, & Allen, 2009), including the linking of pronouns or reflexives with their antecedents. Children with SLI have demonstrated deficits in suppressing irrelevant information, have less resistance to proactive inhibition, a higher rate of interference errors, and poorer coordination between activation and inhibition than their TD peers (Epstein, Hestvik, Shafer, & Schwartz, 2013; Marton, Campanelli, Eichorn, Scheuer, & Yoon, 2014). They may also have weaker binding or release from binding in working memory (Marton et al., 2014).

When listeners encounter anaphora or pronouns, all potential antecedents (i.e., all nouns or all structurally appropriate nouns in the sentence) may be activated, and all but the correct antecedent must rapidly deactivate or be inhibited. Given that children with SLI are more prone to interference, are less well able to inhibit irrelevant information, may have poorer activation of the target antecedent, and have weaker binding and release from binding in working memory, their overall processing may be slower, as we found. The same may be true for other structures that require lexical activation and inhibition in the course of sentence comprehension such as gap filling in object relatives (Hestvik, Schwartz, & Tornyova, 2010).

Given the limitations of the methods used in previous studies of binding, sentence processing paradigms may offer a viable alternative. Although the present study was not designed for this purpose, such paradigms have the potential to explore interference and binding effects in working memory that may affect the comprehension of pronouns, reflexives, and other structures. Although cross-modal priming paradigms have the advantage of examining sentence processing in real time, they examine information activation only at discrete points in time. Methods such as eye tracking or event-related potentials that yield continuous examination of information activation along with end-point comprehension measures may be better suited to these studies.

Acknowledgments

Grants RO1DC003885 and R01DC011041 from the National Institute on Deafness and other Communication Disorders supported the reported research and the article preparation. The authors extend their appreciation to Stacy Asay, who played a significant role in participant recruitment and data collection. We also acknowledge the contributions of Dr. David A. Swinney to our studies of language processing in SLI.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language, 61,* 438–456.

Bishop, D. V. (1983). *Test of reception of grammar*. Manchester, UK: Manchester University.

Bishop, D. V. (1997). *Uncommon understanding: Development and disorder of language comprehension in children*. Hove, East Sussex, UK: Psychology Press.

Brown, L., Sherbenou, R., & Johnsen, S. (1997). *Test of Nonverbal Intelligence, Third Edition*. Austin, TX: Pro-Ed.

Chien, Y.-C., & Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition, 1,* 225–295.

Chomsky, C. (1969). *The acquisition of syntax in children 5 to 10*. Cambridge, MA: MIT Press.

Chomsky, N. (1993). *Lectures on government and binding: The Pisa lectures* (7th ed.). Berlin, Germany: Mouton de Gruyter.

Conroy, A., Takahasi, E., Lidz, J., & Phillips, C. (2009). Equal treatment for all antecedents: How children succeed with Principle B. *Linguistic Inquiry, 40,* 446–486.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition*. Bloomington, MN: Pearson.

Elbourne, P. (2005). On the acquisition of Principle B. *Linguistic Inquiry, 36,* 333–365.

Epstein, B., Hestvik, A., Shafer, V. L., & Schwartz, R. G. (2013). ERPs reveal atypical processing of subject versus object Wh-questions in children with specific language impairment. *International Journal of Language & Communication Disorders, 48,* 351–365.

Franks, S. L., & Connell, P. J. (1996). Knowledge of binding in normal and SLI children. *Journal of Child Language, 23,* 431–464.

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. doi:10.1037/0278-7393.32.6.1304

Grimshaw, J., & Rosen, S. T. (1990). Knowledge and obedience: The developmental status of the binding theory. *Linguistic Inquiry, 21,* 187–222.

Hasher, L., Lustig, C., & Zacks, R. (2007). Inhibitory mechanisms and the control of attention. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York, NY: Oxford University Press.

Hestvik, A., Schwartz, R. G., & Tornyova, L. (2010). Relative clause gap-filling in children with specific language impairment. *Journal of Psycholinguistic Research, 39,* 443–456. doi:10.1007/s10936-010-9151-1

Hollingshead, A. (1975). *Four factor index of social status*. Unpublished manuscript, Yale University, New Haven, CT.

Joanisse, M. F., & Seidenberg, M. S. (2003). Phonology and syntax in specific language impairment: Evidence from a connectionist model. *Brain and Language, 86,* 40–56.

Kirk, S., McArthy, J., & Kirk, W. (1968). *Illinois Test of Psycholinguistic Abilities*. Urbana, IL: University of Illinois Press.

Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). Cambridge, MA: MIT Press.

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

Love, T., & Swinney, D. (1996). Coreference processing and levels of analysis in object-relative constructions: Demonstration of antecedent reactivation with the cross-modal priming paradigm. *Journal of Psycholinguistic Research, 25,* 5–24.

Lust, B. (1986). *Studies in the acquisition of anaphora* (Vol. 1). Dordrecht, the Netherlands: Reidel.

MacKenzie, S., Walenski, M., Love, T., & Shapiro, L. P. (2015). The auditory comprehension of wh-questions in aphasia: Support for the intervener hypothesis. *Journal of Speech,* *Language, and Hearing Research, 58,* 781–797. doi:10.1044/2015_JSLHR-L-14-0099

Marton, K., Campanelli, L., Eichorn, N., Scheuer, J., & Yoon, J. (2014). Information processing and proactive interference in children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 57,* 106–119.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2009). Pronoun co-referencing errors: Challenges for generativist and usage-based accounts. *Cognitive Linguistics, 20,* 599–626.

McDaniel, D., Cairns, H. S., & Hsu, J. R. (1990). Binding principles in the grammars of young children. *Language Acquisition, 1,* 121–138. doi:10.1207/s15327817la0101_5

McKee, C., Nicol, J., & McDaniel, D. (1993). Children's application of binding during sentence processing. *Language and Cognitive Processes, 8,* 265–290.

Miller, C. A., Kail, R., Leonard, L. B., & Tomblin, J. B. (2001). Speed of processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 44,* 416–433.

Oberauer, K., & Lange, E. B. (2009). Activation and binding in verbal working memory: A dual-process model for the recognition of nonwords. *Cognitive Psychology, 58,* 102–136. doi:10.1016/j.cogpsych.2008.05.003

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Reinhart, T. (1983). *Anaphora and sematic interpretation*. London: Croom Helm.

Reinhart, T. (1986). Center and periphery in the grammar of anaphora. In B. Lust (Ed.), *Studies in the acquisition of anaphora: Defining the constraints* (Vol. 1, pp. 123–150). Dordrecht, the Netherlands: Reidel.

Rossion, B., & Pourtois, G. (2001). Revisiting Snodgrass and Vanderwart's object database: Color and texture improve object recognition. *Journal of Vision, 1*(3), 413.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002a). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002b). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Schwartz, R. G. (2009). Specific language impairment. In R. G. Schwartz (Ed.), *Handbook of child language disorders* (pp. 3–43). New York, NY: Psychology Press.

Semel, E., Wiig, E. H., & Secord, W. A. (2004). *Clinical Evaluation of Language Fundamentals–Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Seymour, H. N., Roeper, T. W., & deVilliers, J. (2003). *Diagnostic Evaluation of Language Variation–Screening Test*. San Antonio, TX: The Psychological Corporation.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215.

Solan, L. (1987). Parameter setting and the development of pronouns and reflexives. In T. Roeper & E. Williams (Eds.), *Parameter setting* (pp. 189–210). Dordrecht, the Netherlands: Reidel.

van der Lely, H. K., & Stollwerck, L. (1997). Binding theory and grammatical specific language impairment in children. *Cognition, 62,* 245–290.

Van Dyke, J. A., & Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Language and Linguistics Compass, 6,* 193–211. doi:10.1002/lnc3.330