Running head: Binding in SLI

Processing binding relations in Specific Language Impairment

Richard G. Schwartz[1], David A. Swinney[2], Arild Hestvik[3], Liat Seiger-Gardner,[1,4], and Diana Almodovar[4]

[1]The Graduate Center, City University of New York

[2]University of California, San Diego

[3]University of Delaware

[4]Lehman College

Richard G. Schwartz, Ph.D. Program in Speech-Language-Hearing Sciences, The Graduate Center of the City University of New York; David A. Swinney, Department of Psychology, University of California, San Diego; Arild Hestvik, Department of Linguistics and Cognitive Science, University of Delaware; Liat Seiger-Gardner, Speech-Language-Hearing Science Lehman College; Diana Almodovar, Speech-Language-Hearing Science Lehman College.

Abstract

Purpose: This sentence processing experiment examined the relative abilities of children with specific language impairment (SLI) and children with typical language development (TLD) 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 TLD (7;3 – 10;11) participated in a cross-modal picture priming experiment to determine whether they selectively activated the correct referent at the pronoun or reflexive in sentences. As they listened to sentences, a picture appeared on the screen at the pronoun/reflexive/NP and the child had to judge the animacy of the picture by pressing a button.

Results: The groups did not differ in animacy judgments and exhibited the same pattern of reaction times. Reaction times for both groups were slower for sentences with pronouns than reflexives or noun phrases. The children with SLI were slower overall in their RTs than their typically developing peers.

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 typically developing peers. The overall delay in reaction time 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 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 binding in typically developing children and in children with SLI illustrates this controversy.

Binding Theory (BT) is the set of syntactic constraints on the reference of pronouns, reflexives, and names that entails three principles. Principle A states that an anaphor (i.e., 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. In the sentence, *The mother says that the girl splashed her*, the pronoun refers to *mother*, but in the sentence *John splashed him*, the pronoun cannot refer to *John*. A key finding is that children make the correct assignment when a pronoun antecedent is a quantifier, but not when the antecedent is referential. These binding principles are considered to be part of the innate knowledge of language (Chien & Wexler, 1990; Chomsky, 1969; Lust, 1986; McDaniel, Cairns, & Hsu, 1990). A number of studies using acting out tasks, truth-value judgments (in relation to pictures and sentences), or grammaticality judgments suggest that most young children reliably follow Principle A, but violate Principle B (e.g., referential, but not quantified NPs) and, sometimes, Principle C with notably poorer performance on Principle B until somewhat later in development. However, the findings and the explanations offered are quite mixed. These explanations have varied from a pragmatic principle (Principle P) that young children have not mastered or that Principle B errors may be explained by the *Lexical Learning*

Hypothesis (e.g., (Chien & Wexler, 1990; Reinhart, 1983, 1986) to a processing disorder (Solan, 1987).

This literature is characterized by numerous methodological issues such as the fact that act-out tasks may simply reveal preferences, the bases for grammaticality judgments provided by young children are undetermined, and, in some cases, the stimuli themselves may have influenced performance (Conroy, Takahasi, Lidz, & Phillips, 2009; Grimshaw & Rosen, 1990). For example, Grimshaw and Rosen noted that in two-picture studies, children's errors on sentences involving possessive may reflect a misunderstanding of the possessive rather than binding errors, whereas four-picture studies can distinguish binding and possessive interpretation errors. In their own study, they concluded that children do know binding, but the equivalency of performance on reflexives and pronouns remains in question and subject to the challenges of experiments. 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 children do not obey Principle B (Elbourne, 2008) and findings that children's responses vary with the specific pronoun and 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).

Online studies of binding

Although researchers have made inferences about processing from post-comprehension tasks, few studies have used sentence processing paradigms to examine binding. Love and Swinney (Love & 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 subjects 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 BT (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 typically developing children (4;1 to 6;4). The stimuli were triplets of sentences that differed only in whether the embedded object was a pronoun, a reflexive, or a referential noun phrase (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 subject's task was to decide whether the picture was animate or not. The prediction was that reflexive should trigger reactivation of the embedded clause subject and prime the picture (faster reaction time), whereas the picture judgment at the pronoun would be slower because there is no reactivation of the antecedent. The name (*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 name condition whereas picture judgments in the pronoun condition were not faster than in the name condition. There was no difference between the reflexive and pronoun condition. A subsequent truth-value judgment task (based on Chien & Wexler 1993), allowed them to divide their subjects into two groups, one that exhibited adult-like performance on the judgment task and one that performed more poorly. In a re-examination of the RT data, the first sub-group exhibited the expected difference between pronouns and reflexives, whereas the other group did not exhibit the RT differential. Combined with the behavioral data from this latter group, McKee et al. 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 typically developing children (3;7- 8;6), and 11 SLI children (3;9 - 7;8). Subjects 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 non-structural 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 in two experiments to examine whether children with grammatical SLI interpretation of sentences with reflexives and pronouns appeared to be constrained by Binding conditions. The subjects were 12 children with Grammatical SLI (9;3-12;10) and three control groups (12 children in each) of typically developing children, one group matched on the TROG and the Grammatical Closure Subtest of the ITPA (5;5-6;4) and one group (6;5-7;4) matched on expressive and receptive vocabulary scores, and one group (7;5 -8;9) was 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 SDs of the mean); they scored at least 1.5 standard deviations below the mean on a test of grammatical comprehension, up to 2.5 standard deviations below the mean on TROG, and as many as 5.5 standard deviations below the mean on the Grammatical Closure Test of the ITPA.

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 either a quantifier or a referential noun phrase, and the anaphor was either pronoun or 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, while 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), while the control groups ranged from 90-100% correct. Even though their average percent correct was lower, the average score was 4.5 out of 6 items with an 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 an SD of 1.70) on name-pronouns, and to 55% (chance;

mean score 3.25 with a SD of 1.76) on quantifier-reflexives, while 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 SDs 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 concluded that the performance of SLI children is 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. Any one, or a combination of these factors could account for these findings.

Despite these clearly equivocal findings, van der Lely and Stollwerck 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. Finally, 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 lead to their conclusion that SLI children “do not have the syntactic knowledge characterized by BT (p. 276)”, 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 their 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 on-line processing of sentences.

Current study

In order to test whether SLI children exhibit knowledge of binding theory during on-line processing, we used a cross-modal picture priming paradigm (McKee et al., 1993) to compare children with SLI and their age-matched typically developing peers.

The prediction for the TD group was that they should show significantly faster reaction times to the picture of the embedded clause subject when it occurred after a reflexive that was bound to this subject, than when it occurred after a pronoun referring to the higher clause subject or after a noun phrase. If SLI children don't obey binding theory during real time processing, they should not show this reflexive effect, and there should be a group by condition interaction reflecting this. If SLI children obey binding constraints during real-time processing, there should be no group by condition interaction.

Method

Participants

A total of 48 children (7;3 – 10;11) participated initially: 25 typically developing (TD) children and 23 children with SLI. One SLI subject with only 9% valid responses was excluded because of a large number of very late responses (greater than 2 seconds); such late responses were not typical of other SLI subjects. One TD subject had an accuracy score on comprehension questions of 0.52. The 95% confidence interval of this proportion, assuming a chance model ($p = 0.5$) is [0.45,0.60] is consistent with random responses; therefore, this subject was also removed from analysis. Thus, 46 children were included in the final analysis: 22 children with SLI and 24 children with TLD (see Table 1). The TD group included 13 females and 11 males; the SLI group included 9 females and 13 males. The children in both groups were drawn from a variety of socio-economic backgrounds (SES). 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. Three of the TD children were left-handed and four of the SLI children were left-handed.

All children passed a standard hearing screening, and all had normal articulation and phonological skills in conversational speech, as judged by a speech-language pathologist. The children included did not have reported neurological impairments, seizure disorders, motor deficits, psychological or emotional disorders, or neurodevelopmental disorder (including ADHD).

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 Test of Non-verbal Intelligence – Third Edition (TONI-3) (Brown, Sherbenou, & Johnson, 1997). We originally adopted a criterion of one standard deviation 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, they scored more than one standard deviation below the mean on at least two subtests, were receiving therapy, and had been diagnosed by a speech-language-pathologist as having specific language impairment. All of the remaining children with SLI were also receiving intervention.

The groups differed significantly on all standardized test scores (CELF-R: $t(44) = 6.8$, $p < .0001$; CELF-E: $t(44) = 9.9$, $p < .0001$; CELF: $t(44) = 11$, $p < .0001$; PPVT: t (44)

= 4.8, *p* < .0001; TONI: *t*(44) = 2.1, *p* < .05). 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 TLD had higher non-verbal IQs than the children SLI group. Hierarchical linear modeling (mixed model analysis), allowed us to use TONI and CELF scores as subject level predictors to test whether these scores predicted experimental effects.

Stimuli

We used the 30 sentence triplets from McKee at al. (1993), where each sentence in the triplet differed from the others only by the word *him*, *himself*, or a noun phrase as the embedded object as follows:

Pronoun Condition: *The alligator knows that the leopard with the green eyes is patting **him** on the head with a soft pillow.*

Reflexive Condition: *The alligator knows that the leopard with the green eyes is patting **himself** on the head with a soft pillow*.

Name Condition: *The alligator knows that the leopard with the green eyes is patting **the girl** on the head with a soft pillow*.

The sentences were digitally recorded by a female speaker in a soundproof booth. There were six practice sentences, 14 filler sentences, and 10 pseudo-experimental sentences, which had the same syntactic shape as the experimental sentences, but were matched to non-experimental picture probes.

Instead of the black and white pictures (Snodgrass & Vanderwart, 1980), we used a color and texture modified version of the same picture set (Rossion & Pourtois, 2001), with two changes: BUFFALO was replaced with COW; and TURKEY was replaced with PENGUIN. The same picture was presented with each sentence in a given triplet across each of the presentation conditions. For example, for the aforementioned sentences, a picture of a leopard was presented at the offset of the underlined pronoun, reflexive, or noun.

The thirty 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 order of sentence presentation in a single session was randomized for each subject, so that no subject 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 subject had heard all of the 30 triplets and saw their corresponding pictures.

Procedure

The experiment was divided into three sessions, with approximately a week between each session. Subjects were seated in a comfortable chair in front of a computer screen and a two-button response box. At the beginning of each session, 6 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 1500 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 subject 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 1000 ms. The subject’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 subject pressed the button within 1000 ms, the picture disappeared at the button press. If the subject did not respond within 1000 ms, the picture would disappear, but the subject still had an additional 1000 ms to respond, giving a maximum response window of 2000 ms. If a subject did not respond within 2000 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 subject was then asked a comprehension question for the sentence. The responses were audio recorded. The accuracy of this response was scored but not used as a dependent measure, as it served primarily to maintain the subject'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 connected to the PC. This would be done if the child required an unplanned break, or if the experimenter determined that the subject was inattentive and needed a break with an encouragement to pay attention. Each session lasted approximately 20 minutes.

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. Subjects 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 placed on the computer screen was connected to a tone generator that was triggered when a picture appeared, and another tone when the picture disappeared. We then used a sound-editing program (CoolEdit Pro) to calculate each tone-on/tone-off interval on the digital recording. These measures were compared to 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). A t-test showed that the two measurements did not differ ($t(58) = .16$ $p = .87$). Thus, the E-Prime record of its own events 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 desired presentation time point. With the exception of one trial with a 28 ms discrepancy (a trial for which E-Prime also recorded a 67 ms audio onset delay that was an apparatus error), the maximum discrepancy was +/-14 ms, which is consistent with the refresh rate of the CRT screen. In the most extreme case, if a picture is presented right after the electron beam has passed the center of the screen, it will take 14 milliseconds before it is fully drawn (or 14 ms too early). Typically, this is controlled for by synchronizing stimulus presentation with the onset of offset of the vertical blank on a CRT, but because every stimulus sentence in this experiment had a different latency between the onset of the sentence and the onset of the picture, 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

Of the 3240 trials, 11% were either incorrect or too late responses, and were coded as errors by the software. In addition, 18 trials had reaction times below 300 ms. Trials with RTs below 300 ms were recoded as errors (missing data), as this latency is well below the threshold of possible responses.

Multiple regression analyses for each individual subject were performed to identify slow and fast outliers. We first performed a multiple regression analysis of condition effects within subjects, and then removed outliers within condition for each subject. However, we found that for some subjects, variance differed among conditions, such that what would count as an outlier for one condition would be within 2 standard deviations for another condition. This caused the data to be *un-smoothed* for those subjects (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 presumes that subject variance is consistent within group). This was more conservative and more even by removing RTs above/below a more stable threshold. This removed another 180 trials, and the outliers were not replaced by condition or subject means. The total number of trials that were either incorrect, too slow, or were outliers according to this procedure was 549 (17% of the total).

Statistical analysis

Although three SLI subjects had low accuracy scores ranging from 58% to 73% out of 180 trials, these scores were all better than chance. All other subjects except one (84%) had accuracy scores exceeding 90%.

We conducted two separate statistical analyses: a repeated measures ANOVA, and a Hierarchical Linear Model analysis. For the repeated measures ANOVA, we first calculated a mean per condition for each subject, and then used these means as the three single dependent measures for the within-subject conditions Noun, Pronoun and Reflexive. Group was entered as a between-subject factor, resulting in a 3 x 2 mixed factorial design. The repeated measures ANOVA simply asked whether the three means estimate the same underlying distribution or not. However, our experiment had specific predictions about the direction of those differences. Using the name condition as baseline and control for the reflexive and pronoun condition, the prediction was that the reflexive condition should exhibit priming relative to the name condition, whereas the pronoun condition should be the same as the name condition. The priming in the reflexive condition would occur because of the referential relation between the immediately preceding reflexive pronoun and that subject—allowed by binding theory. The reflexive effectively reactivates the embedded subject leopard, and this in turn should prime the animacy decision of the leopard picture. A pronoun, on the other hand, is blocked by binding theory from referring to this same embedded subject, and in fact must refer to the matrix subject *alligator*; there is no priming and the picture decision should be slower. If SLI children wrongly interpreted the pronoun as referring to the embedded, closest subject, then both reflexives and pronouns should be faster than the name condition. Thus, the name condition served as a baseline for both pronouns and reflexives (McKee et al.,1993); this also entails that reflexives should be faster than pronouns when these two conditions are compared directly. This analysis yielded a main effect of Condition ($F(2,88) = 3.49$, $p = .034$), and a main effect of Group, $F(1,44) =$

6.36, *p* = 0.015. The interaction between Group and Condition was not significant (*F*(2,88 = 0.048, *p* = .95). These effects are illustrated in Figure 1.

We also conducted an orthogonal contrast analyses, first with all subjects together and then separately for each group. With all subjects pooled, the contrast between pronouns and reflexives was significant (*t*(44) = 3.08, *p* = .004). It was also significant for each group analyzed separately (TD effect: 26 ms; t(44) = 2.2, p = .03; SLI effect: 27 *t* ms; (44) = 2.15, *p* = 0.04). Both groups exhibited priming consistent with knowledge of binding condition A.

We next turn to the noun condition. Orthogonal contrast analysis with all subjects pooled showed that nouns were marginally faster than pronouns (39 ms, *t*(44) = 1.86, *p* = .07) . For the SLI group in isolation, however, the contrast was not significant (17 ms difference, *t*(44) = 1.1, *p* = .26), and similarly for the TD group in isolation (22 difference, *t*(44) = 1.5, *p* = .13). Thus, the noun condition does not constitute a baseline on a par with the pronouns. However, when the pronoun condition was contrasted with both reflexives and nouns at the same time in a complex orthogonal contrast, the difference is highly significant (pronouns are 92 ms slower than reflexives and nouns combined, t(44) = 3.04, p = .004). McKee et al found that pronouns and nouns both were significantly slower than reflexives. As is apparent, the main effect of condition is driven by the pronoun condition, where picture decisions are slower than both the reflexive and the name condition. Furthermore, the lack of interaction between group and condition shows that this pattern is the same for both groups. The main effect of group simply reflects that SLI children were significantly slower overall in their reaction time.

Hierarchical Linear Modeling

In order to also account for item as a random factor, as well as conducting more powerful tests of the relationships between subject-level properties such as CELF/TONI scores and the priming effect of reflexives, and for improving the estimates in general by taking into account intra-subject variance, we conducted a mixed effects analysis, using Hierarchical Linear Modeling (Raudenbush & Bryk, 2002). To develop and explain the HLM analysis, we consider first a model for a single subject, and then subsequently develop it to the full model of the set of subjects divided into two groups.

At the level of a single subject, we have 30 x 3 repeated measurements. For each of those 30 sentence sets, the single trial reaction times can be modeled as a "micro-regression equation". The following equation models the reaction times for the three sentences with the leopard as matrix subject and picture and dummy coding for two of the three conditions (with *I* = 1, 2, or 3):

(1) $RT_{i,leopard} = b_{0,leopard} + b_{1,leopard}(PRONOUN) + b_{2,leopard}(REFLEXIVE)$

This equation fits the original design where the noun condition was used as a baseline control condition, and the prediction was that the reflexive coefficient would be a significant predictor of change in reaction time relative to the intercept, whereas the pronoun coefficient would not be so. However, we already know that reflexives and nouns are both faster than pronouns. It would therefore be difficult to interpret the effect of the reflexive coefficient with noun as intercept. Rather, based on the observations we have already made, we used the slowest condition, pronoun, as intercept and test whether reflexives are indeed significantly faster than pronouns. Furthermore, although we don't have a hypothesis that nouns will differ from pronouns, we added noun as a coefficient to compare that reaction time to that of pronouns.

A single subject was then be modeled by 30 equations like (1), one for each sentence set *j*:

(2) $RT_{i,j} = b_{0,j} + b_{1,j}(\text{PRONOUN}) + b_{2,j}(\text{REFLEXIVE})$

This constitutes the Level-1 equation in the sense of HLM. From those thirty equations, three new regression equations were constructed for the thirty intercepts, noun coefficients and reflexive coefficients. Without any predictors at this level, we now had the following set of equations:

(3) Level 1: $RT_{ij} = b_{0j} + b_{1j}(\text{NOUN}) + b_{2j}(\text{REFLEXIVE})$

Level 2: $b_{0j} = g_{00} + u_{0j}$

$b_{1j} = g_{10} + u_{1j}$

$b_{2j} = g_{20} + u_{2j}$

By substituting the level 2 terms for the level 1 terms, we arrived at the combined equation describing a single subject:

(4) $RT_{ij} = g_{00j} + u_{0j} + g_{10j}(\text{NOUN}) + u_{1j}(\text{NOUN}) + g_{20j}(\text{REFL}) + u_{2j}(\text{REFL})$

Once each subject had an equation like this, a third level was constructed that modeled the set of g-intercepts and coefficients. At this level, categorical variables such as group (SLI or TD) and continuous variables (e.g., TONI-scores) were added for tests of interactions between between-subject and within-subject variables. For example, did being an SLI child significantly change the effect of the reflexive co-efficient? Did TONI score make a difference for the strength of the reflexive effect?

We analyzed the data with successive complexity by first testing a model with predictors only at Level-1, as if there were only one group of homogeneous subjects:

(5) $RT_{tij} = g_{000} + g_{100}*NOUN_{tij} + g_{200}*REFL_{tij} + r_{0ij} + r_{1ij}*NOUN_{tij} + r_{2ij}*REFL_{tij} + u_{00j} + u_{10j}*NOUN_{tij} + u_{20j}*REFL_{tij}$

In this model, the reflexive coefficient was significant (26 ms difference from pronouns; $t(44) = -2.51$, $p = .016$); whereas the noun coefficient was marginally significantly different from pronouns (19 ms, t(44) = -1.75, p = .086). We next added group as a coefficient at level 3, testing whether it significantly affected the intercept (i.e., the baseline RT for pronouns), as well as coefficients (i.e. whether there was a group X condition interaction). This resulted in a significant effect of SLI on the intercept (SLI children were 126 ms slower, $t(44) = 2.37$, $p = .022$). In addition, adding SLI as a predictor for nouns removed the marginal significance of the noun coefficient (now a 23 ms effect, but with $t(44) = 1.54$, $p = .13$, the coefficient was no longer significant). Furthermore, the interaction between language status (SLI vs. TLD) and the noun coefficient was not significant (SLI was 8 ms slower, $t(44) = 0.38$, $p = .7$). In other words, being SLI did not make the child significantly slower at the noun condition relative to the pronoun condition. Finally, the reflexive was now 26 ms faster than the pronoun, which was marginally significant by a two-tailed probability (t(44) = -1.87, p = 0.068). However, we predicted that the reflexive should be faster than the pronoun, so with a one-tailed probability, the reflexive coefficient is significant. Finally, adding SLI as a coefficient for the reflexive slope was not significant ($t(44) = 0.032$, $p = .97$). In other words, SLI children did not differ from TD children in showing a significantly faster reaction time to reflexives compared to pronouns.

This result reproduced the original pattern of predictions based on *having knowledge of Binding Theory* (McKee et al., 1993), where pronouns and nouns do not differ significantly, but where reflexives are processed faster than pronouns.

Finally, because we included subjects with high TONI scores, and because the two groups differed significantly on TONI scores, we tested whether TONI score interacted with the coefficients of interest. With TONI scores centered on the grand mean (i.e., with the mean as intercept), TONI had no effect on reaction time (mean RT = 857 ms, effect of TONI was less than 1 ms, $t(44) = -0.35$, $p > .05$). We then added TONI scores as a predictor at the level of subject in a model with reflexive and noun coefficients at level-1. Again, the reflexive slope was significant ($t(44) = -2.52$, $p = 0.015$), but the TONI scores showed no interaction with neither intercept nor coefficients. Non-verbal intelligence had no effect on processing differences between reflexives and pronouns. The same was true for the PPVT-scores.

Children with SLI exhibited the same pattern of priming reflecting knowledge of binding theory as their typically developing peers. This conclusion follows from a direct comparison of reaction times in the reflexive condition versus the pronoun condition. In other words, SLI children are not impaired in their tacit knowledge of binding theory, as shown by these on-line, real-time processing data. As the main effect of group on reaction time indicates, SLI children were significantly slower at computing this relation.

Discussion

The present study examined a core aspect of binding knowledge in children with SLI using an on-line cross-modal picture priming paradigm. Although children with SLI were generally slower in categorizing pictures (*alive/not alive*) than their typically developing peers, they exhibited the same pattern of priming as their typically developing peers. 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. This seems to place the locus of this and perhaps other sentence comprehension deficits in processing rather than in lack of grammatical knowledge.

A notable difference between our findings and those of McKee et al. (1993) is that in our ANOVA, the noun condition was not slower than the reflexive condition. This result appears to contradict their predictions of *knowledge of binding theory*, which were that reflexives should be faster than nouns, and pronouns should be the same as nouns, because reflexives (or activation of the correct antecedent) prime the picture decision, whereas pronouns and nouns do not. Thus, the reflexive condition was expected to be significantly faster than the pronoun condition: this in itself can be interpreted as being a reflection of priming in the reflexive condition as we found.

The HLM analysis, which controlled for variance in both items and subjects, revealed that the noun condition was indeed slower than the reflexive condition; thus, the noun condition did not function as a baseline control by being identical to the pronoun condition. We cannot account for the different results for the noun condition across the two studies, but the noun phrase may not have been the logically correct control condition. Whereas reflexives and pronouns require an antecedent, indefinite noun phrases (e.g., *a girl*) can be introduced without prior referent mention. The definite noun phrases (e.g., *the girl*) we adapted from McKee et al (1993) typically function as discourse anaphors (i.e., mentioned previously), but in these experimental contexts participants may have assumed that the referent had been introduced. Pronouns and reflexives are different; they *seek* and find an antecedent within the sentence and establish a referential relation with the antecedent (the reflexive to the closest subject, and the pronoun to the more distant subject).

If noun processing is governed by a system that is orthogonal to sentence internal reference relations, then there would be no specific prediction concerning the relative speed of noun phrase processing versus reflexive or pronoun processing. Alternatively, because reflexive and pronoun processing require the establishment of syntactic reference chains (i.e., syntactically represented binding relations), whereas noun phrase processing does not, nouns might be processed faster because fewer operations are required. But McKee et al. found the noun condition to be slowest for children and adults and we found that nouns had a faster RT than pronouns but equivalent to reflexives. One underlying issue is that two RTs may reflect a variety of processes even within the same task. For example, the noun condition may have yielded a faster reaction time 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. The critical comparison, for the purposes of investigating binding knowledge, was between reflexives and pronouns. These two elements are governed by mirror-image conditions (Principles A and B), whereas the regulation of nouns phrases, R(referential)-expressions, are governed by Principle C, a qualitatively different constraint.

Recent proposals concerning sentence comprehension point to underlying processes such as interference that 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). Notably, interference, binding of stimuli and responses, and release from binding also play a role in some recent characterizations of working memory, executive functions, and the microstructure of phenomena such as remembering and forgetting (Hasher, 2007; Lewis, Vasishth, & Van Dyke, 2006; Oberauer & Lange, 2009; Van Dyke & Johns, 2012). Children with SLI have demonstrated deficits in suppressing irrelevant information, have less resistance to proactive inhibition, a higher rate of interference errors, and have poorer coordination between activation and inhibition than their typically developing peers (Epstein, Hestvik, Shafer, & Schwartz, 2013; Marton, Campanelli, Eichorn, Scheuer, & Yoon, 2014)

When listeners encounter anaphora, all potential antecedents (i.e., all nouns in the sentence) are activated, and all but the correct antecedent must 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 more limited binding and release from binding, 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 permit exploration of interference and binding effects in the comprehension of pronouns, reflexives and other structures. Although cross-modal priming paradigms have the advantage of examining sentence processing in real-time, they only examine information activation at a single point in time. Methods that yield continuous examination of information activation such as eye tracking or event-related potentials may be better suited to these studies.

Acknowledgements

The research and preparation of the manuscript were supported by Grants RO1 DC003885 and R01DC011041 from the NIDCD. The authors extend their appreciation to Stacy Asay who played a significant role subject recruitment and data collection.
Correspondence should be addressed to Richard G. Schwartz, Ph.D., Ph.D. Program in Speech-Language-Hearing Sciences, The Graduate Center, City University of New York, 365 Fifth Avenue, New York, NY 10016. Email: RSchwartz@gc.cuny.edu.

References

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

Brown, L., Sherbenou, R., & Johnson, S. (1997). Test of Nonverbal Intelligence (3 ed.). 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.

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 (4 ed.). Bloomington, MN: Pearson, Inc.

Elbourne, P. (2008). THE ARGUMENT FROM BINDING. *Philosophical Perspectives, 22*(1), 89-110. doi: 10.1111/j.1520-8583.2008.00142.x

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*(4), 351-365.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *J Exp Psychol Learn Mem Cogn, 32*(6), 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. *Linguiistic Inquiry, 21*, 187-222.

Hasher, L. L., C.; Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. C. C. J. M. K. J. Towse (Ed.), *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*(5), 443-456. doi: 10.1007/s10936-010-9151-1

Leonard, L. B. (2014). *Children with specific language impairment* (Second 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*(1), 5-24.

Lust, B. (1986). *Studies in the acquisition of anaphora* (Vol. 1). Dordrecht: 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*. doi: 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: A Journal of Developmental Linguistics, 1*(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.

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*(1), 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 Publications, Inc.

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: Vol. I. Defining the constraints* (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), 413a.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002a). *E-Prime Reference Guide*. Pittsburgh: Psychology Software Tools, Inc.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002b). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools, Inc.

Schwartz, R. G. (2009). Specific language impairment. In R. G. Schwartz (Ed.), *Handbook of Child Langauge Disorders* (pp. 3-43). New York, NY: Psychology Press.

Semel, E., Wiig, E. H., & Secord, W. A. (2004). Clinical evaluation of language fundamentals (4 ed.). Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.

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. R. E. Williams (Ed.), *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*(3), 245-290.

Van Dyke, J. A., & Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Language and Linguistics Compass, 6*(4), 193-211. doi: 10.1002/lnc3.330

Table 1. Participant Descriptors

Descriptors	TLD (n = 24)	SLI (n = 22)
Age	9;4 (0;11 / 7;7 - 10;11)	8;11 (0;11 / 7;3 - 10;9)
CELF-R	108 (14 / 77 - 137)	82 (11 / 63 - 103)
CELF-E	112 (13 / 77 - 137)	75 (13 / 53 - 102)
CELF	111 (11 / 87 - 140)	75 (11 / 50 - 93)
PPVT-IV	107 (16 / 84 - 150)	87 (11 / 72 - 106)
TONI	107 (16 / 84 - 150)	99 (12 / 83 - 123)

Note: All descriptors except for age are mean standard scores (*SD* / Range). TLD = children with typical language development; SLI = children with specific language development; 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.*

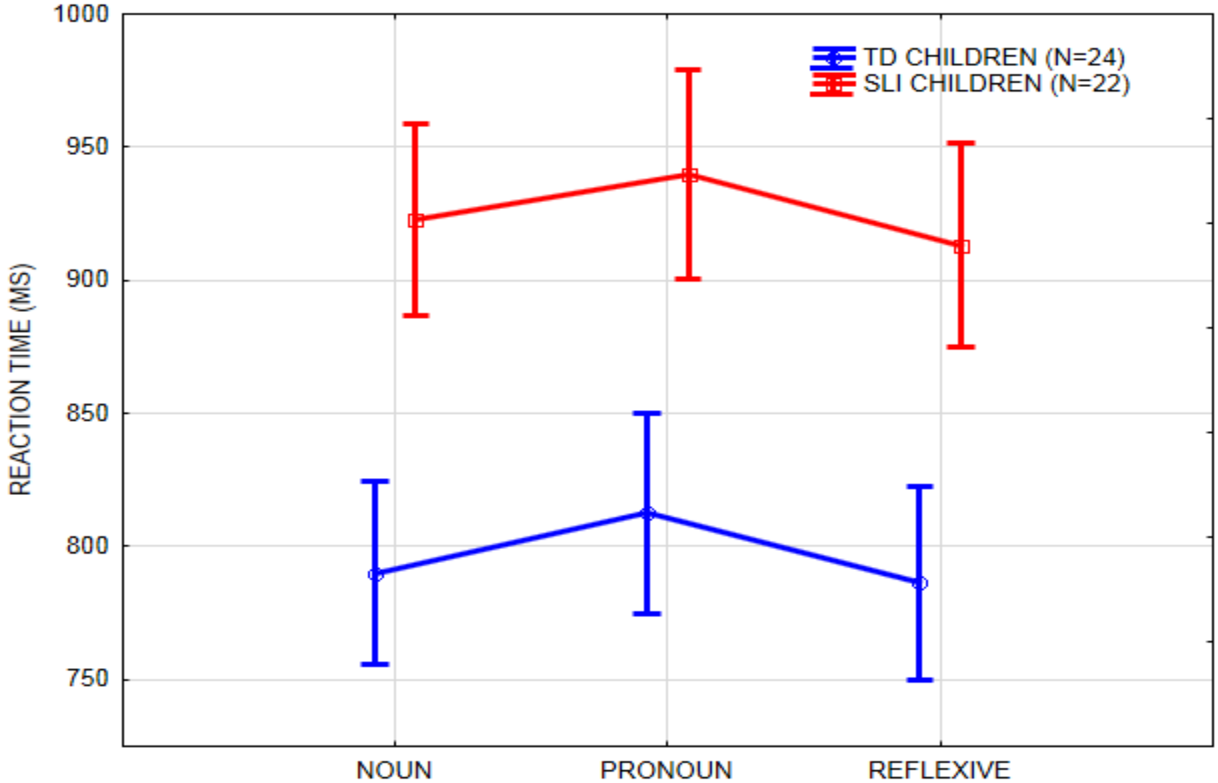

Figure 1: Main effects of group and condition. Error bars represent the 95% confidence interval.

Arild
Dele