

Equal Treatment for All Antecedents: How Children Succeed with Principle B

Anastasia Conroy

Eri Takahashi

Jeffrey Lidz

Colin Phillips

Children have repeatedly been found to exhibit Principle B violations, with some reports that these violations occur only with nonquantified antecedents. This quantificational asymmetry (QA) in the delay of Principle B effect (DPBE) has been taken as support for a theory that restricts the scope of binding theory to bound variable anaphora (Reinhart 1983). However, the QA has been challenged, on the basis of discrepant findings and methodological concerns (Elbourne 2005). Here, we resolve the status of the QA with 3 studies and a review of over 30 previous studies. Using improved experimental materials, we show that children disallow local pronoun binding with both referential and quantificational antecedents when Principle B is at issue (Experiment 1), but not when Principle B is neutralized (Experiment 2). When methodological flaws are reintroduced, we replicate the QA (Experiment 3). Drawing on evidence from adult language processing, we suggest that the role of Principle B as a filter on representations during sentence understanding, in concert with pragmatic infelicities in the tasks used, accounts for the wide variability in the DPBE in the literature.

Keywords: anaphora, binding, coreference, Principle B, quantificational asymmetry, language acquisition

1 Introduction

It is rare that data from child language are taken to constrain models of adult grammatical competence. One such case concerns an asymmetry in children's application of Principle B, a constraint that prohibits local antecedents for pronouns. Many studies, spanning the past 25 years, have reported that 4- to 6-year-old children allow a nonadult interpretation of (1) that is equivalent to *Mama Bear washed herself*.

(1) Mama Bear washed her.

However, when the referential subject in (1) is replaced with a quantificational subject, as in (2), children no longer allow the corresponding anaphoric interpretation.

We are grateful to two anonymous reviewers, Paul Elbourne, Nina Hyams, and many members of the Maryland CNL (Cognitive Neuroscience of Language) Lab for valuable feedback on the issues in this article. This work was supported in part by NSF awards BCS 0196004 to Colin Phillips and BCS-0412809 to Jeffrey Lidz, by NIH award DC-006829 to Jeffrey Lidz, and by a fellowship from the Heiwa Nakajima Foundation to Eri Takahashi.

(2) Every bear washed her.

The observation that, with some types of antecedents, young children are delayed in demonstrating knowledge of Principle B has been termed the *delay of Principle B effect* (DPBE); and the observation that children display knowledge of Principle B with quantificational but not referential antecedents has been termed by Elbourne (2005) the *quantificational asymmetry* (QA).

The theoretical argument, due originally to Chien and Wexler (1990), concerns the scope of Principle B in the grammar. Chomsky's (1981) binding theory treats all cases of anaphora as involving coindexation, thus positing no difference between the treatment of quantificational antecedents and the treatment of referential antecedents with respect to binding. However, Reinhart's (1983) approach distinguishes bound variable anaphora from other cases of coreference. Consequently, the quantificational asymmetry in children's interpretations is taken to mirror the asymmetry that is independently posited in Reinhart's binding theory. In this case, the results from child language acquisition have been taken as strong evidence for a distinction between bound variables and coreferring pronouns in the grammar.

The QA has been an important finding in the study of child language because it appears to decide among leading views of anaphora, lending crucial support to Reinhart's theory. However, Elbourne (2005) has raised several concerns regarding the validity of the QA, although without providing supporting developmental evidence. In this article, we investigate children's knowledge of Principle B, particularly with respect to the methodological concerns raised by Elbourne. We suggest that many of the prior observations concerning the QA and the DPBE reflect shortcomings of the experimental tests used and do not reflect properties of children's grammars. We present a series of three novel experiments designed to examine children's knowledge of Principle B, addressing Elbourne's methodological concerns. We find that appropriately controlled experiments appear to eliminate the QA, a finding that partly vindicates Elbourne's claims. However, we find that 4-year-old children overwhelmingly respect Principle B in sentences with referential and quantificational antecedents alike, disconfirming Elbourne's prediction that children should display a DPBE for all antecedents.

We claim that the QA is an artifact of experimentation, and we suggest that the DPBE is less pervasive than is standardly reported. Children do, in fact, obey Principle B. Our studies therefore remove one of the arguments in favor of Reinhart's theory of binding, although we remain neutral as to whether this theory is actually correct (for recent reviews, see Buring 2005, Elbourne 2008).

Interestingly, our findings align Principle B more closely with other constraints, such as Principle C, a constraint on backward anaphora that rules out coreference in sentences like (3). This constraint has repeatedly been shown to be mastered quite early (Crain and McKee 1986, Crain and Thornton 1998, Guasti and Chierchia 1999/2000, Kazanina and Phillips 2001, Leddon and Lidz 2006).

(3) She washed Mama Bear.

We conclude that there is no QA and that at the level of grammatical representation there is no DPBE. However, an open question remains. If it is the case that children adhere to Principle B,

a grammatical constraint that bans coreference between a pronoun and a local antecedent, then why are they highly susceptible to errors under certain experimental conditions? One might expect that if children's grammars disallow the representation associated with a Principle B violation, then even biased methodologies should not cause them to access interpretations that violate Principle B. We argue that children's greater susceptibility to Principle B errors (over Principle C errors) derives from independently motivated properties of anaphoric dependency processing that are revealed in adult online studies on Principles B and C. This argument, in concert with our findings from children, removes a potential impediment to grammatical theories that give parallel accounts of Principles B and C.

In section 2, we summarize two lines of evidence for the distinction between bound variable anaphora and coreference, drawing on classic theoretical and developmental arguments. We also highlight the contrasting experimental findings regarding children's mastery of Principles B and C. In section 3, we review the methodological assumptions behind widely used tests of children's grammars generally, and Principle B in particular. In section 4, we present three experiments that address the methodological concerns raised about previous experiments. In these experiments, we find no evidence for a QA or for a DPBE. These new results leave us with the question of why findings about children's knowledge of Principle B are so discrepant. In section 5 and appendix A, we review over 30 previous studies on binding constraints in children. We relate these findings to recent studies of binding constraints in real-time language processing in adults, concluding that although children's grammars are apparently intact, they show exaggerated susceptibility to illicit antecedents that are also fleetingly considered in online studies with adults.

2 Asymmetries in Binding Theory

2.1 *Two Types of Anaphoric Relation*

At bottom, the debate over binding theory centers around the formal mechanisms underlying anaphora in the grammar and begins with the discovery that sentences like (4) are three-ways ambiguous, not two.

(4) Al loves his sister.

In one reading, the pronoun refers to a sentence-external antecedent. However, even if we require that the pronoun be anaphoric to the subject of the sentence, an ambiguity remains. The ambiguity can be seen more clearly when we place such sentences in a VP-ellipsis context, as in (5) (Keenan 1971, Sag 1976, Williams 1977).

(5) Al loves his sister and Bill does too.

- a. = Al_i loves his_i sister and Bill_j loves his_j sister too. ('sloppy' reading)
- b. = Al_i loves his_i sister and Bill_j loves his_i sister too. ('strict' reading)

We can interpret the second conjunct as meaning that *Bill loves Bill's sister* or as meaning that *Bill loves Al's sister*. This simple paradigm tells us that there is more than one way for a pronoun to be connected to its antecedent. On the one hand, the pronoun may be treated as a bound

variable, whose interpretation is determined by its antecedent. This gives rise to the “sloppy” interpretation (5a), because the elided VP and the overt VP each contain a bound variable pronoun bound by the subject of the corresponding clause. Alternatively, the pronoun may be understood to have fixed reference that happens to match the reference of the subject of the first clause. This yields the “strict” interpretation (5b). In this case, the elided VP, like its overt counterpart, contains a pronoun that refers to the subject of the first clause. The pronoun in the first clause corefers with the subject of the clause, but unlike a bound variable it is not directly dependent on the subject for its reference. This type of coreference is sometimes called “accidental coreference.”

Further evidence for the ambiguity between bound variable anaphora and coreference comes from cases in which a pronoun takes a quantificational antecedent, as in (6).

- (6) Every linguist loves his sister and Bill does too.
 a. = Every linguist_i loves his_i sister and Bill_j loves his_j sister too. (sloppy)
 b. ≠ Every linguist_i loves his_i sister and Bill_j loves his_j sister too. (strict)

Because quantifiers are not referential, no accidental coreference is possible. Consequently, no strict reading is possible in the VP-ellipsis context (6b). The overt pronoun can only be connected to a quantificational antecedent as a bound variable, and consequently the elided VP must also contain a bound variable, as in (6a). In sum, examples like those in (5) and (6) show that there are two mechanisms by which a pronoun may be linked to its antecedent: one involving variable binding and the other involving coreference (Keenan 1971, Sag 1976, Williams 1977, Evans 1980, Higginbotham 1983).

Reinhart (1983) observes further that the syntactic conditions on bound variable anaphora are stricter than those on accidental coreference. In particular, a bound variable must be c-commanded by its antecedent (7a) (but cf. Büring 2005, Elbourne 2008), whereas no such restriction holds for accidental coreference (7b).

- (7) a. The people who work for him_{*i/j} love every department chair_i.
 b. The people who work for him_{i/j} love AI_i.

Similarly, in VP-ellipsis contexts a pronoun with a non-c-commanding antecedent supports only a coreferential reading and hence only allows the strict reading of the elided VP (8).

- (8) The people who work for AI love him and the people who work for Bill do too.
 a. = . . . and the people who work for Bill love AI.
 b. ≠ . . . and the people who work for Bill love Bill.

2.2 *The Scope of Binding Constraints*

Reinhart (1983) argues that binding theory should apply only to bound variable anaphora, and not to anaphoric interpretations in general. In this section, we summarize one theoretical consideration and an empirical argument that is particularly relevant to the developmental issues that we focus on here.

A primary theoretical motivation for Reinhart's approach comes from the observation that the disjoint reference constraints (i.e., Principles B and C; Chomsky 1981) and bound variable anaphora apply only to c-command relations. Principle B rules out a pronoun with a local antecedent only if c-command obtains between them (9). Similarly, Principle C rules out an R-expression (i.e., nonpronominal NP) that is c-commanded by a coreferential NP (10).

- (9) a. Al_i likes $him_{*i/j}$.
 b. Al_i 's sister likes $him_{i/j}$.
- (10) a. $He_{*i/j}$ thinks that I like Al_i .
 b. $His_{i/j}$ sister thinks that I like Al_i .

This parallel in the domain of applicability is expected if the binding constraints apply only to bound variable anaphora.

The idea that binding constraints apply only to bound variable anaphora correctly predicts that the bound variable interpretation of the pronoun in (11) is blocked by Principle B. However, the prediction is more complicated for sentences like (12) that have a referential subject NP.

- (11) Every candidate_{*i*} likes $him_{*i/j}$.
 (12) Al_i likes $him_{*i/j}$.

Recall that a referential NP may serve as the antecedent of a pronoun either via variable binding or via accidental coreference. Therefore, if accidental coreference is generally available and if Principle B regulates only the distribution of bound variables, then it follows that the accidental coreference interpretation should be available in (12), yielding the interpretation that *Al likes himself*, contrary to speaker judgments.

Reinhart argues that this prediction is not a shortcoming of her theory, but rather is a virtue. She claims that the accidental coreference representation is indeed available, but that special discourse circumstances are required to realize this possibility (Evans 1980, Higginbotham 1983). In a sentence like (13), for example, it is argued to be natural to interpret the pronoun *him* as coreferential with *Bill*, even though it is locally c-commanded by its antecedent in the last conjunct. Therefore, it is important to explain what distinguishes (12) from (13).

- (13) I know what Mary, Sue, and Bill have in common. Mary likes him, Sue likes him, and Bill likes him too.

Reinhart recognizes that her version of Principle B is insufficient to block coreference in (12), owing to the possibility of accidental coreference.¹ She proposes that this binding-theoretic loophole is closed by an additional constraint, labeled Rule I (14).

¹ Indeed, it is precisely this issue that led to Lasnik's (1976) argument that the constraints on pronominalization must be stated in terms of disjoint reference and not coreference. However, Evans (1980) argues that even this restriction is insufficient to block accidental coreference. For alternative accounts of how to distinguish the coreference possibilities in (12) and (13), see Heim 1998 and Levinson 2000.

(14) *Rule I: Intrasentential Coreference* (Grodzinsky and Reinhart 1993:79)

NP A cannot corefer with NP B if replacing A with C, C a variable A-bound by B, yields an indistinguishable interpretation.

In essence, Rule I is an economy condition, stating that accidental coreference is possible only when bound variable anaphora is not. This rule successfully blocks accidental coreference in (12), because the bound variable interpretation and the accidental coreference interpretation have identical truth conditions. Furthermore, Rule I provides an account of why coreference is possible in sentences like (13). The relevant interpretation of (13) asserts that the property shared by Mary, Sue, and Bill is the property of liking Bill—that is, (15).

(15) $\lambda x. x$ likes Bill

However, if the pronoun in the third conjunct is taken to be a bound variable, then that clause asserts that Bill is a self-liker—that is, (16).

(16) $\lambda x. x$ likes x

It is clear that liking oneself and liking Bill are different properties for the members of a group to share, and therefore Rule I does not apply. Consequently, accidental coreference is not trumped by bound variable anaphora in (13) and hence the third conjunct is predicted to allow the interpretation that *Bill likes himself*.

The critical conclusion from this line of argument is that cases of coreference failure like (12), which under classic binding theory (Chomsky 1981) were taken to be violations of Principle B alone, are considered to be ruled out by two separate principles of grammar: Principle B (a condition on the syntax-semantics interface) and Rule I (a condition on the semantics-pragmatics interface).

2.3 *A Developmental Dissociation*

While the theoretical arguments in favor of restricting Principle B to cases of bound variable anaphora can stand on their own, evidence from the developmental pattern of adherence to Principle B is often presented as the best possible evidence for the existence of two different mechanisms underlying Principle B effects (Chien and Wexler 1990, Grodzinsky and Reinhart 1993, Thornton and Wexler 1999). This evidence comes from an apparent dissociation in children's interpretation of sentences traditionally captured by Principle B. Many studies of English and other languages, most notably Dutch and Russian, have reported that children incorrectly allow local binding of a pronoun with a referential antecedent until roughly age 5, but a number of these studies have reported no such delay in sentences with quantificational antecedents (see section 5). For example, in a classic picture judgment study (Chien and Wexler 1990), 5-year-olds accepted a coreferential interpretation for sentences like *Mama Bear is touching her* on 51% of trials, but for sentences like *Every bear is touching her* they accepted it on only 16% of trials.

This developmental dissociation, which Elbourne (2005) calls the quantificational asymmetry (QA), appears to provide striking support for Reinhart's account of binding theory. The theory

claims that different mechanisms restrict quantificational and referential antecedents for pronouns, and the dissociation observed in the child data comports well with this claim. If the children have already mastered Principle B, but either do not know or cannot apply the additional constraint that blocks local anaphora with referential antecedents, then the QA is captured.²

However, there are a number of concerns about the strength of the theoretical conclusions that can be drawn from reports of the QA. First, although the QA has become a part of the received wisdom about language acquisition (e.g., Guasti 2004), the empirical record is not unequivocal (for reviews, see Kaufman 1994, Koster 1994, Elbourne 2005). A number of studies have obtained discrepant findings, regarding both the QA and the strength of the delay of Principle B effect (DPBE). For example, studies of the DPBE have found rates of acceptance of Principle B violations that range from 16% to 82%—far greater variability than would be expected by chance. We discuss the previous literature in more detail in section 5.

Second, examination of the studies that have been used to show the QA raises concerns about the adequacy of the experimental designs used in these studies. Some of these concerns were already raised by Elbourne (2005), and we raise a number of additional concerns in section 3.

Third, the developmental findings do not comport as well with Reinhart's theory of binding as is sometimes suggested. The theoretical arguments that we reviewed for Principle B can also be constructed for Principle C. Indeed, on Reinhart's theory Principle C effects are governed strictly by Rule I. Consider the dialogue in (17).

(17) A: Is that John?

B: It must be. He's wearing John's coat.

Here speakers do not detect a Principle C violation, even though *he* and *John* are taken to be identical in reference. Rule I is satisfied, however, since an assertion that *someone is wearing John's coat* is distinct from an assertion that *someone is wearing his own coat*. The fact that these assertions can be distinguished licenses coreference.

Thus, we would expect that children's failure to apply Rule I, giving rise to apparent Principle B violations, should also give rise to apparent Principle C violations. But this appears not to be the case. Instead, a number of studies have found that Principle C is uniformly obeyed by children at age 4 and even younger (Crain and McKee 1986, Crain and Thornton 1998, Guasti and Chierchia 1999/2000, Kazanina and Phillips 2001, Leddon and Lidz 2006). Thus, to the extent that the QA provides evidence for the necessity of a theoretical construct like Rule I, children's success with Principle C remains a mystery. Put differently, to the extent that the developmental evidence shows an asymmetry between Principle B and Principle C, it casts doubt on the theoretical claim that Principles B and C are governed by common mechanisms. This concern applies not just to Reinhart's theory of binding, but to any approach that treats the disjoint reference rules similarly.

² The literature contains a number of different accounts of the specific cause of the DPBE. For example, Reinhart has argued that children have full adult knowledge, but are unable to perform the computations needed to implement Rule I (Grodzinsky and Reinhart 1993, Reinhart 2006), whereas Wexler and his colleagues have argued that children lack adult knowledge of a pragmatic principle (Principle P) that restricts accidental coreference (Chien and Wexler 1990, Thornton and Wexler 1999). The difference between these accounts does not affect our arguments here.

With these issues in mind, the next section aims to identify the features of a fair test of the QA and the DPBE, using as a focus the truth-value judgment task (TVJT; Gordon 1996, Crain and Thornton 1998), because it is a task that encourages experimenters to make very explicit their assumptions about the discourse context against which a sentence is judged. Although many of the issues discussed below generalize to other experimental measures, and although we believe that the concerns we raise encompass a variety of tasks, we are not able to examine all tasks in as much detail as we do the TVJT. We next turn to a discussion of the basic components of a TVJT test of constraints on anaphora.

3 Truth-Value Judgment Tests

3.1 *The Logic of the Task*

Much of the evidence for the delay of Principle B effect (DPBE), and for the quantificational asymmetry (QA) in particular, has been drawn from studies using the truth-value judgment task (TVJT), a task that confers many advantages but that requires great care in its use and interpretation. In this section, we briefly summarize the desiderata for a fair test of children's knowledge of constraints on binding and coreference, and we raise concerns about how well these desiderata have been satisfied in previous studies, expanding upon the critique by Elbourne (2005). Our discussion here focuses on the TVJT, but similar considerations apply to other experimental tasks that have been used to test children's knowledge of binding constraints.

Suppose that we want to know what interpretations 4-year-old children allow for sentences like (18) and (19), which have been the focus of most research on the DPBE and the QA.

(18) Grumpy painted him.

(19) Every dwarf painted him.

Our interest is in whether children respect the constraint that prevents the pronoun *him* from being anaphoric to a local subject NP in the adult grammar, and whether this constraint affects the two examples equally. In what follows, we use the term *anaphoric* as a cover term for variable binding and (accidental) coreference, reserving the terms *bound* and *coreferential* for the two specific types of referential dependencies. We refer to interpretations in which the pronoun lacks an intrasentential antecedent as *deictic*. We also describe cases of illicit coreference and cases of illicit bound variable anaphora as *Principle B violations*, with no intended prejudice regarding whether these should be handled by one or two mechanisms in the grammar.

Clearly, we cannot ask young children to give explicit judgments about the range of coreference possibilities for sentences like (18) and (19). The TVJT was thus devised to tackle the problem of probing complex grammatical phenomena in young children using a task that is engaging and that requires simple judgments from children (Gordon 1996, Crain and Thornton 1998). In a TVJT, a child and a puppet, such as Kermit the Frog, together watch an experimenter tell a story with props. After the story, the puppet makes a statement about the story and the child's task is to judge whether the statement was accurate. The experimenter can use the child's judgment to draw inferences about the child's interpretation of the target sentence. The task has

many advantages. It can be used to probe complex grammatical representations in young children (most children aged 4 years and up, and some 3-year-olds). It is engaging and nonconfrontational, and it has special advantages for investigating sentences that have multiple potential interpretations. Another advantage of the task is that the test sentence is presented in the context of a discourse, thereby allowing experimenters to manipulate the discourse context and control for its potential effect on sentence interpretations.

When the TVJT is used as a test of binding constraints, its core logic is straightforward: if a child encounters a story in which the anaphoric interpretation of the pronoun in (18) or (19) is both true and prominent in the story, but does not judge (18) or (19) to be a true statement about the story, then the child presumably did not access the anaphoric interpretation of the pronoun. Assuming that the task has been executed properly, we infer that the reason the child did not access the anaphoric interpretation is that a binding constraint made that interpretation inaccessible.

However, the simple judgment that the TVJT requires of a child also creates its greatest challenge. The child makes a judgment about a sentence and a story, and then the experimenter must make an inference concerning the child's grammar. Because the setup of the story is integral to the range of interpretations available to the child, great care must be taken to satisfy the assumptions underlying the task, so as to avoid misleading results (Crain and Thornton 1998).

A widespread assumption in TVJT studies is that children will assent to the truth of a sentence if they can ('Principle of Charity'). In the case of a potentially ambiguous sentence, a test sentence is typically presented in a scenario that makes only one reading of the sentence true. Thus, if a child assents to the truth of the sentence in a scenario, we presume that the child has access to the interpretation made true in that scenario. If the child rejects the sentence in a scenario, then this rejection is taken as evidence that the interpretation made true in that scenario is unavailable. However, such responses justify conclusions about the child's grammar only if we can rule out extralinguistic reasons for the responses. If an interpretation is too strongly biased in the scenario provided, then a child may say 'yes' owing to contextual coercion of an ungrammatical interpretation. Conversely, if an interpretation is not made sufficiently available in the story, then it is possible that the child's rejection results not from grammatical constraints, but from properties of the discourse context. Consequently, if we are to reason about a child's grammar from yes/no responses in a TVJT, then we must ensure that the interpretations under investigation are made equally available in the experimental contexts.

Following this logic, TVJT tests of binding constraints rely on the assumption that the test scenario makes two different interpretations accessible (though not necessarily true): one interpretation corresponds to the anaphoric interpretation of the pronoun, and the other interpretation corresponds to the deictic interpretation of the pronoun. Children then judge the truth of a potentially ambiguous sentence, and the researcher can use the children's judgments to infer which interpretation of the sentence was accessed. If children's judgments show that they systematically fail to access one interpretation, then the researcher may conclude that the children do not allow that interpretation of the test sentence.

However, it is important to recognize that when a child rejects a sentence like (18) or (19) in a TVJT scenario where the anaphoric interpretation is true, the child is not a 'little linguist'

who is directly judging the anaphoric interpretation as ungrammatical. Rather, it is assumed that the child is denying the truth of an alternative, deictic interpretation of the pronoun. When a child does this, it is commonly assumed that the child focused on the deictic interpretation because the anaphoric interpretation was unavailable, and furthermore that it was the child's grammar that made the anaphoric interpretation unavailable. However, this depends on the assumption that the anaphoric interpretation was sufficiently available in the context that only the child's grammar could be responsible for his rejection.

Let us now consider how the assumptions of the TVJT must be satisfied in the context of a Principle B experiment, and the implications for specific experimental designs. At the most general level, these assumptions amount to the need to balance the relative accessibility of the interpretations under investigation. Here it is important to distinguish two notions of accessibility of pronoun interpretations. One applies to the potential referents of pronouns, the other to the propositions that the pronouns appear in.³

Assumption 1: Availability. Pronouns are used to pick out referents that are independently available in a discourse. Therefore, in order to test whether a child's grammar allows or disallows a particular antecedent for a pronoun, it is important that the intended referent for the pronoun be available in the current discourse. This requirement applies equally to antecedents of anaphoric and deictic pronouns, but in tests of binding constraints it is particularly relevant to deictic antecedents, since they need not be explicitly mentioned in the same sentence as the pronoun and hence may be overlooked. If a child is presented with a sentence containing a pronoun for which his grammar excludes an anaphoric interpretation, but the context fails to make a deictic interpretation available, then the child may be "coerced" into choosing the anaphoric antecedent because that is the only discourse-accessible antecedent.

Relatedly, it is important to establish that potential anaphoric antecedents are considered to be potential antecedents by children, once the possible effects of binding constraints are neutralized. This is particularly relevant for tests of bound variable anaphora. If children are reluctant to allow bound variable interpretations for pronouns in general, as has sometimes been suggested (Koster 1994), then they may avoid the anaphoric reading of (19) because of this general bias, independent of Principle B.

Assumption 2: Disputability. The second requirement involves the propositions in which pronouns are used in tests of binding constraints. In order to be submitted for a natural true/false judgment, a proposition should be "under consideration" in the experimental setting. If an interpretation has never been under consideration, then children may have difficulty rejecting that interpretation, even if it refers to an event that did not occur. For example, if a sentence like (19) is presented in a context where the narrative focus is on whether *the dwarves will all paint another individual* (deictic interpretation), and where the possibility that *each of the dwarves might paint*

³ We would like to thank an anonymous reviewer for helping to clarify these issues.

himself is never a live option in the story (anaphoric interpretation), then children are likely to judge the truth of the deictic interpretation, regardless of the effect of binding constraints.

In a TVJT test of binding constraints, children are typically asked to judge sentences containing pronouns following stories in which an anaphoric interpretation of the pronoun is true and a deictic interpretation is false. Practically, this entails that the story should make the deictic interpretation a genuine potential outcome at some point in the story. Crain and colleagues have emphasized the importance of this requirement, which they attribute to Russell (1948); they label it the *Condition of Plausible Dissent* in TVJT designs (Crain and Thornton 1998). Previous studies of children's adherence to Principle B have satisfied—or failed to satisfy—this requirement in a number of ways, and we argue below that this factor is important in understanding variability in past findings. The requirement is commonly satisfied by making the deictic interpretation of the sentence almost become true in the story, although we argue that this is insufficient if the event that almost happens is incidental to the rest of the narrative (see also Hulsey et al. 2004).

In tests of the QA, there is one further consideration that must be taken into account: the scenarios should be maximally similar in the referential and quantificational conditions. If the conditions are poorly matched in this regard, then a spurious QA may be observed. This means that the child's basis for rejecting the test sentences should be maximally similar in the referential and quantificational conditions. The reasoning that we follow in the new studies described below is that if the very same event makes the test sentence false in the referential and quantificational conditions, then the risk of a spurious QA is reduced.

In addition to constraints on the setup of a TVJT scenario, it is important that researchers are confident in their measures of the interpretation accessed by children. For this reason, children's yes/no answers should ideally be followed with requests for explanation. Although most descriptions of the TVJT focus on using children's yes/no answers to draw inferences about their grammatical representations, the real interest lies in how children interpret the test sentences—and for these interpretations, the yes/no answers are but one convenient measure. Children's justifications for their answers provide important additional information about their interpretations, and they sometimes indicate that the yes/no answer is misleading. For example, if a child answers “no,” but his explanation indicates that he interpreted the pronoun anaphorically, then his “no” answer clearly provides no evidence for avoidance of anaphoric interpretations.

Elbourne's (2005) critique of prior studies of the DPBE focuses primarily on one aspect of how these assumptions are satisfied, emphasizing factors such as differential availability of the anaphoric antecedent that may have led to the appearance of a QA. Experiments 1 and 2 below address these concerns by closely matching the test conditions for referential and quantificational conditions and by independently verifying that children can access bound variable interpretations. Issues of disputability, and in particular the question of whether the (ultimately false) deictic interpretation of the test sentence was under consideration in the scenario, are less of a focus in Elbourne's critique, but we suggest that inadequate satisfaction of this assumption may be at least as important as failures to satisfy the availability assumption in tests of Principle B and the QA in children. We show in Experiment 3 that when the disputability assumption is not properly satisfied in the referential condition, the appearance of an asymmetry arises.

3.2 An Example: Thornton and Wexler 1999

We next consider in more detail whether the assumptions outlined above are satisfied in one sample TVJT scenario that has been used as evidence for both the DPBE and the QA. The scenario is drawn from Thornton and Wexler 1999. We focus on this example not because it is better or worse than others, but because the authors provide a detailed description of the study and because it is representative of a design strategy that has been followed in several other studies of the DPBE and the QA, as discussed further in section 5 and appendix B.

The story in (20) gives the outline of one scenario that Thornton and Wexler used to test sentences with referential and quantificational antecedents alike, following a standard TVJT procedure. The child's task was to judge Kermit's statements in (21) and (22).

- (20) Bert and three reindeer friends have a snowball fight, and they all get covered in snow. When they go inside, Bert is shivering, so he asks the reindeer to brush the snow off him. Two of the reindeer (separately) refuse, saying they have too much snow to deal with, and they brush themselves. The third reindeer helps Bert a little bit, but then brushes the snow off of himself. Bert thanks the helpful reindeer for starting to brush him. He says he's sorry he can't reciprocate by helping brush the helpful reindeer; he needs to finish brushing all the snow off of himself because he's still very cold. (Thornton and Wexler 1999:142)

- (21) I think Bert brushed him. *Referential condition*
 (22) I think every reindeer brushed him. *Quantificational condition*

The sentences in (21) and (22) are both true in the scenario in (20) only if the pronoun *him* is interpreted anaphorically, in violation of Principle B. It is true that *Bert brushed himself*, and also true that *every reindeer brushed himself*, but neither of these interpretations is acceptable for adults. Thornton and Wexler report that children accepted (21) as true in 58% of trials and accepted (22) as true in only 8% of trials. These results suggest a strong DPBE for referential antecedents and a clear QA.

The scenario in (20) meets two basic desiderata of a test of binding constraints in children: the anaphoric interpretation of the pronouns in (21)–(22) is true in the story, and there is a deictic interpretation of the pronoun that is false in each case. Furthermore, the Condition of Plausible Dissent is at least partly met in the quantificational and referential conditions. However, when we consider more systematically how the availability and disputability assumptions are satisfied in the two conditions, we find contrasts that may account for the observed QA, independent of Principle B.

The availability assumption is differentially satisfied in the quantificational and referential conditions, owing to the contrasting accessibility of the anaphoric and deictic antecedents assumed in each condition. In the referential condition (21), the anaphoric antecedent is *Bert* and the relevant deictic antecedent is *the third reindeer*, whose help Bert briefly considers reciprocating. In the quantificational condition (22), on the other hand, the anaphoric antecedent is *every reindeer* and the relevant deictic antecedent is *Bert*. Because Bert is clearly the main protagonist of the story, and

because he is the anaphoric antecedent in the referential condition and the deictic antecedent in the quantificational condition, it is perhaps not surprising that children's judgments were based on the anaphoric interpretation on most referential trials and the deictic interpretation on most quantificational trials. Thus, the QA can be derived by assuming that children associate the pronoun *him* with the most prominent referent in the story, with no need to appeal to Principle B.

Elbourne (2005) points to the concern that the main protagonist (Bert) plays a different role in the two conditions, and also raises the concern that children may have favored the deictic interpretation in the quantificational condition (22), perhaps because of a general dispreference for bound variable interpretations of pronouns, an assumption that we directly test in Experiment 2.⁴

Turning to the disputability assumption, we again find that the referential and quantificational conditions differ in how the test sentence relates to the central theme of the story (who will brush Bert?), and do so in a way that could explain children's contrasting responses in the two conditions. In the referential condition, the anaphoric interpretation corresponds to the proposition that *Bert brushed himself*, something that is clearly true and that is closely related to the story's focus on how Bert can remove the snow from his body. In contrast, the deictic interpretation corresponds to the proposition that *Bert brushed the third reindeer*, a possibility that Bert mentions only in passing and that is not directly related to the theme of the story. This may have reduced the accessibility of the deictic interpretation in the referential condition and contributed to children's bias to judge the anaphoric interpretation. In the quantificational condition, the anaphoric interpretation corresponds to the proposition that *every reindeer brushed himself*, something that is clearly true and that is at least indirectly related to the theme of the story, since the reindeer refused to help Bert because they were busy helping themselves. The deictic interpretation corresponds to the proposition that *every reindeer brushed Bert*. Although this proposition does not come close to becoming true, it is closely related to the theme of the story, since Bert does ask each reindeer in turn to help him. Thus, there is an asymmetry in how the test sentence relates to the theme of the story across the two conditions, so that the overall design fails to satisfy the disputability assumption. The deictic interpretation is more relevant to the theme of the story in the quantificational condition than in the referential condition, and so it is perhaps not surprising that children judged the deictic interpretation more frequently in the quantificational condition than in the referential condition.

In light of these concerns, and others raised by the survey of previous studies discussed in section 5, we conducted three experiments that were designed to provide a fairer test of the QA and the DPBE.

⁴ Thornton and Wexler show that children willingly accept a bound variable interpretation in a sentence like *Every reindeer brushed himself*, which replaces the pronoun in (21) with a reflexive. However, the bound variable interpretation is obligatory here, and therefore this does not speak to Elbourne's concern. Koster (1994) also raises a concern about children's willingness to accept bound variable interpretations, although the results of Experiment 2 below suggest that this concern is unwarranted.

4 Experiments

To address Elbourne's concerns, and the additional concerns described in section 3, we conducted three experiments on children's knowledge of locality constraints on pronominal anaphora. The aim of the experiments was to test whether the delay of Principle B effect (DPBE) and the quantificational asymmetry (QA) persist once the assumptions about truth-value judgment task (TVJT) logic outlined above are satisfied. The aim was also to gain a better understanding of the substantial variation found in the results of previous studies of the QA and the DPBE.

Experiment 1 investigates the DPBE and the QA while providing maximally parallel tests for sentences with referential and quantificational antecedents. Experiment 2 provides an independent measure of the availability of bound pronoun interpretations, by pairing the same scenarios used in Experiment 1 with sentences that are not subject to Principle B. Experiment 3 examines the effect of modifying the scenarios from Experiments 1 and 2 so as to reintroduce some of the features that raised concerns about previous studies of the DPBE and the QA. The scenarios used in the experiments are schematized in figure 1.

4.1 Experiment 1

4.1.1 Design and Participants Experiment 1 investigated both the DPBE and the QA, using sentences with a pronoun direct object NP and either a referential or a quantificational subject NP. The experiment used a TVJT, in which a child and a puppet companion, Kermit the Frog, watched the experimenter act out a story with props. When the story was over, Kermit made a statement about it, and the child's task was to reward or correct Kermit based on the accuracy of his statement with respect to the scenario.

The experimental materials consisted of 8 stories, each of which was compatible with test sentences from both the referential condition and the quantificational condition. The 8 stories were assigned to two lists of items in a Latin square design, such that each participant saw all 8 stories, paired with 4 referential and 4 quantificational test sentences, and such that across participants each story was paired equally frequently with referential and quantificational test sentences. Participants were randomly assigned to one of the two lists. The 8 target items were combined with 8 filler items that were intermixed with the target items to create a test consisting of 16 stories. Filler stories were included to provide an independent measure of the children's understanding of the task. Furthermore, the filler sentences were assigned dynamically, such that the experimenter provided either a true or a false target sentence, in order to balance the overall number of true and false sentences presented over the course of the experiment. For individual children, the stories were divided over two sessions of no more than 20 minutes, with each session containing 8 stories. Adults were tested in a single session.

Participants were 16 English-speaking children aged 4;0–5;6 years (mean age 4;6 years) and 16 adult controls. Three additional children were replaced in the design because they made errors on more than two filler trials. The age range for the child participants corresponds to the age range that has been claimed to show strong DPBE and QA effects in previous studies. The children were recruited from preschools at the University of Maryland and in the College Park, MD, area.

We first summarize a sample story in (23) and then review how the story satisfies the assumptions of the TVJT. All other stories were designed following the same template of events. A full list of test sentences can be found in appendix A. The text in (23) describes a scenario presented to children; it does not represent the child-friendly narrative that the children actually heard. Sample videos and slides illustrating the stories can be found on the authors' Web sites.

(23) *The Painting Story*

Characters: Hiking Smurf, Tennis Smurf, Papa Smurf [collectively Smurfs]⁵
 Grumpy, Dopey, Happy [collectively dwarves]

Papa Smurf announces that Snow White is going to have a party, and that she is going to have a painting contest. Papa Smurf declares that he is going to be the judge. Each of the dwarves shows and discusses the color of paint that he is going to use to get painted, as does Tennis Smurf. However, Hiking Smurf does not have any paint, and he wonders whether one of the other characters will be willing to share. He first approaches Happy, who says that he would be glad to help out if any paint remains after he is painted. Fortunately, when Happy is finished some paint remains, and so he paints Hiking Smurf. Hiking Smurf, however, is not yet satisfied, so he approaches Dopey with a similar request, which is similarly successful. Then Grumpy, who is in such a bad mood that he doesn't even want to go to the party, declares that he doesn't need to get painted. The other dwarves really want him to go, and Grumpy agrees to get painted, using all of his paint in the process. After Grumpy is painted, Hiking Smurf approaches him and asks for some paint. Grumpy politely apologizes that he would like to help but cannot, because he has used up all of his paint. Hiking Smurf realizes that his best remaining chance is to ask Tennis Smurf for some extra paint, and Tennis Smurf obliges when he is asked. Finally, everybody is ready for Snow White's party.

Referential lead-in: OK, this was a story about painting. Hiking Smurf didn't have any paint, and Grumpy almost didn't go to the party. Let me see . . . I think . . .

Quantificational lead-in: OK, this was a story about painting. Hiking Smurf didn't have any paint, and all the dwarves looked great. Let me see . . . I think . . .

(24) Grumpy painted him. *Referential condition*

(25) Every dwarf painted him. *Quantificational condition*

The story in (23) attempts to satisfy the assumptions underlying the TVJT logic in a maximally similar fashion in the referential and quantificational conditions, as follows.

⁵ Note that although the two groups of characters in each story had a collective name, they were not described as a group using the collective term in Experiments 1 and 2, in order to ensure that they were adequately individuated. In Experiment 3, the collective names were used, paralleling earlier studies. Note also that although a number of the characters in the stories carried out reflexive actions (e.g., painting themselves), explicit reflexives were never used in telling the stories.

The same stories were used to test the referential and quantificational conditions, and the within-subjects Latin square design ensured that differences between the stories themselves could not be responsible for differences in responses in the two conditions. More importantly, the stories were designed such that the same events were the critical determinants of the truth or falsity of the test sentences in the two conditions, reducing the possibility that the salience of any individual character or event might lead to a spurious QA.

The referential and quantificational conditions are closely matched, both in terms of the accessibility of antecedents/referents (availability assumption) and in terms of the accessibility of the propositions that children were asked to judge (disputability assumption). The central character in the story is Hiking Smurf, and he is the intended deictic referent in both the referential and quantificational conditions. This minimizes any potential concerns about the availability of a suitable deictic referent for the pronoun. Meanwhile, the anaphoric antecedents are matched as closely as possible in the two conditions. Grumpy is the anaphoric antecedent in the referential condition. He is a prominent character in the story and is associated with the most vivid event in the narrative. Also, he is the most prominent of the set of dwarves who are the anaphoric antecedent in the quantificational condition. Hence, the availability assumption is satisfied in a similar fashion in the two conditions.

Turning to the disputability assumption, the referential and quantificational conditions are again closely matched and all relevant propositions are live possibilities during the story. The anaphoric interpretation corresponds to the proposition that *Grumpy painted himself* (referential condition) and that *every dwarf painted himself* (quantificational condition). Both propositions are true, and it is exactly the same event that makes both propositions true. After the first two dwarves paint themselves, Grumpy shows great reluctance to paint himself, temporarily raising the possibility that the anaphoric interpretation will fail to become true in both conditions. When Grumpy finally agrees to paint himself, both propositions become true, and in a similarly vivid fashion. Meanwhile, the deictic interpretation corresponds to the proposition that *Grumpy painted Hiking Smurf* (referential condition) and that *every dwarf painted Hiking Smurf* (quantificational condition). Both propositions are false, and for exactly the same reason, namely, Grumpy's refusal to help.

Furthermore, in response to Elbourne's concern that children may have difficulty in accessing the bound variable interpretation of singular pronouns, we took further steps to promote the accessibility of the anaphoric interpretation in the quantificational condition. First, all characters in the story have a clear individual identity in addition to being part of a group. This contrasts with the relatively undifferentiated reindeer in the story in (20). Second, each character draws attention to his need to paint himself, before offering assistance to Hiking Smurf. (Note, however, that no explicit reflexives were ever used in telling the stories.) Third, both Hiking Smurf and Grumpy/the dwarves are mentioned in the lead-in sentence that precedes the test sentence (this last measure is similar to the procedure in Thornton and Wexler 1999).

Figure 1 illustrates schematically the scenarios used in each of our experiments, highlighting the events that make the anaphoric interpretation of the pronoun true and the events that make the deictic interpretation false, in the referential and quantificational conditions alike. If the design

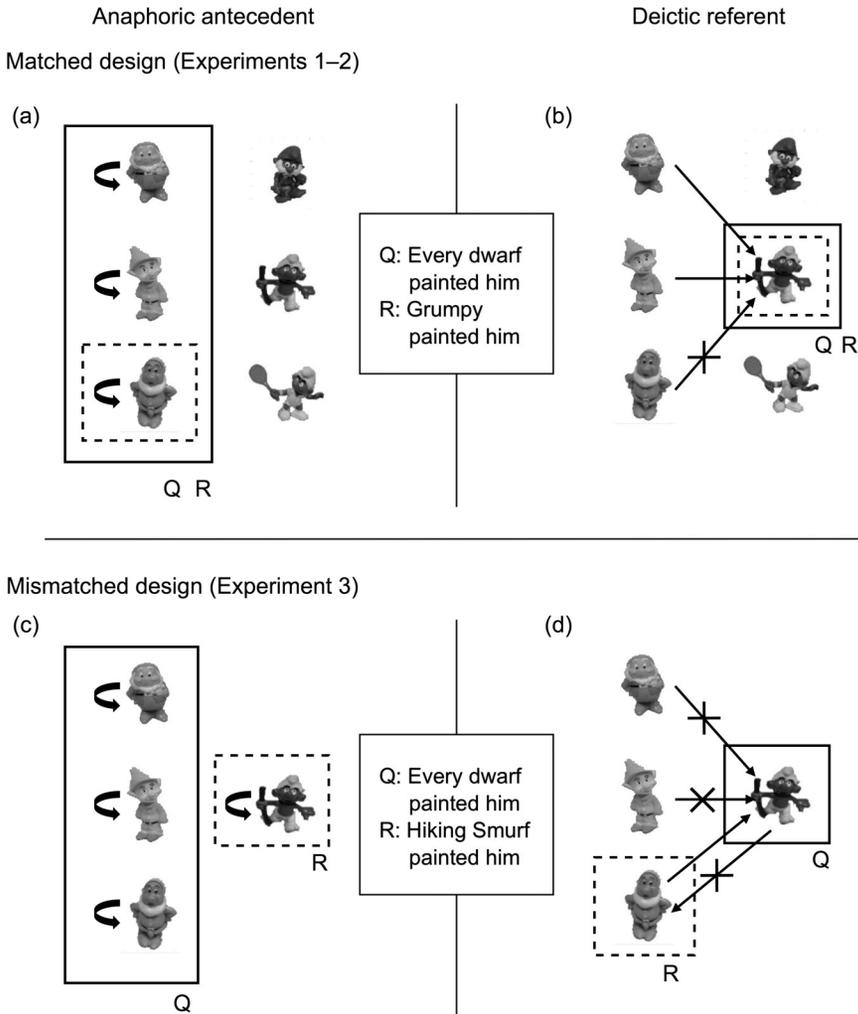


Figure 1

Schematic illustration of characters and scenarios used in Experiments 1–3, highlighting the events that made the anaphoric interpretation of the pronoun true (left column) and the deictic interpretation of the pronoun false (right column), and comparing the referential condition (dashed box) and the quantificational condition (solid box). Reflexive actions are indicated using curved arrows, transitive actions using straight arrows. An arrow with a cross indicates a case where a character refused to carry out an action. (a) and (b) show that in Experiments 1 and 2, the critical events were maximally similar in the referential and quantificational conditions. In Experiment 3, in contrast, the scenarios were modified to make them more similar to those used in some previous tests of the QA, where the critical events were not matched in the referential and quantificational conditions, as shown in (c) and (d).

succeeds in matching the conditions in terms of the accessibility of the relevant referents and propositions, then any observed differences between conditions can more confidently be attributed to the children's grammars. Of course, despite our efforts to ensure that nothing other than Principle B might make the anaphoric interpretation of the pronoun inaccessible to the children, it is difficult to prove this using test sentences that violate Principle B. Experiment 2 provides a more direct test of the accessibility of the anaphoric interpretation.

4.1.2 Results Results are based on the number of trials where the responses reflected an anaphoric interpretation of the pronoun, which was always true in the story, and the number of trials where the responses reflected a deictic interpretation of the pronoun, which was always false in the story. The primary indicator of the pronoun interpretation came from the "yes" and "no" judgments of the puppet's statements, but the children were always asked to explain to Kermit why he was right or wrong, under the guise of helping him to do a better job. In cases where a child's justification conflicted with his or her yes/no judgment, the justification was used to classify the response. For example, if a child gave a "yes" response to the quantificational test sentence *Every dwarf painted him*, but subsequently explained that "only those two did," pointing to the two dwarves who had painted Hiking Smurf, this indicated that the child had interpreted the pronoun deictically. This procedure is consistent with the basic logic of the TVJT (Crain and Thornton 1998). In Experiment 1, children's explanations contradicted their yes/no responses in only 6% of trials (8/128, 6/64 in the quantificational condition).

Results showed that the children and the adult controls consistently avoided the anaphoric interpretation of the pronoun in both conditions. Children accepted the anaphoric interpretation in 11% (7/64) of referential trials and in 14% (9/64) of quantificational trials. This difference was not significant (Wilcoxon signed-ranks, $Z = -0.541$; $p = .59$). No child gave more than 2 nonadultlike responses in either condition, and the nonadultlike responses were contributed by 5 children in the referential condition and by 7 children in the quantificational condition. Adult controls accepted the anaphoric interpretation in only 5% (3/64) of referential trials and 3% (2/64) of quantificational trials, again showing no significant difference between conditions ($Z = -0.447$; $p = .65$). Overall, children accepted more anaphoric interpretations than adults (Wilcoxon signed-ranks, $Z = -2.145$, $p < .05$), but this difference did not interact with experimental condition (Kruskal-Wallis $\chi^2 = 5.12$, $p > .15$).

4.1.3 Discussion The results of Experiment 1 indicate that children and adults consistently avoided the illicit anaphoric interpretation of the pronoun, instead choosing a deictic interpretation of the pronoun that made the test sentence false. The choice of deictic interpretation was consistent across conditions. Assuming that the logic of the TVJT was adequately satisfied, this result suggests that the children and adults avoided the anaphoric interpretation of the pronoun because they respect Principle B. In other words, we find no DPBE and no QA, contrary to the results of a number of previous studies. However, this conclusion depends on the assumption that the children's only reason for avoiding the anaphoric interpretation of the pronoun is Principle B, and it is difficult to confirm this step in the argument using materials that are subject to Principle B. Additionally, the results of this experiment are compatible with the concern raised by Elbourne

and others that children might exhibit a general dispreference for bound variable interpretations of pronouns, independent of Principle B. We conducted Experiment 2 in order to provide an independent test of whether the bound interpretation of the pronoun is available, once the effect of Principle B is neutralized.

4.2 Experiment 2

4.2.1 Design and Participants Experiment 2 was designed to test whether the anaphoric reading of pronouns is grammatically available to children, and whether the meaning that supports this reading is sufficiently prominent in the stories used in Experiment 1. Experiment 2 was identical to Experiment 1 in all respects, except for the object NP in the test sentences used after each story. (26) and (27) have the same truth conditions as the illicit anaphoric interpretation of (24) and (25), but embedding the pronoun as a possessor inside the object NP makes the anaphoric readings fully acceptable. If Principle B were the only reason for rejection of the bound interpretation of the pronoun in Experiment 1, then participants should readily accept the bound interpretation in Experiment 2. Nevertheless, there was still no “correct” or “incorrect” response in this experiment, since the test sentences were fully ambiguous.

(26) Grumpy painted his costume. *Referential condition*

(27) Every dwarf painted his costume. *Quantificational condition*

Note that (26) and (27) provide a fairer test of the availability of the bound reading than does a control in which the pronoun is replaced with a reflexive (e.g., *Every dwarf painted himself*; see Thornton and Wexler 1999). Because a reflexive is obligatorily bound by a local antecedent, a control condition that uses reflexives cannot provide an independent measure of whether a context equally supports the nonbound interpretations of the pronoun.

Participants were 16 English-speaking children aged 4;0–5;4 years (mean age 4;6 years) and 16 adult controls, none of whom had participated in Experiment 1. One child who gave more than two incorrect responses in filler trials was replaced in the design. The children were recruited from preschools at the University of Maryland and in the College Park, MD, area. The procedure was identical to that of Experiment 1.

4.2.2 Results As in Experiment 1, we used both yes/no responses and children’s justifications of their answers to classify responses as evidence for anaphoric versus deictic interpretations of the pronoun. Children’s justifications diverged from the default interpretation of yes/no responses in 5% of trials (6/128 trials, 3/64 in the quantificational condition).

Children accepted the bound interpretation of the pronoun in 80% (51/64) of referential trials and 73% (47/64) of quantificational trials. This difference was not significant (Wilcoxon signed-ranks $Z = -0.836$; $p = .40$). The 13 rejections in the referential condition were contributed by 13 children; no child rejected the target sentence in the referential condition more than once. The 17 rejections in the quantificational condition were contributed by 13 children, 4 of whom gave two rejections. Adults accepted the anaphoric interpretation of the pronoun in 83% (53/64) of referential trials and 67% (43/64) of quantificational trials. This difference was significant ($Z =$

–2.640; $p < .01$). A comparison of adult and child responses showed no main effect of participant group and no group \times antecedent interaction. A comparison of the children's responses in Experiments 1 and 2 showed a highly reliable difference in rates of acceptance of anaphoric interpretations (Kruskal-Wallis $\chi^2 = 42.395$, $p = .001$).

4.2.3 Discussion Experiment 2 was identical to Experiment 1 in all respects, except for the position of the pronoun in the target statements. In Experiment 1, Principle B was potentially active; but by making the pronoun a possessor in Experiment 2, we neutralized any possible contribution of Principle B. In light of the dramatic increase in acceptance of the anaphoric interpretation of the pronoun in Experiment 2, there is good reason to conclude that Principle B was responsible for avoidance of the anaphoric interpretation in Experiment 1. This result leaves little doubt that the test stories make the anaphoric interpretation readily accessible and that children have little difficulty in accepting bound variable interpretations of pronouns, contrary to concerns raised in some previous studies (e.g., Koster 1994, Elbourne 2005). Adults showed a small but reliable tendency to accept anaphoric readings more frequently in the referential condition. Although this might indicate that the anaphoric interpretation was more salient in the referential condition, it should be noted that this difference is proportionally very small compared with the differences that have been presented as evidence for the QA.⁶

In addition, the results of Experiment 2 may provide independent support for the Principle of Charity. In this study, both the deictic and the anaphoric interpretations of the test sentence were available and grammatical, but only one of them was true in the story. Children overwhelmingly said that the sentence was true, which corroborates the basic assumption of the TVJT that children show a bias to give positive answers (Crain and Thornton 1998).

Together, Experiments 1 and 2 satisfy the requirements of a fair TVJT test of the DPBE and the QA, and in so doing also address the concerns raised in Elbourne's (2005) critique. The results confirm Elbourne's prediction that appropriately matched referential and quantificational conditions would show no QA in children. However, our results do not support Elbourne's further prediction that the improved tests would show a clear DPBE in quantificational and referential conditions alike. In fact, our results show that 4-year-old children very rarely chose pronoun interpretations that violate Principle B.

We are now in a position to consider the theoretical implications of the lack of the QA and the DPBE. The QA has been taken to provide a dramatic piece of evidence in favor of Reinhart's (1983) approach to binding theory, which restricts the scope of binding constraints to cases of bound variable anaphora. If children do not show a QA, then one frequently cited argument in favor of Reinhart's approach disappears, although this result does not provide evidence *against* Reinhart's approach, and the theoretical arguments discussed in section 2.2 are unaffected. However, our findings also have more positive consequences, as they open up theoretical possibilities

⁶ The difference in the current study was a 23% higher rate of 'yes' answers in the referential condition than in the quantificational condition. This is very different from the proportional increases attributed to a QA in previous studies (e.g., Chien and Wexler 1990: 220%; Philip and Coopmans 1996: 160%; Matsuoka 1997: 250%; Thornton and Wexler 1999: 625%).

that may have appeared to be closed off by the QA or the DPBE, and they even remove a potential problem for Reinhart's theory. In particular, our findings cast doubt on the surprising asymmetry in children's mastery of Principles B and C that previous studies have reported. The developmental advantage for Principle C is unexpected under theoretical accounts that we are aware of. Under Reinhart's approach, both types of disjoint reference effects fall within the scope of Rule I, and therefore no developmental contrast is expected.⁷

However, our 4-year-olds' apparent lack of a DPBE contrasts with many previous reports of a DPBE in children of a similar age, and this raises the concern that our children's success may have been due to a statistical fluke or a design flaw. We took two steps to begin to address this concern. First, we conducted an additional study (Experiment 3) that was based on the same stories as Experiments 1 and 2, but that reintroduced some of the design features used in previous TVJT tests of the DPBE. Second, we reviewed previous studies of Principle B in children, to assess the reliability of previous findings and the methodologies used in those studies.

4.3 Experiment 3

4.3.1 Design and Participants Experiments 1 and 2 showed that children appear to respect Principle B and display no QA when presented with tests that satisfy the logic of the TVJT. In addition, as we argued in section 3, the QA observed in previous studies may derive from experimental designs that failed to equally satisfy the availability and disputability assumptions in referential and quantificational conditions. We therefore predict that it should be possible to reintroduce the contrasts in children's behavior by altering key features of our experimental designs. Accordingly, in Experiment 3 we modified the stories used in Experiments 1 and 2 in several ways, making them more like the sample story in (20) from Thornton and Wexler 1999.

The story in (28) is a modified version of the story in (23), and the test sentences are shown in (29)–(30). We should emphasize that because the two versions of the story differ in several ways, this experiment cannot identify the exact cause of any differences that might emerge in children's responses. Rather, it serves as an initial test of how much the contextual details of a TVJT experiment influence children's interpretation of pronouns. As in Experiments 1 and 2, the text in (28) presents the plot of a story told to children; it does not give the exact narrative that the children heard. Example videos are available from the authors' Web sites.

- (28) This is a story about three dwarves and Hiking Smurf. Hiking Smurf announces a party at Snow White's house and declares that everybody needs to get painted for the party. He then realizes that he is out of paint and proceeds to solicit help from the dwarves. Hiking Smurf asks the first dwarf to paint him, but he refuses because he is too busy painting himself. Hiking Smurf then approaches the second dwarf, but he also refuses

⁷ Grodzinsky and Reinhart (1993:91–93) recognize this discrepancy and suggest that the evidence for children's mastery of Principle C is unclear, since studies have often confounded the effects of Principle C with the possible effects of a dispreference for backward anaphora. We revisit this issue in section 5, showing that more recent studies have repeatedly confirmed children's early mastery of Principle C.

and paints himself. Hiking Smurf finally asks the third dwarf, who is more forthcoming. He says, “I can give you a little of my paint, but not too much, I need to get painted.” Hiking Smurf thanks the dwarf and remarks that he wishes he could return the favor by helping to paint the dwarf, but cannot because he is too busy getting painted himself.

Referential lead-in: This was a story about dwarves and Hiking Smurf.⁸

Quantificational lead-in: This was a story about dwarves and Hiking Smurf.

- (29) Hiking Smurf painted him. *Referential condition*
 (30) Every dwarf painted him. *Quantificational condition*

The story in (28) and the test sentences in (29)–(30) differ from their counterparts in Experiments 1 and 2 by reintroducing the contrasts between the referential and quantificational conditions found in the story in (20) and the test sentences in (21)–(22).

In terms of the accessibility of antecedents/referents for the pronoun *him* (availability assumption), the referential and quantificational conditions differ. The central figure in the narrative is Hiking Smurf, who fulfills different roles in the two test sentences: he is the intended anaphoric antecedent in the referential condition and the intended deictic antecedent in the quantificational condition. Therefore, if children simply interpret *him* as referring to Hiking Smurf, they should judge the test sentence true in the referential condition and false in the quantificational condition, leading to the appearance of a QA. Meanwhile, there is an additional motivation for children to judge the ungrammatical anaphoric interpretation of the referential condition, since the intended referent, the third dwarf, is relatively insignificant in the story and is also not directly mentioned in the puppet’s lead-in sentence.

In terms of the accessibility of relevant propositions (disputability assumption), the referential and quantificational conditions again contrast, and in a way that could lead to a spurious QA. In the referential condition, the anaphoric interpretation of the pronoun is associated with the proposition that Hiking Smurf painted himself. This is clearly true in the story, although it was not an expected outcome. The deictic interpretation of the pronoun is associated with the proposition that *Hiking Smurf painted the third dwarf*, an eventuality that is unrelated to the plot of the story and is never under consideration until Hiking Smurf mentions in passing that he cannot do it. Thus, there is a likely bias in the referential condition for the anaphoric interpretation. The situation is reversed in the quantificational condition. The deictic interpretation is associated with the proposition that *every dwarf painted Hiking Smurf*. Although this proposition does not become true in the story, it is directly related to the central theme of the story, namely, Hiking Smurf’s request to every dwarf for help. The anaphoric interpretation corresponds to the proposition that *every dwarf painted himself*, something that is clearly true in the story, although it is only indirectly related to Hiking Smurf’s quest.

⁸ As in Experiments 1 and 2, the lead-in sentences first mentioned the group containing the appropriate (deictic) antecedent and then the group containing the inappropriate (anaphoric) antecedent before presenting the test sentence. However, the exact format of the lead-in sentences was changed in order to match our understanding of Thornton and Wexler’s procedure.

It should be noted that the story in (28) nominally conforms to the basic TVJT parameters of a test of a grammatical constraint, since the story makes an ungrammatical reading true and a grammatical reading false and introduces events that potentially make the grammatical interpretation plausible. But as we have emphasized, a fair test of the QA and the DPBE requires more than this.

Apart from the changes in the stories, all other details of the design of the experiment were identical to Experiment 1. The 8 test stories were distributed across two lists in a Latin square design and combined with 8 filler stories to create a task involving 16 stories, which was administered to each child in two testing sessions. Participants were an additional 16 English-speaking children aged 4;1–5;2 years (mean age 4;7). Two children were replaced in the design because they gave more than two incorrect responses in the filler trials. The children were recruited from preschools at the University of Maryland and in the College Park, MD, area.

4.3.2 Results and Discussion When we combine all trials in Experiment 3 on which the child's judgment or justification reflects an anaphoric interpretation, we find a clear contrast between the quantificational and referential conditions. Children showed evidence of evaluating the anaphoric interpretation in 56% (36/64) of referential trials, but in only 16% (10/64) of quantificational trials. This difference was statistically reliable (Wilcoxon signed-ranks $Z = -2.507$; $p = .01$). The total of 36 anaphoric interpretations of referential trials were contributed by 22 children, 14 of whom gave two nonadultlike responses. In the quantificational condition, the total of 10 anaphoric interpretations were contributed by 7 children, 3 of whom gave two nonadultlike responses. Table 1 compares the rates of acceptance of anaphoric interpretations across all three experiments, showing that only in Experiment 3 was there evidence of a DPBE and a QA.

The fact that the DPBE and the QA emerged with the modified stories used in Experiment 3 supports the notion that in previous studies, these effects might have resulted from artifacts of the TVJT designs used. It is impossible to determine the exact cause of the different results in Experiments 1 and 3, because a number of changes were made to the stories. However, a second asymmetry in the results of Experiment 3 suggests that the children may have performed worse because the relevant deictic interpretation was insufficiently accessible in the referential condition. We counted the number of trials on which children justified their answers by referring to the

Table 1
Acceptance rates for anaphoric interpretations of pronouns in Experiments 1–3

% Accept binding/coreference	Experiment 1		Experiment 2		Experiment 3	
<i>Children</i>						
Referential antecedent	7/64	11%	51/64	80%	36/64	56%
Quantificational antecedent	9/64	14%	47/64	73%	10/64	16%
<i>Adults</i>						
Referential antecedent	3/64	5%	53/64	83%		
Quantificational antecedent	2/64	3%	43/64	67%		

events in the story that falsify the target sentence. In the quantificational condition, children gave a relevant rationale on 32 of the 54 trials where they answered “no,” pointing out that two dwarves did not paint Hiking Smurf or that only the last dwarf did paint Hiking Smurf. In the referential condition, in contrast, the relevant events were referred to on only 1 of the 28 trials where a “no” answer was given. That is, almost no children said that the sentence was false because Hiking Smurf was too busy to help the third dwarf. The difference in justification type between conditions likely reflects the contrasting accessibility of the referents and propositions associated with the deictic interpretation in the two conditions.

In the referential condition of this experiment, there were a particularly high number of trials in which children answered “no” but gave a justification that was odd in two respects. First, these justifications were not consistent with the “no” response. Second, unlike in Experiment 1, these justifications did not make reference to the event in the story that falsified the target sentence on the noncoreferential interpretation.

With respect to the first point, it appeared to us that these justifications were most consistent with an anaphoric interpretation of the pronoun. For example, in 30% (19/64) of trials, children responded to *Hiking Smurf painted him* by saying something like “He [first dwarf] didn’t have enough time, he [second dwarf] wouldn’t, then he [third dwarf] painted him.” This justification consists of three conjoined sentences with identically interpreted VPs. In each of these VPs, it is obvious that the object of the verb refers to Hiking Smurf. Under the assumption that the justification reflects the child’s interpretation of the target sentence, it follows that the child also interpreted the pronoun in the target sentence as referring to Hiking Smurf. This leads us to believe that these justifications reflect an anaphoric interpretation of the target sentence, and so we scored them in that fashion. Of course, this reasoning is valid only to the extent that the children’s justifications indeed reflect their interpretation of the target sentence, as is standardly assumed in these types of tasks (Crain and Thornton 1998). However, if this assumption does not hold, we have no independent confirmation that the children are giving yes/no responses based on their interpretation of the sentence relative to the context.

Returning to the second point, if the children’s “no” responses reflect genuine noncoreferential interpretations, these justifications are incongruent in a different respect. In Experiment 1, a typical response for a noncoreferential interpretation described the event in the story that falsifies the target sentence. In response to a sentence like *Hiking Smurf painted him*, where the grammatical antecedent is the dwarf, children typically responded by saying, “No, because he [the Smurf] didn’t have time [to paint the dwarf].” This response serves two functions. First, by referring to the event in the story that falsifies the grammatical interpretation, this response illustrates that children interpreted the pronoun in an adultlike fashion. Second, the justification elaborates coherently on the children’s yes/no response, giving us more confidence in our result. These two features of children’s justifications are criterial for taking their judgments to reflect their understanding of the sentence-context pair. It is striking that in Experiment 3, responses that refer to the falsification event are essentially missing, as noted above. Because the yes/no responses and justifications were not aligned, we chose to score this type of response on the basis of the most coherent interpretation of the justification with respect to the target sentence.

Of course, the “no” responses that we are taking to reflect an anaphoric interpretation are still “no” responses, a fact that highlights the oddity of the experimental context in Experiment 3. That is, the misalignment between the yes/no response and the justification can itself be taken as additional evidence that the contexts in which these sentences were presented were somehow unnatural. Such misalignments were found only in the referential condition of Experiment 3 and not in any of the other five conditions that we tested, supporting our contention that apparent Principle B violations in the previous literature might reflect imperfectly designed experimental materials and not a lack of knowledge on the part of children.⁹

One further piece of evidence suggests that the children had difficulty relating the test sentences to the story. In some trials, children responded simply by retelling the story, as if they were not sure which events were relevant. This occurred on 6 trials in Experiment 3 (9%, 5 referential trials, 1 quantificational trial), and never occurred in Experiments 1 and 2.

Thus far, we have shown that when presented with suitably balanced experimental conditions, children appear to abide by the disjoint reference constraints imposed by Principle B, showing no evidence of a QA. We characterized the prerequisites for a fair test of children’s knowledge in terms of accessibility of antecedents/referents (availability) and relevant propositions (disputability), and we showed in Experiment 3 that when we reintroduced contrasts in how these requirements are satisfied, based on scenarios used by Thornton and Wexler (1999) and others, both a QA and a DPBE reemerged. Nevertheless, we should note that our experiments do not allow us to conclusively establish the relative importance of availability and disputability, since our focus is on the improvements that obtain when both are satisfied.

Although our experiments support the notion that children know Principle B (and associated constraints, such as Rule I), our findings are at odds with the received wisdom on this topic. Therefore, we next survey previous studies on the DPBE and the QA in order to determine whether the received wisdom is consistent with what has been found, and whether variation in previous results can be understood in terms of the experimental design factors that we have identified here.

5 Previous Findings

We have shown that under appropriate experimental conditions, children abide by Principle B, showing no effects of a quantificational asymmetry (QA) and no delay of Principle B effect (DPBE), contrary to the received wisdom on this topic. In this section, we survey more than 30 previous studies and find that their results diverge widely and that little evidence for a QA remains once matching of events and antecedents is taken into account. We then examine the evidence

⁹One thing that remains mysterious is that when our child participants were confused by the referential condition, they answered “no,” whereas the prior literature has found a higher rate of “yes” responses in such conditions. We suggest that this difference may be explained by other properties of the experimental sessions, possibly by the nature of the fillers. In our task, the truth of the filler sentences was determined dynamically to ensure a balance of “yes” and “no” responses across the experimental session, whereas other studies (e.g., Thornton and Wexler 1999) fixed the truth of the filler sentences independently of the child’s responses to be the opposite of the grammatical response. Since the grammatical response in previous work was associated with a “no” response, all fillers consequently had “yes” responses, possibly giving rise to an overall “yes” bias.

for the DPBE in English and other languages. Here we find that although some studies are subject to methodological concerns, there is good evidence that children do accept interpretations that violate Principle B, albeit at somewhat lower rates than is commonly supposed. Finally, we face the remaining question of why children, if they have knowledge of Principle B, are so susceptible to interpretations that violate this constraint but not to violations of other constraints on anaphora. We suggest that this contrast may be related to a parallel contrast between Principles B and C found in recent studies on the real-time processing of anaphora.

Here, we summarize the overall findings from the survey and some comments that are relevant to a number of previous studies. Further discussion of the specifics of certain individual studies can be found in appendix B, which is available on the authors' Web sites.

5.1 *The Quantificational Asymmetry*

We examined 19 studies that tested for a QA, among which 10 report a QA and 9 do not.¹⁰ These studies are summarized in table 2. Among the studies that report a QA, the rates of acceptance in the referential condition vary from 31% to 93%, and the rates of acceptance in the quantificational condition vary from 0% to 27%. In some studies that do not report a QA, children performed well with referential and quantificational antecedents alike (Kaufman 1988, Hestvik and Philip 1999, current studies), whereas in others, children showed similarly high error rates for both types of antecedent (Lombardi and Sarma 1989, Boster 1991 Exp. 2, Avrutin and Wexler 1992, Utakis 1995, Grolla 2005). Thus, the results of previous studies vary considerably (see also Kaufman 1994, Koster 1994, Elbourne 2005 for earlier reviews), and the finding of a QA is certainly not consistent.

Among the studies in our survey, Kaufman 1988 is one of the earliest tests of the QA, and it is also possibly the best example of a study that satisfies our criteria for a fair test of the QA. Kaufman used a TVJT in which the scenarios used for the quantificational and referential conditions were very similar in structure. In particular, the deictic antecedent for the pronoun was closely matched in the two conditions, and the event that made the deictic interpretation false was similar across conditions. It is therefore striking that Kaufman reports almost identical rates of acceptance of around 16% for the two conditions, thereby providing evidence against both the QA and the DPBE. One other study shows very low error rates for both types of antecedent. In a picture verification task with Norwegian-speaking 4- and 5-year-olds, Hestvik and Philip (1999/2000) found high rates of success in referential and quantificational conditions alike. The authors report a small difference in acceptance of local anaphora in their two conditions (referential, 3%; quantificational, 9%) and suggest that this reflects a QA. However, these are among the lowest error rates observed in any study of Principle B in children, and they therefore imply early mastery

¹⁰ Our survey includes studies on groups of typically developing monolingual children for which we were able to find at least some details of the methods used. We excluded case studies based on very small numbers of children, and studies of second language learners or atypically developing children. In cases where an investigator presents multiple studies based on very similar tests, we present only a representative example. Our sample of studies that tested clitic pronouns in Romance and other languages is also not comprehensive, since the finding of improved performance with clitic pronouns is not at odds with the findings in our own studies.

Table 2

Summary of results from tests of the quantificational asymmetry in children. (TVJT = truth-value judgment task)

Study	Language	Age	<i>N</i>	Accept referential	Accept quantificational	Method
<i>Studies reporting no quantificational asymmetry</i>						
Kaufman 1988	English	2;7–3;11	30	23%	18%	TVJT
		5;0–6;5	30	10%	13%	
Lombardi and Sarma 1989	English	4;0–6;2	11	55%	49%	Act-out, TVJT
Boster 1991, Exp. 2	English	3;3–4;9	24	38%	42%	Picture verification
Avrutin and Wexler 1992	Russian	4–7	16	52%	41%	TVJT
Utakis 1995	English	3;4–9;5	30	37%	40%	TVJT
Baauw, Escobar, and Philip 1997	Spanish; clitic	mean 5;6	45	10%	10%	Picture verification
Hamann, Kowalski, and Philip 1997	French; clitic	3;5–4;8	9	22%	30%	Picture verification
		5;3–5;11	8	0%	12%	
Hestvik and Philip 1999/2000	Norwegian	4;5–5;11	15	9%	3%	Picture verification
Grolla 2005	English Brazilian, Portuguese	3;7–5;11	23	52%	46%	Picture verification
		3;4–6;6	40	44%	49%	
<i>Studies reporting a quantificational asymmetry</i>						
Chien and Wexler 1990, Exp. 4	English	2;6–3;11	48	70%	54%	Picture verification
		4;0–4;11	45	60%	40%	
		5;0–5;11	44	51%	16%	
		6;0–7;0	40	24%	14%	
McDaniel, Cairns, and Hsu 1990	English	2;9–6;7	19	44%	19%	Grammaticality judgment
Thornton 1990	English; <i>who</i>	3;7–4;8	12	49%	8%	TVJT
Boster 1991, Exp. 1	English; <i>who</i>	4;6–6;0	10	38%	4%	TVJT
McDaniel and Maxfield 1992	English	3;1–6;10	37	41%	25%	Grammaticality judgment
Avrutin and Thornton 1994	English; collective vs. distributive	3;10–4;10	33	93%	27%	TVJT

Table 2 (*continued*)

Study	Language	Age	<i>N</i>	Accept referential	Accept quantificational	Method
Philip and Coopmans 1996	English	3;6–7;0	19	68%	26%	Picture verification
	Dutch; strong pronoun	4;3–6;11	37	66%	50%	
Matsuoka 1997	English	3;10–6;0	19	70%	20%	TVJT
Savarese 1999	English	3;5–5;11	25	31%	N/A	TVJT
		4;3–6;1	26	N/A	0%	
Thornton and Wexler 1999	English	4;0–5;1	19	58%	8%	TVJT

of binding constraints. Hestvik and Philip discuss some grammatical properties of Norwegian pronouns that may have caused the children's unusually good performance, but the specific cause remains uncertain.

Another group of studies shows no QA, while finding similarly high rates of errors in referential and quantificational conditions. Many of the errors can be attributed to limitations of the experimental designs used. For example, Grolla (2005) used a picture verification task in which only an anaphoric antecedent was provided in the test scenarios, leaving children with no alternative candidate referent for the pronoun. Avrutin and Wexler (1992) conducted a TVJT study on Russian that shares many design features with the scenario in (20) from Thornton and Wexler 1999 that we discussed at length in section 3. It is therefore not surprising that Avrutin and Wexler found error rates in the referential condition that were similar to those reported by Thornton and Wexler (1999), although it is surprising that the Russian-speaking children showed similarly high error rates in the quantificational condition. Lombardi and Sarma (1989) found high rates of nonadultlike responses in an act-out task, which we discuss further in appendix B.

A number of the studies that report a QA used TVJTs or similar designs that are subject to the same methodological concerns discussed in section 3. Matsuoka (1997) used story formats that closely parallel the example in (20) from Thornton and Wexler 1999, and hence it is unsurprising that Matsuoka found a similar QA. A pair of studies by McDaniel and colleagues found a QA in tasks that are described as grammaticality judgment tasks, although they are very similar to tasks that are described elsewhere as TVJTs (McDaniel, Cairns, and Hsu 1990, McDaniel and Maxfield 1992). Given the similarity with TVJTs, it is very relevant that these tasks did not make the deictic reading of the target sentences accessible. This could be responsible for the high rates of nonadultlike responses in the referential conditions, although it is not clear why a QA should have arisen in this study. In appendix B, we discuss further TVJT-like studies that have reported a QA, and we offer specific suggestions about the source of the observed asymmetries in those studies (Thornton 1990, Boster 1991, Avrutin and Thornton 1994, Savarese 1999).

It is important to comment also on the relation between the TVJT method that we have discussed at length here and the picture selection or picture verification tasks that have been used in many studies of Principle B in children, including the best-known report of a QA (Chien and Wexler 1990). We contend that picture-based tasks are subject to the same constraints that we have discussed for TVJTs, except that it is more difficult in picture-based tasks to assess how well the constraints are satisfied. In a picture verification task, as in a TVJT, children are placed in a situation where they could choose to interpret a pronoun either deictically or anaphorically. As in a TVJT, the choice of whether to interpret the pronoun deictically or anaphorically may depend on a number of factors in addition to the child's grammar, including the accessibility of suitable deictic antecedents and expectations about what events are likely to be commented upon. The primary difference between a TVJT and a picture judgment task is that in a TVJT the experimenter uses the narrative to explicitly control availability and disputability, whereas in a picture-based task a greater burden is placed on the child to conjure up a relevant context in which to interpret the picture. Elbourne (2005) offers a number of suggestions about how a QA may have arisen in Chien and Wexler's classic study, in which children were shown a line drawing, told "These are the bears; this is Goldilocks," and asked "Is every bear touching her?" (For specific comments on that study, see Elbourne 2005.)

We should emphasize that despite our criticisms of particular TVJT studies, we do not take these examples to show that the TVJT is fundamentally flawed or that other experimental measures are superior. We believe that our critique and our own studies follow closely the underlying logic of the TVJT, as laid out by Crain and his colleagues. Our criticisms apply to specific studies, not to the task itself.

Summarizing, our studies are by no means unique in failing to find a QA, and most previous studies that have reported a QA are amenable to alternative explanations that do not invoke a grammatical asymmetry between coreference and variable binding. We therefore consider it well justified to doubt the received wisdom about the existence of a QA in children.

5.2 *The Delay of Principle B Effect*

Our experimental results, together with those of Kaufman (1988), suggest that there is no DPBE and that children perform well across all types of antecedents. However, this conclusion is at odds with many previous studies. In addition to the 19 studies that tested the QA, our survey included a further 14 studies that tested children's adherence to Principle B with referential antecedents only. The results of these studies are summarized in table 3. Even if we restrict our attention initially to studies on English, we find 13 studies from table 2 and an additional 7 studies from table 3 that report a DPBE, with acceptance rates for local antecedents of pronouns that range from 16% to 82%.

The high degree of variability in acceptance rates across studies suggests that children's responses are not simply the product of a relatively stable grammar. If each study had presented a test of Principle B that was immune to extragrammatical biases, then we should have expected to observe more consistent results across studies. This implies that the variability reflects specifics

Table 3

Summary of selected tests of the delay of Principle B effect, excluding studies already covered in table 2 by the survey of tests of the quantificational asymmetry. (TVJT = truth-value judgment task)

Study	Language	Age	<i>N</i>	Accept referential	Task
Jakubowicz 1984	English	4	10	30% ^a	Act-out
		5	11	25%	
Wexler and Chien 1985	English	2;6–6;6	129	43%	Picture selection
Deutsch, Koster, and Koster 1986	Dutch; strong pronoun	6	32	46%	Picture selection
Solan 1987	English	4–7	37	57%	Act-out
Chien and Wexler 1990, Exps. 1–2	English	2;6–6;6	298	29% ^b	Act-out
Grimshaw and Rosen 1990	English	4–5	12	42%	TVJT
Padilla 1990	Spanish; clitic	3;0–3;11	20	37%	Act-out
		5;0–5;11	20	30%	
McKee 1992	English	2;6–5;3	60	82%	TVJT
	Italian; clitic	3;7–5;5	30	15%	TVJT
Sigurjónsdóttir and Hyams 1992	Icelandic	4;0–4;6	10	45%	TVJT
		4;6–5;0	10	43%	
Baauw 1999	Dutch; weak pronoun	4;2–5;3	15	47%	Picture verification
Varlokosta 2000	Greek; clitic	3;7–5;6	20	5%	TVJT
	Greek; strong pronoun			13%	
Varlokosta and Dullaart 2001	Greek; clitic	3;3–7;6	10	5%	TVJT
	Greek; strong pronoun			5%	
	Dutch; weak pronoun			55%	
	Dutch; strong pronoun			45%	
Kiguchi and Thornton 2004	English	4;1–5;10	13	27%	TVJT
Spender, Smits, and Hendriks 2007	Dutch	4;4–7;7	83	25% ^c	Picture verification

^a The percentages for Jakubowicz 1984 are estimates derived from published histograms.

^b The average shown for Chien and Wexler 1990 is a nonweighted average derived from the mean of all age groups.

^c The average for Spender, Smits, and Hendriks 2007 is a mean across three sentential contexts.

of the tasks used in individual studies. Indeed, in our survey we found that a good deal of the variability in previous results could be explained by task differences, and in particular by the extent to which the task provided a clear deictic alternative to the illicit anaphoric interpretation of the pronoun.¹¹ Nevertheless, we find that there is a ‘residue’ of the DPBE that is a real effect and not an experimental artifact, and we propose an account for this effect in section 5.3.

¹¹ We also found substantial variation across studies in the level of detail provided in the experimental descriptions. Many studies did not give enough information to allow their methods to be adequately assessed.

In tests of the DPBE where the QA is not at stake, it remains important to satisfy the assumptions of availability and disputability for the anaphoric and deictic interpretations of the pronoun. However, it is more straightforward to satisfy these assumptions in this case, because there is no need to also closely match quantificational and referential conditions. In all TVJT tests of the DPBE that we are aware of, the anaphoric interpretation of the pronoun is made true and readily accessible. The primary variation in the experimental designs lies in whether a deictic antecedent is readily available, and in whether the proposition corresponding to the deictic interpretation of the pronoun is a live possibility in the scenario, despite ultimately turning out to be false.

In TVJT studies where the deictic interpretation of the pronoun is accessible in the context, we find relatively low rates of nonadultlike judgments, although the rates are often too high to be dismissed as experimental “noise” (Kaufman 1988, 16% acceptance;¹² Thornton 1990, 29% acceptance across conditions; Boster 1991, 21% acceptance across conditions; Savarese 1999, 31% acceptance; Kiguchi and Thornton 2004, 27% acceptance; current study, 11% acceptance). In these studies, the deictic interpretation of the pronoun corresponds to a prominent character in the story, and the event corresponding to the deictic interpretation is explicitly rejected in the story.

In a number of other studies using the TVJT or similar tasks, we find that the experimental design either fails to make a deictic antecedent available or provides a deictic interpretation that is not seriously under consideration in the scenario. In these studies, we typically find much higher rates of acceptance of illicit anaphoric interpretations of pronouns (Grimshaw and Rosen 1990, 42% acceptance; McDaniel, Cairns, and Hsu 1990, 44% acceptance; McDaniel and Maxfield 1992, 41% acceptance; McKee 1992, 82% acceptance; Matsuoka 1997, 70% acceptance; Thornton and Wexler 1999, 58% acceptance; current study, Exp. 3, 56% anaphoric interpretations). We offer more specific remarks on some of these studies in appendix B, but the overall generalization from TVJT studies of the DPBE is clear: children are more likely to give an ungrammatical anaphoric interpretation in tasks where a grammatical deictic interpretation of the pronoun is not readily accessible.

A number of early studies of the DPBE used act-out tasks (Jakubowicz 1984, Solan 1987, Lombardi and Sarma 1989). These tasks have the limitation that they track a child’s preferred interpretation of a test sentence and cannot readily distinguish dispreferred from illicit interpretations. However, a distinct advantage is that the child’s act-out provides more direct evidence of his interpretation than does the yes/no response used in TVJT and picture verification tasks. Jakubowicz (1984) used an act-out task in one of the earliest demonstrations of the DPBE, and showed a relatively low rate of anaphoric interpretations of the pronoun (25%–30%). Chien and Wexler (1990, Exps. 1–2) also found a relatively low rate of anaphoric interpretations (29%) in an act-out task conducted with around 300 children. Rather higher rates of anaphoric interpretation have been reported in other act-out studies, but in at least one of these cases the design may have

¹² Kaufman’s descriptions suggest that the events corresponding to the deictic interpretation of the pronoun were explicitly avoided by characters in her stories, but it is not clear whether this was a consistent feature of the stories.

led to exaggeration of the number of anaphoric interpretations (Lombardi and Sarma 1989; see appendix B).

A number of studies of the DPBE have used picture verification tasks, and these have revealed similarly broad variability in acceptance of illicit anaphoric interpretations of pronouns, ranging from 9% to 68% (Chien and Wexler 1990 Exp. 4, Boster 1991, Philip and Coopmans 1996, Hestvik and Philip 1999/2000, Spenader, Smits, and Hendriks 2007). As discussed above, picture verification tasks rely on a logic similar to that of TVJTs, with the difference that it is harder to control the context against which the child judges the test sentence. (For discussion of factors that may affect how children choose to interpret pronouns in these tasks, see Elbourne 2005 and Spenader, Smits, and Hendriks 2007.) A variant on the picture verification task is the picture selection task used in an important early study by Wexler and Chien (1985). In that study, children listened to sentences like *Cinderella's sister points to her* and had to find a picture that showed the scenario described in the sentence. In this task, the children chose pictures that corresponded to Principle B violations on 43% of trials. The similarity between this task and picture verification tasks depends on how a child chooses to undertake the task. The child may treat the task as a series of picture verification tasks, looking at each picture in succession and deciding whether the test sentence accurately describes the picture. In this case, exactly the same considerations apply as in TVJTs and picture verification tasks. On the other hand, if the child carries out the task by first constructing an interpretation of the test sentence and then looking for a picture that matches that interpretation, the task is a little different. This way of approaching the task may make requirements such as disputability or plausible dissent irrelevant, but it does not remove the need for a viable deictic antecedent for the pronoun. Because of this uncertainty and the limited information about the materials used in this study, we cannot offer firm suggestions on the cause of children's nonadultlike responses. However, we speculate that it may not have been immediately apparent to children that the possessor *Cinderella* was a viable referent for the pronoun (see Spenader, Smits, and Hendriks 2007 for remarks on applying Centering Theory (Grosz, Joshi, and Weinstein 1995) to Principle B studies).

Summarizing previous studies on the DPBE in English, we find that the DPBE is weaker than often supposed. In the studies that we take to present the fairest tests of the DPBE, we find that children accept illicit anaphoric readings of pronouns in only 15%–30% of trials. We do not find evidence that children “guess” when presented with potential Principle B violations, nor do we find evidence that children misanalyze pronouns as elements that require or strongly prefer local binding. In this respect, the results from our own experiments are consistent with previous literature.

However, we cannot conclude from the survey of DPBE studies that the effect is artifactual. Even in many studies that present fair tests of binding constraints, we find that some form of DPBE remains. If Principle B acted as a strong constraint on children's interpretations, then we should expect it to be sufficiently powerful to make children “blind” to illicit anaphoric interpretations of pronouns, something that appears not to be the case. In addition, we must acknowledge that some studies in other languages, particularly languages with clitic pronouns, have shown that similar or identical tests elicit much lower rates of Principle B violations. In the

next sections, we consider the source of the residual DPBE effect and why children behave differently in tests of Principle C and in tests involving clitic pronouns.

5.3 *Principle B versus Principle C*

Although our survey supports the conclusion from our own experiments that the QA is an experimental artifact, our conclusions about the DPBE are more nuanced. Our own studies indicate that Principle B has a strong impact on 4-year-olds' judgments, since the children rarely accepted anaphoric interpretations in Experiment 1, where Principle B was relevant, and frequently accepted anaphoric interpretations in Experiment 2, where Principle B was neutralized. However, we are still left with a number of studies that appear to present fair tests of children's knowledge of Principle B and that show acceptance of illicit antecedents on ~15%–30% of trials. This is a weaker DPBE than is often assumed in the literature (e.g., Reinhart 2006), but it cannot easily be dismissed as "noise." A useful comparison can be found in TVJT studies of Principle C, which have typically shown error rates of around 10%–20% (Crain and McKee 1986, 12% acceptance; Guasti and Chierchia 1999/2000, 11% acceptance; Kazanina and Phillips 2001, 17% acceptance).¹³ These findings have been taken to indicate that children know Principle C by the age of 3 to 5. See also Lust, Loveland, and Kornet 1980, Solan 1983, and McDaniel, Cairns, and Hsu 1990 for studies using other techniques, and Lust, Eisele, and Mazuka 1992 for a review of earlier studies on this topic.

Therefore, the "residue" of the DPBE appears to be slightly elevated error rates in Principle B contexts relative to Principle C contexts. We cannot reasonably argue from such small differences that 4-year-olds know Principle C but do not know Principle B (and Rule I and related constraints). It has sometimes been suggested that children appear to perform better on tests of Principle C because of a general dispreference for backward anaphora. However, a number of studies have shown that children freely accept backward anaphora once the effect of Principle C is neutralized (Crain and McKee 1986, Guasti and Chierchia 1999/2000, Kazanina and Phillips 2001), and hence it is difficult to dismiss children's success in tests of Principle C as an experimental artifact. We must therefore look elsewhere for an explanation of children's slightly degraded performance on tests of Principle B.

One possible explanation derives from a similar contrast between Principle B and Principle C that has been found in recent online studies of pronoun resolution in adults. These studies have asked whether binding constraints act as an "initial filter" on the processing of pronouns, such

¹³ Grimshaw and Rosen (1990) report a TVJT study that showed a much higher rate of acceptance of anaphoric interpretations. Children watched movie clips and judged statements about the clips. For example, in a scene that showed Ernie hitting himself, children were told, "Ernie was fighting with Big Bird. He hit Ernie." Grimshaw and Rosen's description of their study suggests that they used a strict coding scheme in which all instances of "yes" and "no" were considered relevant to the experimental hypothesis, even when the children's full answers suggested otherwise, such as when children said things like, "No, because hitting isn't right." It is therefore possible that some of the children's "yes" answers reflected an inference from the lead-in sentence that mentioned that Ernie and Big Bird were fighting, rather than acceptance of interpretations that violate Principle C.

that the parser is blind to potential antecedents in grammatically illicit positions, or whether they act as a “late filter,” such that comprehenders temporarily consider grammatically illicit antecedents for a pronoun before excluding them from consideration. Existing evidence suggests that Principle C acts as an initial filter, such that comprehenders do not attempt to link pronouns to R-expressions that they c-command (Cowart and Cairns 1987, Kazanina et al. 2007), whereas the results for Principle B are more mixed. Some studies using cross-modal priming and self-paced reading methods have presented evidence that Principle B acts as an initial filter (Nicol and Swinney 1989, Clifton, Kennison, and Albrecht 1997, Lee and Williams 2006), but a number of more recent studies using eye-tracking and self-paced reading measures have found evidence for temporary consideration of ungrammatical antecedents in Principle B contexts (Badecker and Straub 2002, Kennison 2003, Runner, Sussman, and Tanenhaus 2003, 2004, 2006, Sturt, Kreiner, and Garrod 2005).

Although the adult results indicate merely fleeting access to ungrammatical antecedents in online studies, whereas the results from children indicate “offline” judgments that violate Principle B, there is good reason to think that these might be related. A recurring finding in studies of children’s language processing is that children show greater difficulty than adults in inhibiting and recovering from incorrect initial interpretations of sentences (e.g., Hamburger and Crain 1984, Trueswell et al. 1999). Therefore, what appears in adults as transient effects of ungrammatical antecedents might appear in children as ungrammatical interpretations that persist.

Next, we can ask why Principle B and Principle C should affect the online search for pronoun antecedents in different ways. Although we cannot exclude the possibility that the constraints themselves are qualitatively different from one another, there are independent reasons why the search for antecedents might proceed differently in the two cases. These differences are all related to the fact that Principle B primarily constrains forward anaphora, whereas Principle C primarily constrains backward anaphora. In backward anaphora contexts, a pronoun precedes its antecedent, and encountering a pronoun initiates an active search for a suitable antecedent (Kazanina et al. 2007). During this search, the parser is able to consider potential antecedents one at a time as they appear in the input, with no need to retrieve antecedents from memory. Additionally, the parser can identify that a given structural domain cannot contain an antecedent for the pronoun, because of Principle C, before it encounters any of the NPs in that domain. In contrast, the forward anaphora contexts that are normally used in tests of Principle B place different demands on the reference resolution process. The parser encounters the pronoun only after it has encountered its potential antecedents, and it must therefore conduct a retrospective search of referents in memory. Furthermore, the contexts that are typically used in tests of Principle B in children and adults force the parser to consider multiple candidate antecedents (intrasentential or extrasentential) in parallel. Both of these factors may increase the likelihood of error in the search for a grammatically appropriate antecedent.

Finally, we can consider the cause of the significantly increased acceptance of Principle B violations in the studies that fail to make a grammatical deictic interpretation of the pronoun sufficiently accessible. Children in these studies who entertain the anaphoric interpretation of the pronoun receive strong semantic support for that interpretation, given its prominence in the story,

and they do not have a readily available deictic interpretation that can inhibit the anaphoric interpretation. This might account for the acceptance rates of 40%–80% observed in these studies.

Summarizing, evidence from many different studies with children indicates that 4-year-olds show good knowledge of the disjoint reference requirements imposed by Principles B and C, but that children are more prone to error in Principle B contexts. We suggest that this difference may reflect an independently motivated contrast in the search for pronoun antecedents that has been observed in online studies with adults. Whereas Principle C appears to act as a constraint on the generation of representations, Principle B may sometimes act as a filter on representations that are at least temporarily generated (see also Grimshaw and Rosen 1990).

5.4 *Pronouns versus Clitics*

We must also consider the frequently reported finding that the DPBE is much weaker in languages with clitic pronouns. Although there are some studies of clitic languages where children's improved performance may simply reflect an experimental design that better satisfies the disputability assumption (e.g., Varlokosta 2000), there are other studies that show strong cross-language differences using the same tasks, suggesting that the effect of clitic pronouns on the DPBE is genuine.¹⁴ For example, McKee (1992) reports that the Italian-speaking children in her study rarely accepted anaphoric interpretations of a clitic pronoun (15% acceptance), whereas English-speaking children accepted the anaphoric interpretation on a majority of trials (82% acceptance). Although the high acceptance rate among English-speaking children could be attributed to the lack of an accessible deictic antecedent in McKee's stories, this cannot explain the strong cross-language difference. Here we suggest that the cross-language difference may be related to the availability of cases of accidental coreference. English pronouns can be used in the examples of local accidental coreference that escape Rule I, such as (13) above. In Italian and other clitic languages, such examples require tonic pronouns and disallow clitic pronouns. Therefore, for an Italian-speaking child, the possibility of using a clitic pronoun for local coreference does not exist. For an English-speaking child, the possibility of using a pronoun for local coreference does exist in certain Rule I-escaping contexts, and thus, English-speaking children may mistakenly use pronouns to locally corefer outside of these contexts. This difference between clitic and tonic pronouns may affect the way in which children (and adults) access and inhibit potential antecedents during language comprehension.

6 Conclusion

The relation between grammatical knowledge and linguistic behavior is complex. In any experimental task, participants must access their linguistic knowledge in real time and relate it to a host of nonlinguistic properties of the experimental context. Given this complexity, behavior in a linguistic experiment (and, for that matter, in the real world) may be determined by (a) the

¹⁴ Nevertheless, Padilla (1990) reports 30%–40% choice of anaphoric interpretations of Spanish clitic pronouns in an act-out task, and studies in Dutch using weak and strong pronouns have not reported consistent differences (see table 3). Therefore, the empirical record is not yet unequivocal.

grammar, (b) the parser, (c) pragmatic influences on the interpretation of the context, and (d) world knowledge. Thus, in order to assess whether a behavior reflects grammatical structure, an experimenter must take great care to neutralize any possible influence from extragrammatical factors. We have argued that prior findings showing the delay of Principle B effect (DPBE) and the quantificational asymmetry (QA) leave room for extragrammatical explanation, and that once extragrammatical factors are removed, preschoolers show little evidence of deficit in their knowledge of Principle B.

Our argument was based on two kinds of data. First, a survey of the existing literature indicates that the empirical support for the DPBE and the QA is less robust than it is often presumed to be. There is substantial variability across experiments in the evidence for either effect, and corresponding variability in the experimental control over extralinguistic factors. Second, we conducted three experiments in which we found no evidence for either the DPBE or the QA. However, we did find a QA when we deliberately introduced extralinguistic factors that we considered as possible causes of a spurious QA in previous studies. Children in the first two experiments behaved in a way consistent with knowledge of Principle B, for quantificational and referential antecedents alike.

The dissolution of the QA also resolves an apparent conflict in the previous experimental literature. Whereas there are widespread reports of 4- to 5-year-olds making errors in Principle B contexts, many other results indicate that children of the same age show clear mastery of Principle C. This asymmetry between Principle B and Principle C introduces a problem for the pragmatic explanation of the QA. The standard account of the QA consists of two parts: first, the claim that there are two mechanisms governing the interpretation of pronouns in Principle B contexts, one dealing with syntactic variable binding, the other dealing with pragmatic conditions on the appropriate use of pronouns; second, the claim that children's difficulties in Principle B contexts with referential antecedents derive from difficulty in applying the relevant pragmatic rule. The problem with this kind of explanation is that the same pragmatic rule is assumed to apply in Principle C contexts, and thus we should expect to find the same difficulty with Principle C, contrary to fact. To the extent that our results have eliminated the developmental asymmetry between Principles B and C, they also eliminate theoretical problems that the contrast created.

Nevertheless, there appears to be some residual basis for the contrast between Principles B and C, as our survey of previous studies shows. Although we have argued that experimental design factors may account for much of the variability found in previous studies of the DPBE, we cannot overlook the fact that children appear to be more susceptible to interpretations that violate Principle B than they are to interpretations that violate Principle C. We have suggested that this contrast derives from independently motivated differences in how these constraints affect real-time language processing. Evidence from adult psycholinguistic studies suggests that NPs in the c-command domain of a pronoun are never even considered as possible antecedents, such that comprehenders are effectively "blind" to interpretations that violate Principle C. In contrast, a number of studies of pronouns in Principle B contexts indicate that both licit and illicit antecedents are at least temporarily considered. Children's susceptibility to Principle B violations is compatible with knowledge of Principle B, just as it is for adults.

Appendix A: Experimental Materials

The following list presents the target sentences used in the eight experimental stories. Each participant saw all eight stories, paired with the quantificational (Q) or referential (R) target sentence, in a Latin square design. Sample slides and videos illustrating the stories are available from the authors' Web sites.

Experiment 1	Experiment 2	Experiment 3
1 Q I think that every space guy decorated him.	I think that every space guy decorated his costume.	I think that every space guy decorated him.
R I think that Storm Trooper decorated him.	I think that Storm Trooper decorated his costume.	I think that Alien decorated him.
2 Q I think that every superhero squirted him.	I think that every superhero squirted his body.	I think that every knight squirted him.
R I think that Robocop squirted him.	I think that Robocop squirted his body.	I think that Dog squirted him.
3 Q I think that every lizard sprayed him.	I think that every lizard sprayed his body.	I think that every lizard sprayed him.
R I think that Blue Lizard sprayed him.	I think that Blue Lizard sprayed his body.	I think that Butterfly sprayed him.
4 Q I think that every Smurf stamped him.	I think that every Smurf stamped his shirt.	I think that every Smurf stamped him.
R I think that Painting Smurf stamped him.	I think that Painting Smurf stamped his shirt.	I think that Dog stamped him.
5 Q I think that every dwarf painted him.	I think that every dwarf painted his costume.	I think that every dwarf painted him.
R I think that Grumpy painted him.	I think that Grumpy painted his costume.	I think that Smurf painted him.
6 Q I think that every troll labeled him.	I think that every troll labeled his shirt.	I think that every troll labeled him.
R I think that Orange Troll labeled him.	I think that Orange Troll labeled his shirt.	I think that Grey Bear labeled him.
7 Q I think that every turtle wiped him.	I think that every turtle wiped his hair.	I think that every turtle wiped him.
R I think that Blue Turtle wiped him.	I think that Blue Turtle wiped his hair.	I think that Mickey wiped him.
8 Q I think that every M&M fanned him.	I think that every M&M fanned his body.	I think that every M&M fanned him.
R I think that Hat M&M fanned him.	I think that Hat M&M fanned his body.	I think that Barney fanned him.

Appendix B: Discussion of Previous Experiments

Appendix B is available on the authors' Web sites.

References

- Avrutin, Sergey, and Rosalind Thornton. 1994. Distributivity and binding in child grammar. *Linguistic Inquiry* 25:165–171.
- Avrutin, Sergey, and Kenneth Wexler. 1992. Development of Principle B in Russian: Coindexation at LF and coreference. *Language Acquisition* 2:259–306.
- Baaauw, Sergio. 1999. The role of the clitic–full pronoun distinction in the acquisition of pronominal coreference. In *Proceedings of the 23rd annual Boston University Conference on Language Development*, ed. by Annabel Greenhill, Heather Littlefield, and Cheryl Tano, 32–43. Somerville, MA: Cascadilla Press.
- Baaauw, Sergio, María A. Escobar, and William Philip. 1997. A delay of Principle B effect in Spanish speaking children: The role of lexical feature acquisition. In *Language acquisition: Knowledge representation and processing. Proceedings of GALA '97*, ed. by Antonella Sorace, Caroline B. Heycock, and Richard Shillcock. Edinburgh: HCRC.
- Badecker, William, and Kathleen Straub. 2002. The processing role of structural constraints in the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28:748–769.
- Boster, Carole Tenny. 1991. Children's failure to obey Principle B: Syntactic problem or lexical error? Ms., University of Connecticut, Storrs.
- Büring, Daniel. 2005. *Binding theory*. Cambridge: Cambridge University Press.
- Chien, Yu-Chin, and Kenneth Wexler. 1990. Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition* 1:225–295.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Clifton, Charles, Shelia Kennison, and Jason Albrecht. 1997. Reading the words *her*, *him*, and *his*: Implications for parsing principles based on frequency and on structure. *Journal of Memory and Language* 36:276–292.
- Cowart, Wayne, and Helen S. Cairns. 1987. Evidence for an anaphoric mechanism within syntactic processing: Some reference relations defy semantic and pragmatic constraints. *Memory and Cognition* 15:318–331.
- Crain, Stephen, and Cecile McKee. 1986. The acquisition of structural restrictions on anaphora. In *Proceedings of North East Linguistic Society (NELS) 16*, ed. by Stephen Berman, Jae-Woong Choe, and Joyce McDonough, 94–110. Amherst: University of Massachusetts, Graduate Linguistic Student Association.
- Crain, Stephen, and Rosalind Thornton. 1998. *Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.
- Deutsch, Werner, Charlotte Koster, and Jan Koster. 1986. What can we learn from children's errors in understanding anaphora? *Linguistics* 24:203–225.
- Elbourne, Paul. 2005. On the acquisition of Principle B. *Linguistic Inquiry* 36:333–365.
- Elbourne, Paul. 2008. The interpretation of pronouns. *Language and Linguistics Compass* 2:119–150.
- Evans, Gareth. 1980. Pronouns. *Linguistic Inquiry* 11:337–362.
- Gordon, Peter. 1996. The truth-value judgment task. In *Methods for assessing children's syntax*, ed. by Dana McDaniel, Cecile McKee, and Helen Cairns, 211–232. Cambridge, MA: MIT Press.
- Grimshaw, Jane, and Sara Rosen. 1990. Knowledge and obedience: The developmental status of the binding theory. *Linguistic Inquiry* 21:187–222.
- Grodzinsky, Yosef, and Tanya Reinhart. 1993. The innateness of binding and coreference. *Linguistic Inquiry* 24:69–101.
- Grolla, Elaine. 2005. Pronouns as elsewhere elements: Implications for language acquisition. Doctoral dissertation, University of Connecticut, Storrs.

- Grosz, Barbara, Aravind Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21:203–226.
- Guasti, Maria Teresa. 2004. *Language acquisition: The growth of grammar*. Cambridge, MA: MIT Press.
- Guasti, Maria Teresa, and Gennaro Chierchia. 1999/2000. Reconstruction in child grammar. *Language Acquisition* 8:129–170.
- Hamann, Cornelia, Odette Kowalski, and William Philip. 1997. The French ‘‘delay of Principle B’’ effect. In *Proceedings of the 21st annual Boston University Conference on Language Development*, ed. by Elizabeth Hughes, Mary Hughes, and Annabel Greenhill, 205–219. Somerville, MA: Cascadilla Press.
- Hamburger, Henry, and Stephen Crain. 1984. Acquisition of cognitive compiling. *Cognition* 17:85–136.
- Heim, Irene. 1998. Anaphora and semantic interpretation: A reinterpretation of Reinhart’s approach. In *The interpretive tract*, ed. by Uli Sauerland and Orin Percus, 205–246. MIT Working Papers in Linguistics 25. Cambridge, MA: MIT, MIT Working Papers in Linguistics.
- Hestvik, Arild, and William Philip. 1999/2000. Binding and coreference in Norwegian child language. *Language Acquisition* 8:171–235.
- Higginbotham, James. 1983. Logical Form, binding, and nominals. *Linguistic Inquiry* 14:547–593.
- Hulsey, Sarah, Valentine Hacquard, Danny Fox, and Andrea Gualmini. 2004. The question-answer requirement and scope assignment. In *Plato’s problem: Problems in language acquisition*, ed. by Aniko Csirmaz, Andrea Gualmini, and Andrew Nevins, 71–90. MIT Working Papers in Linguistics 48. Cambridge, MA: MIT, MIT Working Papers in Linguistics.
- Jakubowicz, Celia. 1984. On markedness and binding principles. In *Proceedings of North East Linguistic Society (NELS) 14*, ed. by Charles Jones and Peter Sells, 154–182. Amherst: University of Massachusetts, Graduate Linguistic Student Association.
- Kaufman, Diana. 1988. Grammatical and cognitive interactions in the study of children’s knowledge of binding theory and reference relations. Doctoral dissertation, Temple University, Philadelphia, PA.
- Kaufman, Diana. 1994. Grammatical or pragmatic: Will the real Principle B please stand? In *Syntactic theory and first language acquisition: Cross-linguistic perspectives*. Vol. 2, *Binding, dependencies, and learnability*, ed. by Barbara Lust, Gabriella Hermon, and Jaklin Kornfilt, 177–200. Hillsdale, NJ: Lawrence Erlbaum.
- Kazanina, Nina, Ellen Lau, Moti Lieberman, Masaya Yoshida, and Colin Phillips. 2007. The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language* 56:384–409.
- Kazanina, Nina, and Colin Phillips. 2001. Coreference in child Russian: Distinguishing syntactic and discourse constraints. In *Proceedings of the 25th annual Boston University Conference on Language Development*, ed. by Anna H.-J. Do, Laura Domínguez, and Aimee Johansen, 413–424. Somerville, MA: Cascadilla Press.
- Keenan, Ed. 1971. Names, quantifiers, and the sloppy identity problem. *Papers in Linguistics* 4:211–232.
- Kennison, Shelia. 2003. Comprehending the pronouns *her*, *him*, and *his*: Implications for theories of referential processing. *Journal of Memory and Language* 49:335–352.
- Kiguchi, Hirohisa, and Rosalind Thornton. 2004. Binding principles and ACD constructions in child grammars. *Syntax* 7:234–271.
- Koster, Charlotte. 1994. Problems with pronoun acquisition. In *Syntactic theory and first language acquisition: Cross-linguistic perspectives*. Vol. 2, *Binding, dependencies, and learnability*, ed. by Barbara Lust, Gabriella Hermon, and Jaklin Kornfilt, 201–226. Hillsdale, NJ: Lawrence Erlbaum.
- Lasnik, Howard. 1976. Remarks on coreference. *Linguistic Analysis* 2:1–22.
- Leddon, Erin, and Jeffrey Lidz. 2006. Reconstruction effects in child language. In *Proceedings of the 30th annual Boston University Conference on Language Development*, ed. by David Bamman, Tatiana Magnitskaia, and Colleen Zaller, 328–339. Somerville, MA: Cascadilla Press.

- Lee, Ming-Wei, and John N. Williams. 2006. The role of grammatical constraints in intra-sentential pronoun resolution. Ms., London Metropolitan University and University of Cambridge.
- Levinson, Stephen. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Lombardi, Linda, and Jaya Sarma. 1989. Against the bound variable hypothesis of the acquisition of Principle B. Paper presented at the annual meeting of the Linguistic Society of America, Washington, DC.
- Lust, Barbara, Julie Eisele, and Reiko Mazuka. 1992. The binding theory module: Evidence from first language acquisition for Principle C. *Language* 68:333–358.
- Lust, Barbara, Katherine A. Loveland, and R. Kornet. 1980. The development of anaphora in first language: Syntactic and pragmatic constraints. *Linguistic Analysis* 6:359–391.
- Matsuoka, Kazumi. 1997. Binding conditions in young children's grammar: Interpretation of pronouns inside conjoined NPs. *Language Acquisition* 6:37–48.
- McDaniel, Dana, Helen Cairns, and Jennifer Hsu. 1990. Binding principles in the grammars of young children. *Language Acquisition* 1:121–139.
- McDaniel, Dana, and Thomas Maxfield. 1992. Principle B and contrastive stress. *Language Acquisition* 2: 337–358.
- McKee, Cecile. 1992. A comparison of pronouns and anaphors in Italian and English acquisition. *Language Acquisition* 2:21–54.
- Nicol, Janet, and David Swinney. 1989. The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18:5–19.
- Padilla, Jose. 1990. *On the definition of binding domains in Spanish*. Dordrecht: Kluwer.
- Philip, William, and Peter Coopmans. 1996. The double Dutch delay of Principle B effect. In *Proceedings of the 20th annual Boston University Conference on Language Development*, ed. by Andy Stringfellow, Dalia Cahana-Amitay, Elizabeth Hughes, and Andrea Zukowski, 576–587. Somerville, MA: Cascadilla Press.
- Reinhart, Tanya. 1983. *Anaphora and semantic interpretation*. London: Croom Helm.
- Reinhart, Tanya. 2006. *Interface strategies*. Cambridge, MA: MIT Press.
- Runner, Jeffrey, Rachel Sussman, and Michael Tanenhaus. 2003. Assignment of reference to reflexives and pronouns in picture noun phrases: Evidence from eye-movements. *Cognition* 89:B1–B13.
- Runner, Jeffrey, Rachel Sussman, and Michael Tanenhaus. 2004. The influence of binding theory on the on-line reference resolution of pronouns. In *Proceedings of North East Linguistic Society (NELS) 34*, ed. by Keir Moulton and Matthew Wolf. Amherst: University of Massachusetts, Graduate Linguistic Student Association.
- Runner, Jeffrey, Rachel Sussman, and Michael Tanenhaus. 2006. Processing reflexives and pronouns in picture noun phrases. *Cognitive Science* 30:193–241.
- Russell, Bertrand. 1948. *Human knowledge: Its scope and limits*. London: Allen and Unwin.
- Sag, Ivan. 1976. Deletion and Logical Form. Doctoral dissertation, MIT, Cambridge, MA. Published, New York: Garland (1980).
- Savarese, Frederick. 1999. Studies in coreference and binding. Doctoral dissertation, University of Maryland, College Park.
- Sigurjónsdóttir, Sigríður, and Nina Hyams. 1992. Reflexivization and logophoricity: Evidence from the acquisition of Icelandic. *Language Acquisition* 2:359–413.
- Solan, Lawrence. 1983. *Pronominal reference: Child language and the theory of grammar*. Dordrecht: Reidel.
- Solan, Lawrence. 1987. Parameter setting and the development of pronouns and reflexives. In *Parameter setting*, ed. by Thomas Roeper and Edwin Williams, 189–210. Dordrecht: Reidel.
- Spenader, Jennifer, Erik-Jan Smits, and Petra Hendriks. 2007. Coherent discourse solves the pronoun interpretation problem. Ms., University of Groningen.

- Sturt, Patrick, Hamutal Kreiner, and Simon Garrod. 2005. Talk presented at the University of Maryland, College Park.
- Thornton, Rosalind. 1990. Adventures in long-distance moving: The acquisition of complex *wh*-questions. Doctoral dissertation, University of Connecticut, Storrs.
- Thornton, Rosalind, and Kenneth Wexler. 1999. *Principle B, VP-ellipsis, and interpretation in child grammar*. Cambridge, MA: MIT Press.
- Trueswell, John, Irina Sekerina, Nicole Hill, and Marian Logrip. 1999. The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition* 73:89–134.
- Utakis, Sharon. 1995. Quantification and definiteness in child grammar. Doctoral dissertation, CUNY, New York.
- Varlokosta, Spyridoula. 2000. Lack of clitic-pronoun distinctions in the acquisition of Principle B in child Greek. In *Proceedings of the 24th annual Boston University Conference on Language Development*, ed. by S. Catherine Howell, Sarah Fish, and Thea Keith-Lucas, 738–748. Somerville, MA: Cascadilla Press.
- Varlokosta, Spyridoula, and Joanna Dullaart. 2001. The acquisition of pronominal reference by Greek-Dutch bilingual children: Evidence for early grammar differentiation and autonomous development in bilingual first language acquisition. In *Proceedings of the 25th annual Boston University Conference on Language Development*, ed. by Anna H.-J. Do, Laura Domínguez, and Aimee Johansen, 780–791. Somerville, MA: Cascadilla Press.
- Wexler, Kenneth, and Yu-Chin Chien. 1985. The development of lexical anaphors and pronouns. *Papers and Reports on Child Language Development* 24:138–149.
- Williams, Edwin. 1977. Discourse and Logical Form. *Linguistic Inquiry* 8:101–139.

Department of Linguistics
University of Maryland
1401 Marie Mount Hall
College Park, MD 20742

staceyc@umd.edu
takahash@umd.edu
jlidz@umd.edu
colin@umd.edu