Where are the Cookies? Two- and Three-year-olds use Number-Marked Verbs to Anticipate Upcoming Nouns

Cynthia Lukyanenko

and

Cynthia Fisher

University of Illinois at Urbana-Champaign

Author note:

This research was supported by NIH grant HD054448, NSF grant BCS 06-20257, and NIH training grant T32-MH19990. A preliminary report of Experiment 2 appeared in the Proceedings of the 38th Annual Boston University Conference on Language Development. We are grateful to the children and parents who participated in this research, and also thank Renée Baillargeon, Kay Bock, Laurel Brehm, Gary Dell, and Kara Federmeier for many helpful comments.

Address for Correspondence: Cynthia Lukyanenko, Center for Language Science, The Pennsylvania State University, University Park, PA 16802. (cul31@psu.edu).

ABSTRACT

We tested toddlers' and adults' predictive use of English subject-verb agreement. Participants saw pairs of pictures differing in number and kind (e.g., one apple, two cookies), and heard sentences with a target noun naming one of the pictures. The target noun was the subject of a preceding agreeing verb in *informative* trials (e.g., *Where <u>are</u> the good cookies?*), but not in *uninformative* trials (*Do you see the good cookies?*). In Experiment 1, 3-year-olds and adults were faster and more likely to shift their gaze from distractor to target upon hearing an informative agreeing verb. In Experiment 2, 2.5-year-olds were faster to shift their gaze from distractor to target in response to the noun in informative trials, and were more likely to be fixating the target already at noun onset. Thus, toddlers used agreeing verbs to predict number features of an upcoming noun. These data provide strong new evidence for the broad scope of predictive processing in online language comprehension.

Abstract word count: 159

Keywords: language acquisition; syntax acquisition; agreement; grammatical number; comprehension; visual world

1. Introduction

Both young children and adults understand language incrementally: They integrate many aspects of the linguistic and non-linguistic context to assign meaning to words and sentences as they unfold, and even to make implicit predictions about upcoming words (e.g., Altmann & Kamide, 2007; Altmann & Steedman, 1988; Federmeier, 2007; Fernald, Zangl, Portillo, & Marchman, 2008; Friedrich & Friederici, 2005; Huang, Zheng, Meng, & Snedeker, 2013; Kamide, 2008; Kidd, White, & Aslin, 2011; Kukona, Fang, Aicher, Chen, & Magnuson, 2011; Mani & Huettig, 2012; Swingley, Pinto, & Fernald, 1999; van Berkum, 2008). Such predictions speed computation of the speaker's meaning, making possible rapid language comprehension and allowing children and adults to find sentences meaningful (despite frequent linguistic ambiguity) long before they are complete.

Recent influential proposals also suggest that *prediction may drive language learning itself*. Error-based learning accounts propose that learners routinely use their existing knowledge of language to predict upcoming words or their features, and that the error signal generated when inaccurate predictions are falsified by reality drives implicit learning about the categories and dependencies of the native language (e.g., Chang, Dell & Bock, 2006; Dell & Chang, 2014; Fine & Jaeger, 2013; Misyak, Christiansen & Tomblin, 2010; Saffran, 2001). On an error-based learning account, we learn language in part by (a) predicting future words or their features, (b) comparing our predictions to the words we actually encounter, and (c) refining our predictions to do better next time.

1.1. Prediction in early comprehension: Prior findings

To understand the role of error-based learning in acquisition, we must explore the nature and scope of prediction in young children's language processing. Error-based learning implies a strong form of prediction, the pre-activation of words or features not yet encountered (see Dell & Chang, 2014; Federmeier, 2007; Kutas, DeLong, & Smith, 2011, for reviews). Evidence that children can sometimes predict upcoming words before they begin is therefore important support for the plausibility of an error-based learning account.

To illustrate, young children use the rich semantic constraints provided by the content words in a sentence to anticipate likely upcoming words. Toddlers and preschoolers are quicker to look at a target object if its name follows a semantically restrictive verb (e.g., *drink the juice* as opposed to *take the juice*; Fernald et al., 2008; Mani & Huettig, 2012), and they unite the semantic constraints of agent and verb to anticipate a plausible direct object (e.g., *The pirate chases* … predicts an object that is both pirate-related and chaseable; Borovsky & Creel, 2014; Borovsky, Elman, & Fernald, 2012; Borovsky, Sweeney, Elman, & Fernald, 2014). These effects can be measured before the target word begins, yielding powerful evidence for the pre-activation of words or their features given strong semantic constraints, and thus initial evidence for the plausibility of error-based learning as a mechanism for language learning.

But error-based learning requires that this kind of pre-activation be widespread. To learn the syntactic categories and dependencies of their native languages, learners must generate predictions based on many kinds of linguistic evidence, and do so at many levels of constraint, not only in those cases licensing the strongest predictions (see Smith & Levy, 2013; Szewczyk & Schriefers, 2013, for discussion).

Children's ability to generate predictions from function words in a broad range of contexts is of particular interest, because of the potential for function words to guide syntax acquisition (e.g., Bernal, Lidz, Millotte, & Christophe, 2007; Cyr & Shi, 2013; Gerken, Landau, & Remez, 1990; Maratsos & Chalkley, 1980). Predictions from function words (such as determiners and auxiliary verbs) contrast sharply with those driven by content words (such as nouns and verbs). The verb *drink* could license predictions by virtue of both its grammatical properties and its concrete semantic content: It is a transitive verb, and thus in many contexts predicts a following noun phrase, but it also prompts thoughts of beverages, permitting listeners to activate a focused set of semantically-related nouns. In contrast, function words have very abstract semantic content, and thus may license predictions chiefly by virtue of their grammatical properties. For example, a determiner predicts that a noun will soon appear, and can cue abstract dimensions such as grammatical gender, number or definiteness—but not whether the noun might name something to drink or to chase. Thus predictions from function words and content words may drive learning about different aspects of the linguistic system.

There is compelling evidence that toddlers recruit determiners in online comprehension, despite their abstract nature (e.g., Bernal, Dehaene-Lambertz, Millotte & Christophe, 2010; Gerken & McIntosh, 1993; Kedar, Casasola, & Lust, 2006; van Heugten & Johnson, 2011; Zangl & Fernald, 2007). For example, 3-year-old Spanish speakers used gender-marked determiners to facilitate noun comprehension: They were quicker to look at a target picture (*Encuentra la pelota*$_{\text{FEM}}$, 'Find the ball') if the distractor picture's name differed in grammatical gender and thus could not follow the same determiner (*el zapato*$_{\text{MASC}}$; 'the shoe' vs. *la galleta*$_{\text{FEM}}$, 'the cookie'; Lew-Williams, & Fernald, 2007). Similarly, French-learning 2-year-olds used the gender (Melançon & Shi, in press; van Heugten & Shi, 2009) and number features of determiners in online comprehension (Robertson, Shi, & Melançon, 2012), shifting their gaze more quickly to a target picture if its name was preceded by function-word cues that were informative in the referential context.

1.2. The case of subject-verb agreement

Building on these findings, in the present work we sought to broaden the evidence for early predictive processing based on function words by focusing on the case of English subject-verb agreement. In English, as in many languages, the verb agrees in number with its subject noun phrase. In two experiments using a looking-while-listening task (e.g., Fernald et al., 2008), we examined 2.5- and 3-year-olds' and adults' comprehension of simple inverted sentences in which a number-marked verb preceded its subject, as in (1), and thus could provide advance information about the subject's grammatical number.

(1) a. Where <u>are</u> the good cookies?

b. Where <u>is</u> the good apple?

In the experimental condition, each sentence accompanied a pair of pictures differing in number and object-type (e.g., two cookies, one apple), as shown in the left column of Figure 1. In this referential context, if children routinely use function words to predict upcoming material, they might use *Where <u>are</u>* (as opposed to *Where <u>is</u>*) to predict a plural noun phrase, and thus reject a single apple as a referent even before the target word began. An uninformative adjective (e.g., *good*) was added to allow more time for such predictions to be measured. In the control condition (Figure 1, right column) no predictions were possible because the two pictures matched in number (e.g., two cookies and two apples in plural trials; one cookie and one apple in singular trials). This referential context renders the agreeing verb-form uninformative; thus any anticipatory effects of the agreeing verb should disappear. Finding that children can exploit the constraints of agreeing verbs in this way would provide evidence that early language comprehension is inherently and broadly anticipatory, and thus new support for the hypothesis that we learn language in part by learning to predict the future.

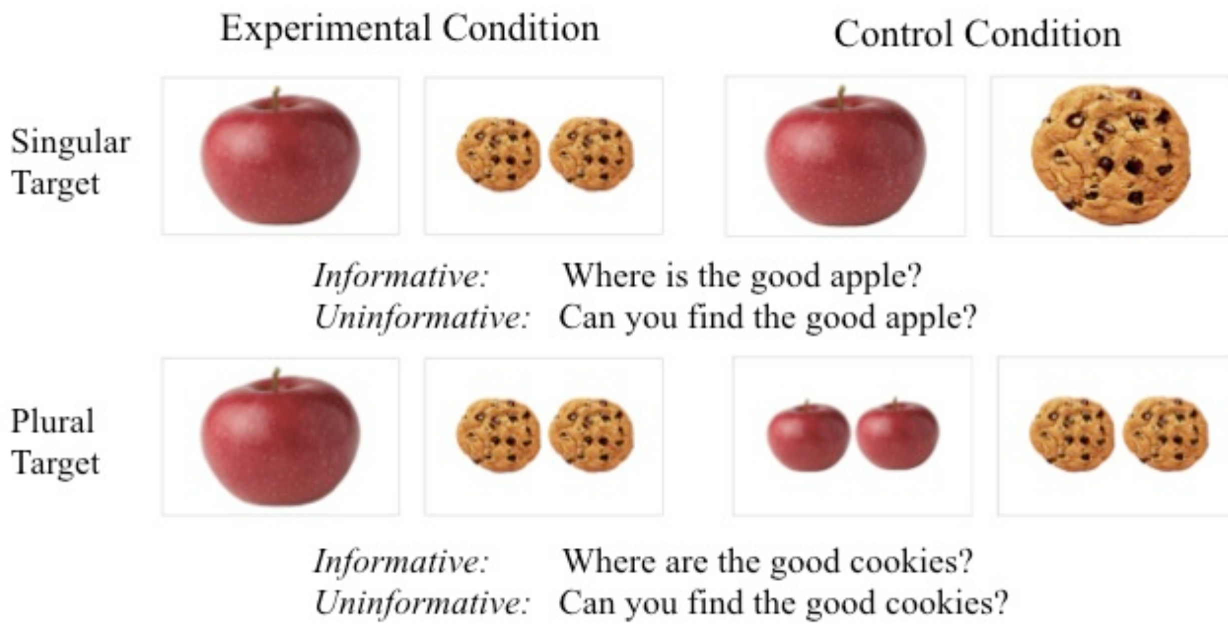

Figure 1. Trial structure. Children were presented with two pictures and an accompanying sentence. In Informative trials, the sentence contained a number-marked verb that preceded its (plural or singular) subject noun-phrase. The Control condition involved the same recorded sentences, but differed in that the distractor and target pictures always matched in number.

Subject-verb number agreement provides a particularly useful test for the scope of predictive processing in early comprehension for two reasons.

1.2.1. Agreeing verbs can precede or follow their subjects. First, unlike many other function words, subject-verb agreement spans two major sentence constituents, the subject and the verb phrase. In this respect, it differs from the determiner-noun dependency that has been the focus of past research on toddlers' use of function words in online comprehension, and provides a more stringent test of children's ability to use function words predictively. The determiner-noun dependency is syntactically local (within the noun phrase) and fixed in order (e.g., English or Spanish determiners precede nouns). A determiner therefore permits a strong directional prediction about the grammatical category of an upcoming word (Arnon & Ramscar, 2012; Mariscal, 2009). For example, Thorpe and Fernald (2006) reported that English *the* was immediately followed by a noun 93% of the time in an analysis of child-directed speech. In contrast, agreeing verbs follow their subjects in English declarative sentences, but precede them in sentences with inverted orders, including WH-questions as in (1), and locatives (*There are the good cookies!*)[1]. Such inversions are common in speech to children (e.g., Broen, 1972; Newport, Gleitman & Gleitman, 1977). To illustrate, in a corpus of child-directed speech (see Appendix A), *are* and *is* followed their subjects 35% of the time, and preceded them 65% of the time[2]. Because they can either precede or follow their subjects, agreeing verbs permit much less consistent predictions about the grammatical features of upcoming words than do determiners.

1.2.2. Notional vs. grammatical number. Second, like other function words, agreeing verb forms such as *is* and *are* license relatively abstract predictions about their subjects (*singular* vs. *plural*, not *beverage* vs. *food*), and do so chiefly by virtue of their grammatical properties, not their semantic content. In principle, plurality could be computed in notional or grammatical terms: Children might use *are*, for example, to predict that an upcoming subject noun phrase will refer to more than one thing (notional plurality), that it will be grammatically plural, or both. However, the relationship between verb-form and notional plurality holds only when the subject is a count noun, and also depends on grammatical person. Non-count nouns (*toast, scissors*) and second-person subjects (*you*) select verb forms that do not vary with real-world number: We say *Where are the scissors?* and *Where are you?* whether we seek one referent or many. Second-person subjects are common in casual speech, so *are* often appears in contexts where it does not reflect the notional number of the subject. In the analysis of child-directed speech mentioned

---

[1] Main verbs other than the copula do not precede their subjects in questions (though they can in locative inversions as in *There goes the bunny!*), but their tense and agreement affixes do, hosted by auxiliary verbs (e.g., *Do_ they like jam? vs. Does she like jam?*).

[2] These figures were based on counts from maternal speech to Adam (Adam01-Adam55, aged 2;3 - 5;2; Brown, 1973), available in the CHILDES database (MacWhinney, 2000). This sample, which comprised 20,344 maternal utterances, and 92,318 word-tokens, was searched for occurrences of the words *are* and *is*. This search yielded 745 tokens of *are,* and 1645 tokens of *is*.

above, *are* appeared with *you* as its subject in 49% of cases.

In sum, an agreeing verb provides a pointer to the grammatical (rather than notional) number of its subject, which may precede or follow it. If we find that young children nevertheless use subject-verb agreement to pre-activate features of an upcoming word, this will provide strong new evidence for the broad scope of prediction in early language comprehension,.

1.3 Prediction vs. integration

A second goal of the present research was to strengthen the evidence for pre-activation based on function-word cues. As noted earlier, the clearest evidence for prediction comes from cases in which anticipation of a target word can be measured before that word is heard. Effects of preceding context measured after target-word onset are often open to alternative interpretations: Rather than driving *prediction*, the preceding context could ease the *integration* of the target word into the representation of the sentence, once the target word is encountered (e.g., Federmeier, 2007; Kutas et al., 2011; Smith & Levy, 2013). The difference between prediction and integration is subtle. Both mechanisms permit incremental interpretation; they differ only in whether the linguistic context is used to pre-activate words that have not yet been experienced (prediction), or whether the context constrains word identification only after word candidates are activated by bottom-up perceptual evidence (integration). This distinction is central to error-based learning accounts, which propose that the difference between what we predict and what we perceive provides an error signal for learning.

At present, clear evidence of advance prediction in toddlers' sentence comprehension is mostly limited to cases in which the semantic content of content words strongly restricts plausible continuations (as in *drink the juice;* e.g., Borovsky et al., 2012; Fernald et al., 2008; Mani & Huettig, 2012). Children's use of determiners in online sentence interpretation (Lew-

Williams & Fernald, 2007; Melançon & Shi, in press; Robertson et al., 2012; van Heugten & Shi, 2009) could reflect prediction, and is often described in this way. However, in most prior studies, facilitation due to the determiner was assessed in analysis intervals that included the noun, yielding no direct evidence for pre-activation of suitable nouns based on the determiner.

A few studies have tested for such pre-activation, and here the findings are mixed. Robertson et al. (2012) found effects of French number-marked determiners (*les* vs. *le*) on 2-year-olds' comprehension after, but not before, target-noun onset; this pattern is thus open to both prediction and integration accounts. Interestingly, Dahan, Swingley, Tanenhaus and Magnuson (2000) found the same pattern in adults' incremental use of gender-marked determiners: Adults used gender-marked determiners to limit the set of phonological competitors they considered in interpreting a target word's onset (thus avoiding looks to *le bouton*$_{\text{MASC}}$ upon hearing the initial sounds of *la bouteille*$_{\text{FEM}}$), but showed no tendency to look toward objects with gender-matching names in response to the determiner itself, before target-noun onset. This yielded evidence that listeners used the determiner to interpret the initial sounds of the word, but not that they predicted grammatically suitable words in advance. Two other reports, in contrast, yielded evidence of prediction based on function words. Melançon and Shi (in press) found that 2.5-year-old French-learners looked longer at a familiar target picture when its name was preceded by an informative gender-marked article, and did so before target-noun onset. Likewise, Tsang and Chambers (2011) found that Cantonese-speaking adults used shape classifiers to predict grammatically compatible object names, looking more at objects whose names could follow the classifier they heard, before target-noun onset.

What might explain this mixed evidence? One possibility is that language comprehension is broadly predictive in both toddlers and adults, but that the short intervals typically tested (the interval between the onset of a determiner and the following noun) did not provide enough time, given the abstract predictions licensed by function words, to generate evidence of pre-activation (see Dahan et al., 2000).

There is some support for this view in the existing data. The reports that yielded evidence for pre-activation based on a function word included longer intervals for measurement between the function word and the onset of the target noun than those that yielded negative results (see also Kukona et al., 2011). The classifiers tested by Tsang and Chambers (2011) spanned a slightly longer interval than the gender-marked determiners used by Dahan et al. (2000), and Melançon and Shi (in press) interposed an uninformative adjective between the determiner and noun (e.g., *le joli ballon* 'the pretty ball'). Perhaps previous failures to detect predictive effects of function words are in part a consequence of the brief observation intervals typically tested, rather than a fundamental fact about language processing.

1.4. The development of subject-verb agreement

To use agreeing verbs predictively, children must know something about the English dependency between subject and verb, and about the referential consequences of linguistic number. Existing data suggest that children possess both of these prerequisites early in life.

First, both listening-preference and production data show sensitivity to the distributional consequences of subject-verb agreement. Infants under 2 years old listen longer to grammatical than to ungrammatical combinations of number-marked nouns and verbs (English: Soderstrom, 2002; Soderstrom, Wexler & Jusczyk, 2002; Soderstrom, White, Conwell & Morgan, 2007; French: Nazzi, Barrière, Goyet, Kresh & Legendre, 2011; van Heughten & Shi, 2010). The production data are complicated by toddlers' well-known tendency to omit function words (Brown, 1973; de Villiers & de Villiers, 1973). However, when children do produce inflected verbs, copulas, or auxiliaries, they typically use the right form (e.g., Keeney & Wolfe, 1972; Leonard, Eyer, Bedore & Grela, 1997; Wexler, 2011). The speech of children 2.5 to 3 years old shows emerging command of agreement, though error rates tend to be higher with plural forms and in inverted contexts (e.g., Leonard et al., 1997; Rispoli, Hadley & Holt, 2009; Rissman, Legendre & Landau, 2013; Rubino & Pine, 1998; Theakston, Lieven, Pine, & Rowland, 2005; Theakston & Rowland, 2009).

Second, comprehension data show that young English-speakers attach number meaning to at least some linguistic markers of number by about age 2. For example, at 2.5 years, children used number morphology on nouns to identify the referents of novel nouns (*Look at the gooms*; Jolly & Plunkett, 2008). Two-year-olds used multiple cues to number in sentences such as *There are some blickets* (vs. *There is a blicket_*) to direct their attention to a many- as opposed to a single-object display (Kouider, Halberda, Wood, & Carey, 2006; see also Wood, Kouider, & Carey, 2009). Moreover, analyses of the 2-year-olds' visual fixations over time showed that a preference for the target (one object in singular trials, multiple objects in plural trials), emerged before the onset of the number-marked novel noun itself, showing the predictive use of function words — either the agreeing verb, the quantifier, or both (*are some* vs. *is a*). The sentences in this experiment, however, made it difficult to disentangle effects of the agreeing verb from those of the adjacent quantifier or determiner. This result thus provides additional evidence that toddlers use function words predictively, but could be driven by children's interpretation of a number-marked determiner rather than the less direct pointer provided by an agreeing verb.

Some evidence that seems to support this concern comes from another line of research suggesting that verb agreement is difficult to interpret on its own. In comprehension tasks with an explicit judgment component, preschoolers often fail to use verb agreement itself as a cue to notional number (e.g., Brandt-Kobele & Höhle, 2010; Johnson, de Villiers & Seymour, 2005; Keeney & Wolfe, 1972; Leonard, Miller & Owen, 2000; Miller, 2012). For example, 3- and 4-year-old English speakers had great trouble judging which of two pictures matched a sentence when the critical number information was indicated only by an auxiliary verb (*The sheep is/are jumping*) or by regular agreement inflections on verbs (*The sheep jump/jumps*), but performed quite well if number was marked within the noun phrase, by a number-marked determiner (*This/These sheep jumped*) or the nominal plural inflection itself (*The girl/girls jumped*; Nicolaci-da-Costa & Harris, 1983, 1984).

Why might 3- and 4-year-olds fail to use agreeing verbs to infer notional number in comprehension, even though they honor subject-verb agreement in their own speech[3]? We would argue that this pattern emerges because, as discussed above, verb agreement does not directly reflect notional number, but primarily reflects the grammatical number of the subject (e.g., Bock & Middleton, 2011; Brandt-Kobele & Höhle, 2010; Johnson et al., 2005; Keeney & Wolfe, 1972). Number words (*one, two*), quantifiers (*a, some, this, these*) and nominal plurals, in contrast, are linked more directly with number semantics and are more readily interpreted as direct indicators of number meaning. This difference might particularly affect performance in comprehension tasks requiring an explicit judgment, which might prompt children to rely on metalinguistic knowledge about which words convey notional number.

Taken together, these findings support the hypothesis that verb agreement in English

[3] The exception to this pattern is the success of 2.5-year-old French learners (Legendre, Barriere, Goyet & Nazzi, 2010; Legendre et al., 2014): Children more often chose a video of two boys (rather than one) kissing a novel object when hearing *Ils embrassent le tac* (‘They kiss the tac’) as opposed to *Il embrasse le tac* (‘He kisses the tac’). These sentences differ only in the consonant *–z* (of *Ils*) joined via liaison to the vowel-initial verb *embrasser*. Legendre et al. argue that *ils*, traditionally analyzed as a clitic pronoun, is becoming a pre-verbal agreement marker. Such a historical shift, from free-standing pronoun to clitic, and then clitic to agreement marker, is cross-linguistically common. If *ils* has completed this change, then these French-learning toddlers used verb agreement to infer number meaning in an explicit task. However, another interpretation is that the children still interpret *ils* as a plural pronoun (Smouse, Gxilishe, de Villiers & de Villiers, 2012).

plays a primarily syntactic role in language processing: Verb agreement morphology reflects the grammatical number of the subject, but itself carries little or no number meaning. However, the status of agreement as a primarily syntactic cue does not preclude its predictive use in comprehension. Children could use a number-marked verb form to anticipate the grammatical features of an upcoming subject noun-phrase. Evidence that they can do so would provide new support for the broad role of prediction in early comprehension.

2. Experiment 1

In Experiment 1 we tested 3-year-old children's and adults' ability to use an agreeing verb-form to anticipate the number features of its subject noun phrase in inverted sentences as shown in Figure 1. If listeners hear *Where are the good cookies?* or *Where is the good apple?* while viewing two cookies and one apple*,* they could use the form of the verb (*are* vs. *is*) to anticipate which of the two pictures is about to be named. These anticipatory effects should disappear in the control condition, in which the two pictures show the same number of objects.

2.1. Method

2.1.1. Participants. Sixty-four 3-year-olds (34.2-42.4 months; *M* = 37.8; 32 girls) and 48 college students (18-27 years; M = 20.1; 30 women) participated; all were native English speakers. Half of the children and adults were randomly assigned to the experimental and control conditions. Another 5 children were excluded because of a reported language delay (1), or because they declined to participate (1) or were missing too many trials due to inattentiveness or parental interference (3; see Coding below). Another 2 adults were excluded because of missing trials due to inattentiveness (1) or because glare on their glasses made coding impossible (1). Children's productive vocabularies, measured using the short form of the MacArthur-Bates CDI (Level III; Fenson et al. 2007), ranged from 12 to 99 (median = 75).

2.1.2. Stimuli. The stimuli were sentences containing eight familiar object-names, all count nouns, accompanied by photographs, as in Figure 1. Sentences were recorded by a female native English speaker. The target noun was always at the end of the sentence. In informative trials, the target noun phrase was the subject of a preceding number-marked verb (e.g., *Where are the good cookies?*); in uninformative trials, it was not (e.g., *Can you find the good cookies?*). To increase variety, we used two informative sentence preambles, a WH-question (*Where is/are the…?*) and a locative (*There is/are the…!*), and two uninformative preambles (*Can you find the…?; Oh, look at the…!*). Familiar adjectives preceded the target nouns, allowing time to observe anticipatory looks before the noun itself could be identified. In the experimental condition the pictured objects always differed in number, as shown in the left column of Figure 1 (e.g., one apple, two cookies). Because informative and uninformative trials involved different recorded sentences, we included a control condition in which another group of children heard the same recorded sentences, but saw pictures that always matched in number (e.g., one apple, one cookie in singular trials; or two apples, two cookies in plural trials; right column of Figure 1). In this referential context, the number-marked verbs were unhelpful, and children should be no quicker to look at the target in trials with a number-marked verb than in those without[4].

Pictured items appeared in yoked pairs (*dog-baby, cat-turtle, bike-truck, apple-cookie*); each item served 4 times as target and 4 times as distractor. Each picture showed either one large single object or two slightly smaller identical objects (see Figure 1). Each object kind was represented by two exemplars (e.g., different dogs) that appeared in different trials. Participants

[4] We manipulated referential context between participants because in pilot work using a within-participants design, in trials in which the number of objects on the two sides differed, children showed a baseline bias for the plural picture. Such a bias might reduce our ability to measure effects of an agreeing verb. The current design, in which the number of objects either never matched in number (experimental) or always did (control), eliminated that baseline bias. We speculated that this predictable trial structure made it easier for children to determine the number and identity of the objects on each screen.

received 16 informative and 16 uninformative trials, each with 8 singular and 8 plural targets. The left-right position of the target and of the two-object picture in the experimental condition were counterbalanced with item and trial-type. Trial order was pseudo-randomized such that item pairs were never repeated in adjacent trials, there were no more than 3 trials of the same type in a row (singular vs. plural, informative vs. uninformative, target on left vs. right, two-object picture on left vs. right), and each item appeared equally often as target and distractor in the first and second halves of the procedure. Three single-picture filler trials were interspersed as breaks among the 32 critical trials. In addition, in the adult version of the task, 32 trials for a separate experiment were pseudo-randomly interspersed as fillers. These were included in part to pre-test items for a planned experiment with children.

2.1.3. Apparatus and procedure. Participants sat about 4 feet from a 50-inch television in a dimly lit room. Curtains to the child's left and right, and behind the television, hid the rest of the room. Children sat on a parent's lap, and parents wore opaque glasses. On each trial, two pictures (11.25" tall x 17.5" wide) appeared about 8 inches apart, aligned with the left and right edges of the screen. A camera beneath the screen recorded participants' eye-movements. Adult participants were asked to point to the named target picture in each trial; this was intended to promote attentiveness. Child participants were given no explicit task.

In each trial, the pictures were visible for 7 s. Recorded sentences were aligned so that the onset of the determiner (*Where are <u>the</u> good cookies?*; *Can you find <u>the</u> good cookies?*) occurred 3 s after the pictures appeared; thus speech began approximately 2 s into the trial. The critical sentence in each trial was followed by a second question that began 5 s after picture onset (e.g., *Do you like it/them?*); these were included to help maintain children's attention to the pictures. Trials were separated by a 1 s blank-screen interval.

2.1.4. Coding. We coded where participants looked (left, right, away) during each trial, from silent video. Coders viewed videos frame-by-frame in Quicktime, and coded visual fixations by hand, recording the onset of each shift in fixation. Reliability was assessed for 25% of the participants; coders agreed on 96% of video frames for the children and 94% for the adults. Trials were eliminated from analysis if more than 50% of the 7-s trial was spent looking away or was uncodeable (children: 145 trials; adults: 23 trials), if participants' or parents' speech obscured the critical sentence (children: 51 trials), or, for adult participants who were asked to point, if the participant pointed to the distractor (9 trials). This resulted in the elimination of 196 of 2048 trials from the children's data (9.6%), and 32 of 1536 trials from the adults' data (2.1%). Four participants' data were excluded because more than 4 of the 8 trials of each type (informative-plural, informative-singular, uninformative-plural, uninformative-singular) were eliminated (3 children, 1 adult).

2.1.5. Measures. Before the informative verb arrived, listeners had no information about which picture would be named. Thus, at this point participants should be fixating the target and the distractor about equally often. For trials in which the participant was fixating the distractor (distractor-initial trials), the correct response to an informative agreeing verb was to look away from the distractor toward the target (e.g., moving away from a single apple upon hearing *Where are the...*). Accordingly, in distractor-initial trials we measured (a) the *latency* of the first gaze shift from distractor to target (e.g., Fernald et al., 2008; Lew-Williams & Fernald, 2007), and (b) the *probability of shifting from distractor to target* in a pre-noun window (e.g., Altmann & Kamide, 1999; Thorpe & Fernald, 2006).

Distractor-initial trials were identified based on fixations at determiner onset. We chose determiner onset as our anchor-point because the determiner was present in both informative

(*Where are* <u>*the*</u> *good cookies?*) and uninformative sentences (*Can you find* <u>*the*</u> *good cookies?*), and in informative sentences it immediately followed the number-marked verb. 892 of the children's trials (48% of 1852 included trials) and 732 of the adults' trials (49% of 1504 included trials) were distractor-initial trials.

The latency of the first distractor-to-target shift was analyzed only in those distractor-initial trials with a direct shift to the target, eliminating trials in which participants looked away before fixating the target. Analyses of eye-movements also typically exclude saccades launched less than 200 ms after stimulus onset for adults, and less than 300 ms after stimulus onset for children (e.g., Fernald et al., 2008). In our materials, the number-marked verb in informative trials began on average 211 ms (range 156-265 ms, *is* M = 234 ms, *are* M = 189 ms) before the determiner. For 3-year-olds, we therefore calculated the latency of the first shift to the target in a 1500-ms window extending from 100 to 1600 ms after determiner onset; this captured latencies about 300 to 1800 ms after verb onset. For adults, we calculated latencies in a window of equal length, but beginning at determiner onset; this captured latencies about 200 to 1700 ms after average verb onset. 797 of children's distractor-initial trials (89% of 892 distractor-initial trials) and 697 of the adults' (95% of 732 distractor-initial trials) contained a direct distractor-to-target shift within the 1500-ms interval. Because average noun onset occurred well before the end of the latency window (M = 576 ms after determiner onset; range 433-700), our latency measure included shifts generated on the basis of information about the noun itself. Thus, this measure allowed us to ask whether children and adults used agreeing verbs incrementally (via either prediction or integration processes) to facilitate identification of the following subject noun, as shown for determiners in prior work (e.g., Lew-Williams & Fernald, 2007). If so, listeners should be quicker to shift to the target in informative than uninformative trials, in the experimental condition only.

The probability of a distractor-to-target shift before target-noun onset was measured in a 633-ms pre-noun window extending from 100 ms after determiner onset through 300 ms after the earliest target-noun onset for children, and from determiner onset through 200 ms after the earliest target-noun onset for adults. For each distractor-initial trial, we measured whether participants shifted their gaze to the target at any point within this window (a correct response), or did not shift to the target before noun onset. Trials were excluded if participants shifted their gaze before the start of the pre-noun window, or if they looked away from both pictures and did not look back before the end of the window. Unlike a latency measure, this accuracy measure permitted us to include trials both with and without a distractor-to-target shift, eliminating problems of data sparseness that would result from excluding trials without a shift in this brief window[5]. If listeners use number-marked verbs to predict features of an upcoming noun, distractor-to-target shifts should be more frequent in this pre-noun window in informative than in uninformative trials, in the experimental condition only.

2.2. Results

Figure 2 shows looks to the target as a proportion of looks to either picture in 33-ms time intervals measured from determiner onset, by trial-type (informative vs. uninformative), condition (experimental vs. control), and age group (3-year-olds vs. adults). As the Figure shows, both children and adults looked about equally at the two pictures at determiner onset, but in the experimental groups, target fixations increased earlier in informative than in uninformative trials.

[5] We calculated this accuracy measure only in distractor-initial trials because we had no clear prediction about responses in target-initial trials. In target-initial trials, the correct response to an informative agreeing verb might be to stay on target, suppressing shifts to an incongruent distractor picture. However, a number of findings suggest that neither children nor adults routinely keep a detailed representation of a display in working memory, but rather interrogate it by moving their eyes to the relevant areas when they need information (Swingley & Fernald, 2002; Ballard, Hayhoe & Pelz, 1995). Thus, to the extent that participants did not hold the non-fixated picture in working memory, there was no reason for them to suppress target-to-distractor shifts.

This informative-trial advantage began well before noun onset. No such informative-trial advantage appeared in the control condition, in which the referential context made the agreeing verbs uninformative.

To preview our results, this pattern was reflected in both measures described above: Listeners in the experimental but not the control group were faster to shift to the target, and more likely to shift to the target in the pre-noun window in informative than in uninformative trials.

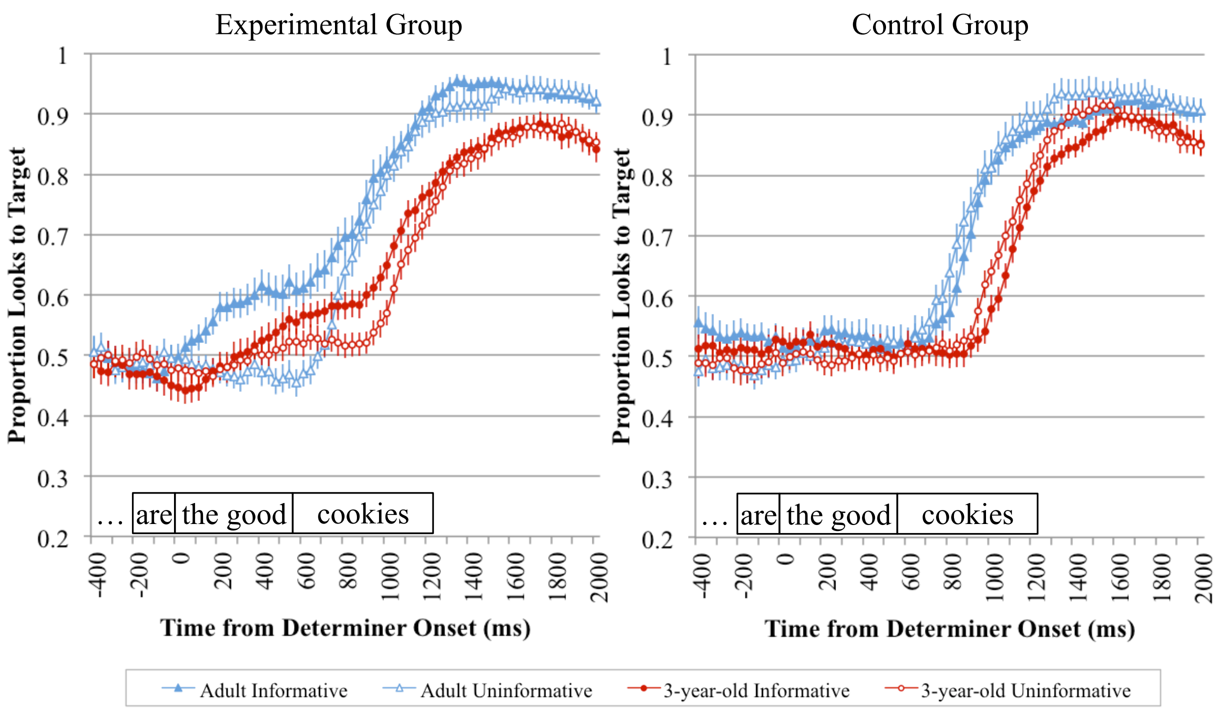

Figure 2. Mean (se) proportion fixations to the target in Experiment 1 by trial type (informative vs. uninformative), age group (adult vs. child) and condition (experimental vs. control). The x-axis shows time measured from determiner onset; average verb onset in informative trials preceded determiner onset by approximately 200 ms.

2.2.1. Latency of first distractor-to-target shift. As shown in Figure 3, latencies to shift from distractor to target in distractor-initial trials, measured from determiner onset, were shorter in informative than in uninformative trials for both 3-year-olds and adults in the experimental but not the control condition. This suggests that both children and adults used the agreeing verb to facilitate processing; for the children, the effect appeared stronger in trials with plural targets. Such a plural-singular asymmetry is not unexpected, as in some sense a singular is inherently less informative in our referential context than a plural: A picture of two cookies also shows one cookie, which a speaker might in principle single out for mention.

To test this pattern, we fit a mixed-effects regression model of shift latency in R (R Core Team, 2013) using the *lmer()* function of the *lme4* package (Bates, Maechler & Bolker, 2013). Predictor variables included the within-participants factors trial-type (informative, uninformative) and target plurality (singular, plural), the between-participants factors condition (experimental, control) and age group (3-year-olds, adults), and all interactions. All predictors were coded using mean-centered effects coding. Random intercepts were included for participants and target nouns. The model with the maximal random effects structure was attempted but failed to converge; we therefore used a forward best-path method to determine which random slopes to include (Barr, Levy, Scheepers & Tily, 2013). None of the random slopes met the inclusion criterion ($\alpha = .2$). The syntax for all main models is in Appendix B. The results are shown in Table 1, with χ^2 statistics and p-values obtained by model comparison.

This analysis revealed a significant main effect of age group, reflecting children's slower latencies (and the difference in measurement-window offset for children and adults, see Measures, above), a main effect of trial-type, and crucially, a significant interaction of condition and trial-type, showing that the difference between informative and uninformative trials differed across the experimental and control conditions. We also found a marginal 3-way interaction of condition, trial-type, and target plurality; this presumably reflects the numerical plural-singular asymmetry in the children’s data, visible in Figure 3. The four-way interaction of condition, trial-type, target plurality, and age was not significant.

Follow-up comparisons were conducted using treatment coding to extract the simple main effect of trial-type within each cell of the design. The simple main effect of trial-type (informative vs. uninformative) was significant in the experimental condition for both 3-year-olds ($\chi^2(1) = 15.4$, $p < .0001$) and adults ($\chi^2(1) = 6.1$, $p = .01$), but not in the control condition (3-year-olds: $\chi^2(1) = 0.1$, $p = .76$; adults: $\chi^2(1) = 0.6$, $p = .44$). Thus, as predicted, children and adults in the experimental condition demonstrated an informative-trial advantage.

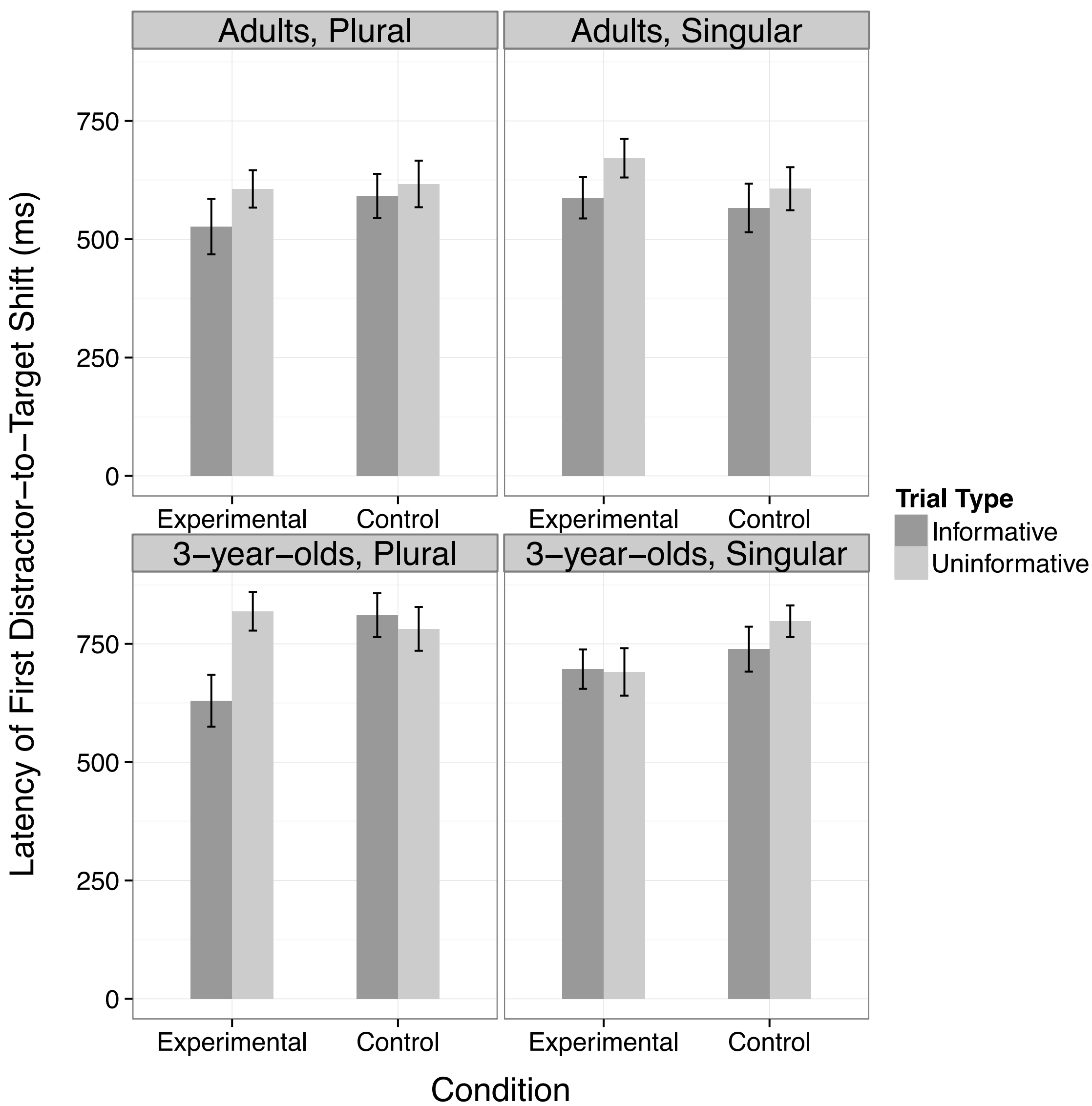

Figure 3. Mean (se) latency of the first distractor-to-target shift in distractor-initial trials,

Experiment 1, measured from determiner onset, by trial type (informative vs. uninformative trials), target plurality (singular vs. plural), age group (adults vs. children) and condition (experimental vs. control).

Table 1. Estimates for mixed effects model of distractor-to-target shift latency measured from determiner onset, Experiment 1. Trial type and target plurality are within-participants factors, and condition and age group are between-participants factors. Statistically significant and marginal effects are shown in bold. (N=1494).

Fixed Effects

	Model Summary			Model Comparison		
	β	SE	t	χ^2	df	*p*
Intercept	668.6	18.9	35.41	**44.23**	**1**	**<.0001**
Age Group	-135.2	28.2	-4.79	**21.56**	**1**	**<.0001**
Condition	-35.5	28.1	-1.26	1.64	1	0.20
Trial Type	-68.9	18.3	-3.77	**14.29**	**1**	**0.0002**
Plurality	-9.3	18.3	-0.51	0.26	1	0.61
Age Group × Condition	78.8	56.4	1.40	2.00	1	0.16
Age Group × Trial Type	14.2	36.6	0.39	0.15	1	0.70
Age Group × Plurality	-48.1	36.7	-1.31	1.73	1	0.19
Condition × Trial Type	-95.8	36.5	-2.62	**6.92**	**1**	**0.009**
Condition × Plurality	-58.8	36.6	-1.61	2.58	1	0.11
Trial Type × Plurality	-57.6	36.6	-1.57	2.52	1	0.11
Age Group × Condition × Trial Type	63.1	73.2	0.86	0.76	1	0.38
Age Group × Condition × Plurality	-93.7	73.3	-1.28	1.66	1	0.20
Age Group × Trial Type × Plurality	49.4	73.3	0.67	0.45	1	0.50
Condition × Trial Type × Plurality	-138.3	73.2	-1.89	**3.63**	**1**	**0.057**
Age Group × Condition × Trial Type × Plurality	192.7	146.7	1.31	1.73	1	0.19

Random Effects

	s^2
Participant (intercept)	12546.6
Target (intercept)	1277.8

Previous work has revealed correlations between vocabulary and measures of speech processing efficiency (e.g., Lew-Williams & Fernald, 2007; Mani & Huettig, 2012): Children with higher vocabularies tend to be quicker to use available linguistic cues to locate a target picture. Unfortunately, the present data did not permit us to assess such effects: The children's vocabularies clustered near the top of the 100-item scale provided by the CDI-III (with a median of 75), limiting the degree to which these scores could correlate with other measures. We will return to this point in Experiment 2.

In sum, both children and adults were quicker to shift their gaze to the target picture in trials with an informative agreeing verb. This effect closely resembles the facilitation from determiners or classifiers observed in previous experiments in other languages (Lew-Williams & Fernald, 2007; Melançon & Shi, in press; Robertson et al., 2012; Tsang & Chambers, 2011; van Heugten & Shi, 2009), and provides new evidence for the online use of morphosyntactic cues in comprehension. However, because the 1500-ms window in which we measured shift latencies included the noun itself, this informativeness advantage could reflect predictive use of the agreeing verb or the ease of integrating information about the noun into the preceding context. To seek strong evidence for pre-activation, we examined effects of the informative verb in the pre-noun window, before the target noun itself could have influenced fixations.

2.2.2. Shift probability. Figure 4 shows the mean proportion of distractor-initial trials that included a shift to the target during the pre-noun window, by trial-type, condition, age group and target plurality. Both children and adults made more distractor-to-target shifts in informative than uninformative trials in the experimental but not the control condition. In the children's data, this tendency again appeared more pronounced in the plural trials.

This pattern was supported by a binomial mixed-effects regression model of shift probability, fit in R using the *glmer()* function of the *lme4* package. Predictor variables were as described above. A model with the full random-effects structure again failed to converge, and no random slopes met the inclusion criterion (α = .2). Model results are shown in Table 2.

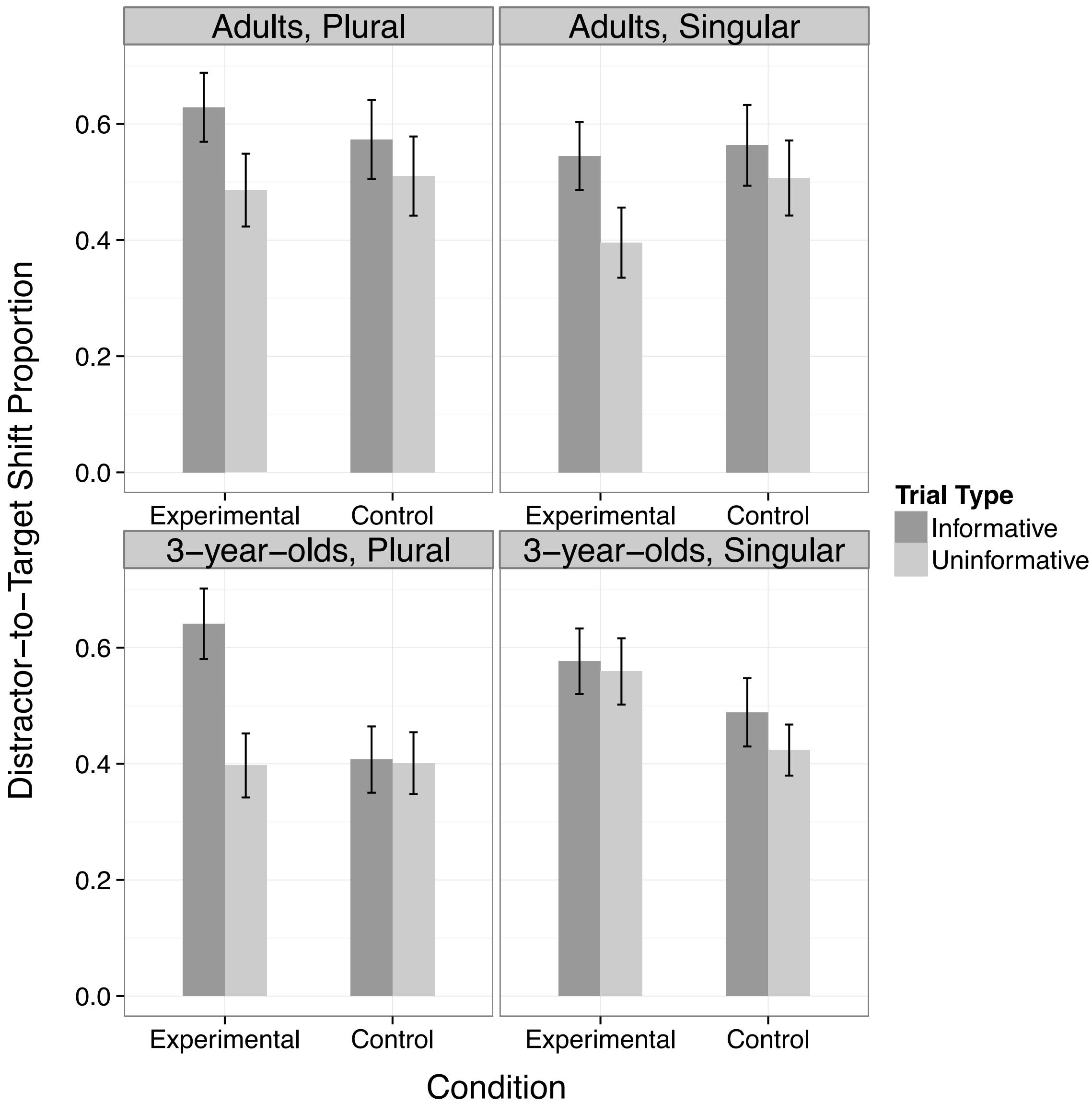

Figure 4. Mean (se) proportion of trials including a distractor-to-target shift in fixation during the pre-noun window in Experiment 1, by trial type (informative vs. uninformative), target plurality (singular vs. plural), age group (adults vs. children) and condition (experimental vs. control).

Table 2. Estimates for binomial mixed effects model of distractor-to-target shift probability, Experiment 1. Trial type and target plurality are within-participants factors, and condition and age group are between-participants factors. Statistically significant and marginal effects are shown in bold. (N=1544).

Fixed Effects

	Model Summary			Model Comparison		
	β	SE	z	χ^2	df	*p*
Intercept	0.056	0.095	0.59	0.34	1	0.56
Age Group	0.057	0.157	0.36	0.13	1	0.72
Condition	0.204	0.156	1.31	1.67	1	0.20
Trial Type	0.443	0.107	4.14	**16.60**	**1**	**<.0001**
Plurality	0.044	0.107	0.41	0.16	1	0.69
Age Group × Condition	-0.554	0.313	-1.77	**3.04**	**1**	**0.08**
Age Group × Trial Type	-0.029	0.214	-0.14	0.02	1	0.89
Age Group × Plurality	0.343	0.215	1.60	2.45	1	0.12
Condition × Trial Type	0.568	0.214	2.66	**6.80**	**1**	**0.009**
Condition × Plurality	0.260	0.215	1.21	1.41	1	0.23
Trial Type × Plurality	0.339	0.215	1.58	2.40	1	0.12
Age Group × Condition × Trial Type	-0.259	0.429	-0.61	0.35	1	0.55
Age Group × Condition × Plurality	0.454	0.430	1.06	1.08	1	0.30
Age Group × Trial Type × Plurality	-0.232	0.430	-0.54	0.28	1	0.60
Condition × Trial Type × Plurality	0.530	0.429	1.24	1.47	1	0.23
Age Group × Condition × Trial Type × Plurality	-0.836	0.859	-0.97	0.91	1	0.34

Random Effects

	s^2
Participant (intercept)	0.354
Target (intercept)	0.023

This analysis revealed a significant main effect of trial-type, the crucial significant interaction of condition and trial-type, and a marginal interaction of age group and condition, but again no significant interactions involving trial-type, condition and age group, $\chi^2 < 1$, and no significant interactions involving target plurality. The simple main effect of trial type (informative vs. uninformative) was significant in the experimental condition for both children ($\chi^2(1) = 14.6$, $p = .0001$) and adults ($\chi^2(1) = 8.4$, $p = .004$), but not in the control condition for either age group (children: $\chi^2(1) = 0.3$, p = .60; adults: $\chi^2(1) = 0.9$, $p = .35$). This reveals sensitivity to the agreeing verb in this pre-noun window: Upon hearing an informative agreeing verb, children and adults were reliably more likely to shift their gaze from the distractor to the target before target-noun onset.

Because shift probability was calculated within an interval preceding the earliest target-noun onset in our stimuli, the informative-trial advantage observed in the experimental condition is evidence that both children and adults used the number-marked verb *predictively*, to reject a number-mismatching picture and shift to the target before its label was heard.

2.3. Discussion

In Experiment 1, listeners used an agreeing verb to predict the number features of its subject; this anticipatory effect was seen in shorter latencies to shift to the target in a 1500-ms measurement interval anchored on the determiner, and in a greater likelihood of shifting to the target within the pre-noun window. The effect of the agreeing verb disappeared in a control condition in which the target and distractor pictures always showed the same number of objects. This confirms the previous finding that young children used the combination of an agreeing verb and quantifier in online comprehension (Kouider et al., 2006), and establishes for the first time that 3-year-olds and adults can use an agreeing verb alone as a cue to pre-activate the number features of an upcoming noun.

These findings provide new evidence for the broad scope of predictive processing in young children's and adults' language comprehension. As noted earlier, prior investigations of young children's use of function words in online comprehension have focused on the close and consistently-ordered dependency between determiners and nouns. The present data show that the use of morphosyntactic cues to predict upcoming words is not limited to the case of determiners, but extends to the less consistent cue provided by an agreeing verb. This suggests that children's and adults' language comprehension is inherently and broadly predictive, and therefore that young children have access to the two advantages of anticipatory language processing with which we began. First, when children's predictions are correct, predictive use of function words demonstrably increases the speed and accuracy of sentence comprehension. Second, a tendency to predict may lead to a useful error signal when the predictions are false. New evidence of the broad scope of predictive processing therefore supports the plausibility of an error-based learning account.

Children's success in our task contrasts with the well-documented difficulty English-learning children have in using agreement as a direct cue to subject number in explicit judgment tasks (e.g., de Villiers & Johnson, 2007; Johnson et al., 2005; Miller, 2012). Brandt-Kobele and Höhle (2010) found a similar mismatch between implicit and explicit comprehension tasks in German-learning 3- and 4-year-olds. They used the homophony of the German singular feminine and plural pronouns to construct sentences in which only verb agreement disambiguated the number of the subject (*Sie füttert einen Hund*, 'She is feeding a dog' vs. *Sie füttern einen Hund*, 'They are feeding a dog'). Children used this cue to find the target in a looking-while-listening task, but their sensitivity to the verb-form vanished when they were asked to point to the target picture. These findings, and our data from younger children learning English, both suggest that young children readily use verb agreement in online comprehension *to serve its syntactic function*—that of implicitly predicting or checking the number of the subject—, even though they do not explicitly recognize verb agreement as a cue to number.

To explore the time-course of the informative-verb effect, we asked whether an agreeing verb also affected processing after noun information became available. Such an effect could reflect prediction (though too little or too late to effectively drive anticipatory looks to the target), but also integration processes. In trials in which participants were looking at the distractor at noun onset, we analyzed the latency of the first distractor-to-target shift in a 1500-ms window from 300 ms to 1800 ms (200 ms to 1700 ms for adults) after noun onset. This analysis, reported in Appendix C, showed a pattern somewhat similar to the latency measure anchored at the determiner, in that latencies tended to be shorter in informative trials in the experimental condition only (Figure C.2). However, the differences were small, and inferential statistics accordingly less robust, suggesting that the effect of the agreeing verb was mostly over by the time the noun began. In particular, the crucial interaction of condition and trial type was marginal ($\chi^2(1) = 3.51$, $p = .06$). The numerical resemblance between this and earlier measures, however, hints that the use of an informative agreeing verb can be extended in time (see Appendix C for discussion). Three-year-olds and adults can use an agreeing verb predictively, but may not always do so, or may do so too late or too weakly to pre-activate the target before its name is heard. In these cases, some of the verb's influence lingers after noun onset. We return to the time-course of our effects in Experiment 2.

Interestingly, the effect of agreeing verbs in our task appeared to be carried by the plural trials, especially for the 3-year-olds (although we did not find the significant 3-way interaction of trial-type, condition, and plurality that would support a claim of distinct response patterns in plural and singular trials)[6]. This numerical pattern might reflect inherent differences in the referential constraints of singular versus plural forms: As noted earlier, a picture of two cookies always shows one cookie that could be singled out. A similar point has been made regarding the inherent informativeness of novel transitive vs. intransitive verbs: A transitive verb (e.g., *She's kradding her*) requires a two-participant referent event, but an intransitive verb (*She's kradding*) could refer either to a one-participant event or to a component of a two-participant event (e.g., Yuan, Fisher & Snedeker, 2012). The apparent power of plurals in the present task is therefore not surprising, and suggests that we should expect a similar pattern with the younger children tested in Experiment 2.

3. Experiment 2

In Experiment 1, 3-year-olds and adults used an agreeing verb to facilitate the processing of a later noun and even to pre-activate its number features. This tells us that anticipatory use of the grammatical constraints of function words is not limited to the case of determiners, but extends to the more distant and less reliably marked dependency between subject and verb. We argued that this provides new evidence for the anticipatory nature of early comprehension and therefore for the plausibility of an error-based learning account of acquisition.

In Experiment 2, we extended this investigation to younger children, 2.5-year-olds. For error-based learning to guide syntax acquisition, children must generate predictions even before they know much about the morphosyntactic features of their native language; this prompts us to investigate children's predictive use of subject-verb agreement earlier in development. As noted

[6] Distractor-to-target shift latencies measured from noun onset were the only exception to this pattern. There, as shown in Figure C.2, the informative-trial advantage in the experimental condition appeared larger in singular rather than plural trials. The reversed singular-plural asymmetry in this late measure supports the impression that the early predictive effect of the verb was stronger (or faster) in plural trials: Accordingly, any remnant of the informative-trial advantage in this late measure was carried by singular trials.

earlier, 3-year-olds produce subject-verb agreement fairly reliably in their own speech, although they still omit tense and agreement markers in many contexts. 2.5-year-olds, in contrast, are much more likely to omit markers of tense and agreement, including the copula and auxiliary BE (e.g., Brown, 1973; Cazden, 1968; Rowland, Pine, Lieven & Theakston, 2005; Schütze, 2004; Theakston & Rowland, 2009; Wexler, 1994). Evidence that an agreeing verb nonetheless facilitates noun comprehension, or even permits pre-activation in our task, would suggest that children can make incremental use even of their less stable command of verb morphology.

In Experiment 2 we repeated the design of Experiment 1 with 2.5-year-olds, but anticipated that younger children might be slower to use an informative agreeing verb to make predictions. The speed of word identification increases with age (Swingley et al., 1999), as does efficiency in integrating each word into a representation of the sentence or discourse context (Fernald, Thorpe & Marchman, 2010; Song & Fisher, 2007). Therefore, in addition to the two main measures defined for Experiment 1, we also analyzed distractor-to-target shift latencies anchored at noun onset (where the older children and adults of Experiment 1 still showed remnants of an informativeness advantage). If the younger children use agreeing verbs to facilitate online noun identification (even if they sometimes fail to generate a correct prediction in time), they should need less information to identify the noun when it follows an informative agreeing verb, leading to shorter latencies. In addition, we analyzed the younger children's likelihood of fixating the target as opposed to the distractor at the onset of the target noun. This measurement point was chosen because it is the latest point at which we can seek evidence of predictive effects. At this point, children have received no information about the noun, but in informative trials they have already heard an agreeing verb. If children can use the agreeing verb predictively, then at noun onset they should be more likely to already be looking at the target in trials with an informative agreeing verb.

3.1. Method

3.1.1. Participants. Sixty-four 2.5-year-olds participated (28.1-32.3 months, *M* = 29.8; 28 girls). All were learning English as their first language. Half of the children were randomly assigned to the experimental condition and half to the control condition. Another 9 children were excluded because of a reported language delay (5), refusal to participate (1), or too many missing trials due to inattentiveness or parental interference (3; see Coding below). Children's productive vocabularies, measured using the short form of the MacArthur Bates CDI (Level III; Fenson et al. 2007), ranged from 0 – 92 (median = 55.5).

3.1.2. Stimuli, apparatus, and procedure. Experiment 2 was identical to Experiment 1 with three exceptions. First, the auditory stimuli were new recordings of the same sentences, spoken by the same speaker at a slightly slower rate, suitable for younger children. Second, to make room for the slightly longer stimulus sentences, the second question in each trial began 5.3 s (160 frames) after picture onset (e.g., *Do you like it/them?*). Third, Experiment 2 included seven rather than three single-picture filler trials among the 32 critical trials.

3.1.3. Coding. Coding was carried out as in Experiment 1. Reliability was calculated for 25% of the data. Coders agreed on 96% of all video frames. Individual trials were eliminated (174 of 2048 possible trials, 8.5%) if more than 50% of the trial was spent looking away or was uncodeable (123 trials), or if the child's or parent's speech obscured the critical sentence (51 trials). Three children's data were excluded because such eliminations left fewer than 4 of the 8 possible trials of each trial-type/plurality combination (informative singular, informative plural, uninformative singular, uninformative plural).

3.1.4. Measures. As in Experiment 1, we examined (a) the *latency* of the first distractor-to-target shift after the verb, measured from determiner onset, and (b) the *probability of distractor-to-target shifts* in the pre-noun window. We also analyzed (c) *the probability of target fixations* at noun onset, and (d) the *latency* of the first distractor-to-target shift after noun onset.

Windows for these measures were determined using the same criteria as in Experiment 1. In Experiment 2, the average length of the verb was 219 ms (range 135-295 ms, *is* M = 261 ms, *are* M = 177 ms). Accordingly, the latency of the first distractor-to-target shift after the verb was calculated in a 1500-ms window extending from 67 ms (2 frames) after determiner onset through 1567 ms after determiner onset; this captured latencies about 300 to 1800 ms after verb onset. Children were looking at the distractor at determiner onset in 894 trials (48% of 1874 included trials); 748 of these (84%) trials contained a direct distractor-to-target shift within the 1500-ms measurement interval. If participants used the number-marked verb to guide identification of the target noun, latencies should be shorter in informative trials in the experimental condition only.

The probability of a distractor-to-target shift before target-noun onset was measured in an 867-ms pre-noun window extending from 67 ms (2 frames) after determiner onset through 300 ms after the earliest target-noun onset. This window captures shifts beginning about 300 ms after average verb onset and excludes shifts that could be influenced by the target noun; this longer pre-noun window (867 ms vs. 633 ms in Experiment 1) reflects the slower speech in the recordings for Experiment 2.

Latency of the first distractor-to-target shift after noun onset was calculated in a 1500-ms window extending from 300 ms to 1800 ms after the onset of each target noun. Noun onsets occurred an average of 728 ms after determiner onset (range 621-810 ms). Children were fixating the distractor at noun onset in 862 trials (46% of 1874 included trials); 701 of these (81%) contained a direct distractor-to-target shift within the 1500-ms measurement interval. If the presence of an informative number-marked verb facilitates noun processing and integration, then latencies should be shorter in informative trials in the experimental condition.

3.2. Results

Figure 5 shows looks to the target as a proportion of looks to either picture in informative and uninformative trials over time, separately for the experimental and control conditions. As in Experiment 1, children's fixations to the target picture increased earlier in informative than in uninformative trials, in the experimental condition only.

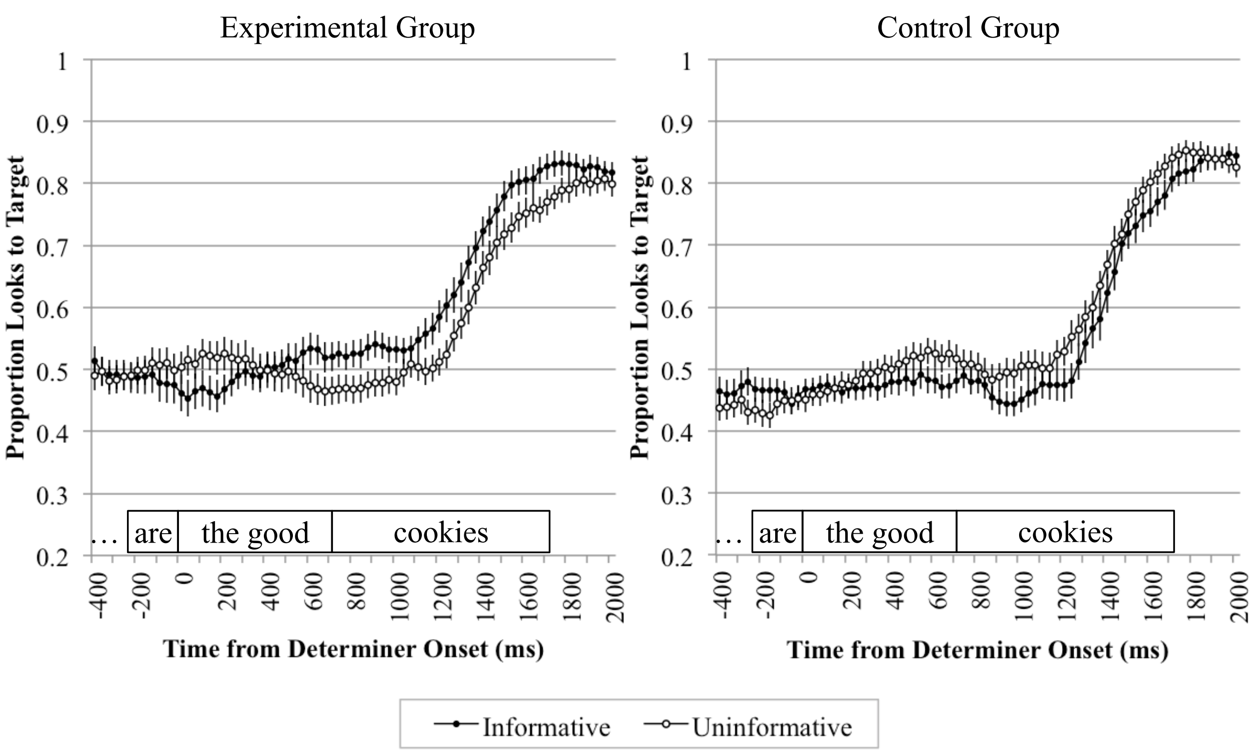

Figure 5. Mean (se) proportion fixations to the target in Experiment 2, plotted by trial type (informative vs. uninformative), and condition (experimental vs. control). The x-axis shows time measured from determiner onset; average verb onset in informative trials preceded determiner onset by approximately 233 ms. Text boxes indicate average onset and offset for each region of the sentence.

To preview our results, this pattern was numerically true of all four measures. However, we found robust effects of an informative agreeing verb on the two later measures anchored at the noun, not those anchored at the determiner, suggesting that these younger children used agreeing verbs predictively, but did so more slowly or less consistently than did the 3-year-olds in Experiment 1. Given an informative agreeing verb, the 2.5-year-olds were (a) only marginally faster to shift to the target in the 1500-ms latency window anchored at determiner onset and (b) not reliably more likely to shift from distractor to target in the pre-noun window. However, they were (c) reliably more likely to be fixating the target at noun onset, and (d) reliably faster to shift to the target in the 1500-ms latency window anchored at noun onset, in informative trials in the experimental as compared to the control condition.

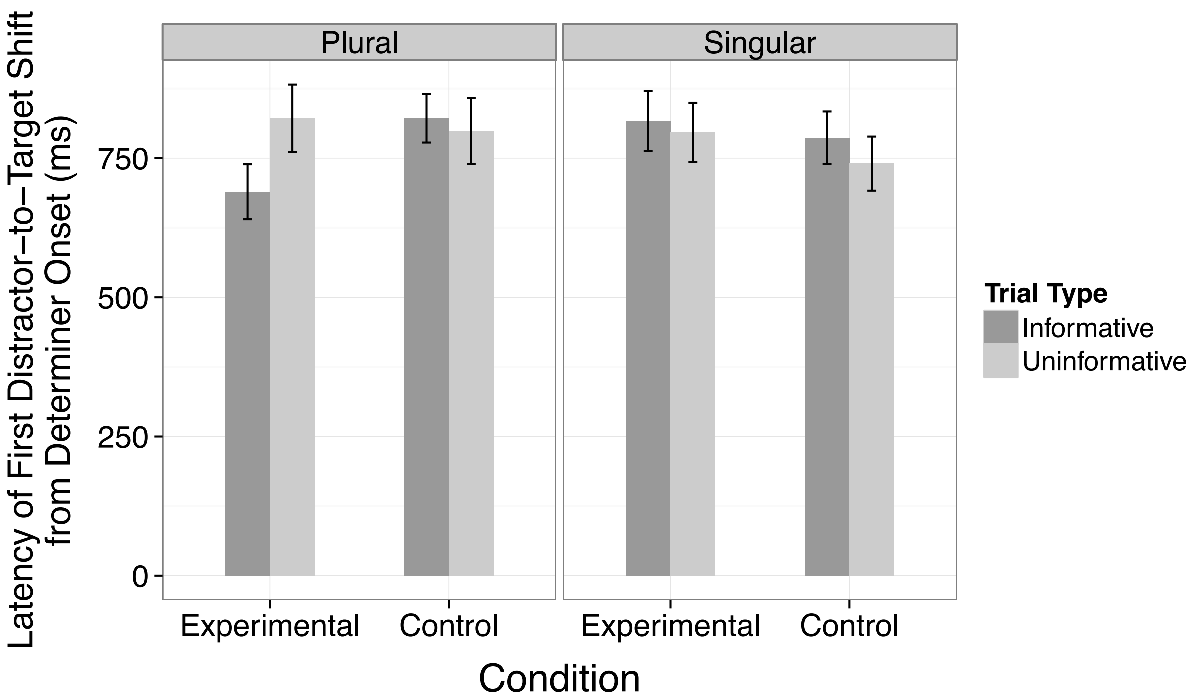

Figure 6. Mean (se) latency of first shift from distractor to target after the verb, measured from determiner onset, in Experiment 2 (30-month-olds), plotted by condition (experimental, control), trial type (informative, uninformative) and plurality (singular, plural).

Table 3. Estimates for mixed effects model of distractor-to-target shift latency measured from determiner onset, Experiment 2. Trial type and target plurality are within-participants factors and condition is a between-participants factor. Statistically significant and marginal effects are shown in bold. (N=748).

Fixed Effects

	Model Summary			Model Comparison		
	β	SE	t	χ^2	df	*p*
Intercept	769.1	24.5	31.39	**38.73**	**1**	**<.0001**
Condition	0.7	32.5	0.02	0.00	1	.99
Trial Type	-1.4	32.6	-0.04	0.00	1	.98
Plurality	-7.6	32.5	-0.24	0.05	1	.82
Condition × Trial Type	-122.7	65.1	-1.88	**3.57**	**1**	**.059**
Condition × Plurality	-126.9	65.0	-1.95	**3.81**	**1**	**.051**
Trial Type × Plurality	-112.6	65.2	-1.73	**3.02**	**1**	**.082**
Condition × Trial Type × Plurality	-2.0	130.2	-0.02	0.00	1	.98

Random Effects

	s^2
Participant (intercept)	0.000
Target (intercept)	2689.4

3.2.1. Latency of first distractor-to-target shift after the verb. As shown in Figure 6, distractor-to-target shift latencies measured from determiner onset were shorter in informative than in uninformative trials in the experimental condition; this pattern appeared only in plural trials. To test this pattern, we fit a mixed-effects model of latency as in Experiment 1. Predictor variables were the between-participants factor condition (experimental, control), and the within-participants factors trial-type (informative, uninformative) and target plurality (singular, plural).

Predictors were coded using mean-centered effects coding. Random intercepts were included for subjects and target nouns. No random slopes met the inclusion criterion (α = .2). Model results are shown in Table 3, and syntax for all models is in Appendix B. The crucial interaction of condition and trial type was marginal, as were the interactions of condition and plurality, and of trial type and plurality. The simple main effect of trial-type (informative vs. uninformative) was not significant in the experimental condition ($\chi^2(1) = 2.5$, $p = .11$) or the control condition ($\chi^2(1) = 2.3$, $p = .13$; note the numerical difference in the control condition is in the wrong direction).

In sum, the latency of the first distractor-to-target shift after the verb showed similar numerical patterns in Experiments 1 and 2, but the tendency for shifts to be faster in informative trials in the experimental condition only was marginal among the younger children.

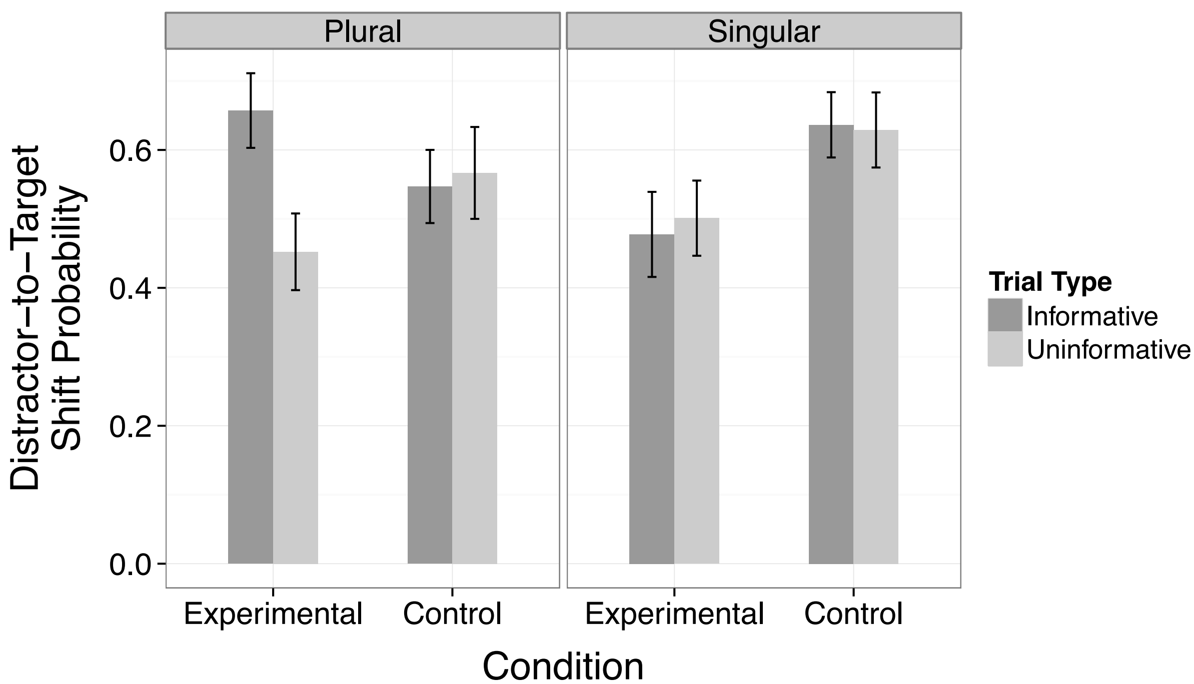

Figure 7. Mean (se) proportion of distractor initial trials with a shift to target in the pre-noun window in Experiment 2 (30-month-olds), plotted by condition (experimental, control) and trial type (informative, uninformative).

Table 4. Estimates for binomial mixed effects model of shift probability, in Experiment 2. Trial type and target plurality are within-participants factors, and condition is a between-participants factor. Statistically significant and marginal effects are shown in bold. (N=870)

Fixed Effects

	Model Summary			Model Comparison		
	β	SE	z	χ^2	df	*p*
Intercept	0.37	0.16	2.34	**4.46**	**1**	**.03**
Condition	-0.13	0.20	-0.67	0.44	1	.51
Trial Type	0.19	0.14	1.29	1.61	1	.20
Plurality	0.01	0.14	0.05	0.00	1	.96
Condition × Trial Type	0.44	0.29	1.52	2.22	1	.14
Condition × Plurality	0.65	0.29	2.25	**4.89**	**1**	**.03**
Trial Type × Plurality	0.45	0.29	1.57	2.38	1	.12
Condition × Trial Type × Plurality	0.82	0.57	1.43	1.97	1	.16

Random Effects

	s^2
Participant (intercept)	0.29
Target (intercept)	0.12

3.2.2. Shift Probability. Figure 7 shows the proportion of distractor-initial trials that included a shift to the target during the pre-noun window, by trial type, condition, and target plurality. As the Figure shows, children in the experimental but not the control condition made more distractor-to-target shifts in informative than uninformative trials in plural trials only. This pattern was tested using a binomial mixed effects model, with predictors and coding as described above. No random slopes met the inclusion criterion (α = .2). Model results are in Table 4. This analysis revealed a significant interaction of condition and plurality, but the key interaction of condition by trial type was not significant ($\chi^2(1) = 2.22$, $p = .14$). Follow up analyses conducted for comparison with Experiment 1 revealed that the simple effect of trial-type was significant in the experimental ($\chi^2(1) = 3.9$, $p = .048$), but not the control condition ($\chi^2(1) = .02$, $p = .9$).

Though the critical interaction between condition and trial-type was not significant for this early measure of anticipatory shifts, these analyses echoed the pattern found for 3-year-olds and adults in Experiment 1, including the observation that any effect of informative verb agreement was carried by the plural trials.

3.2.3. Probability of fixating the target at noun onset. As Figure 8 shows, at noun onset children in the experimental condition were already more likely to be looking at the target in informative than in uninformative trials. No such difference appeared in the control condition; in fact, the Figure shows a trend in the opposite direction.

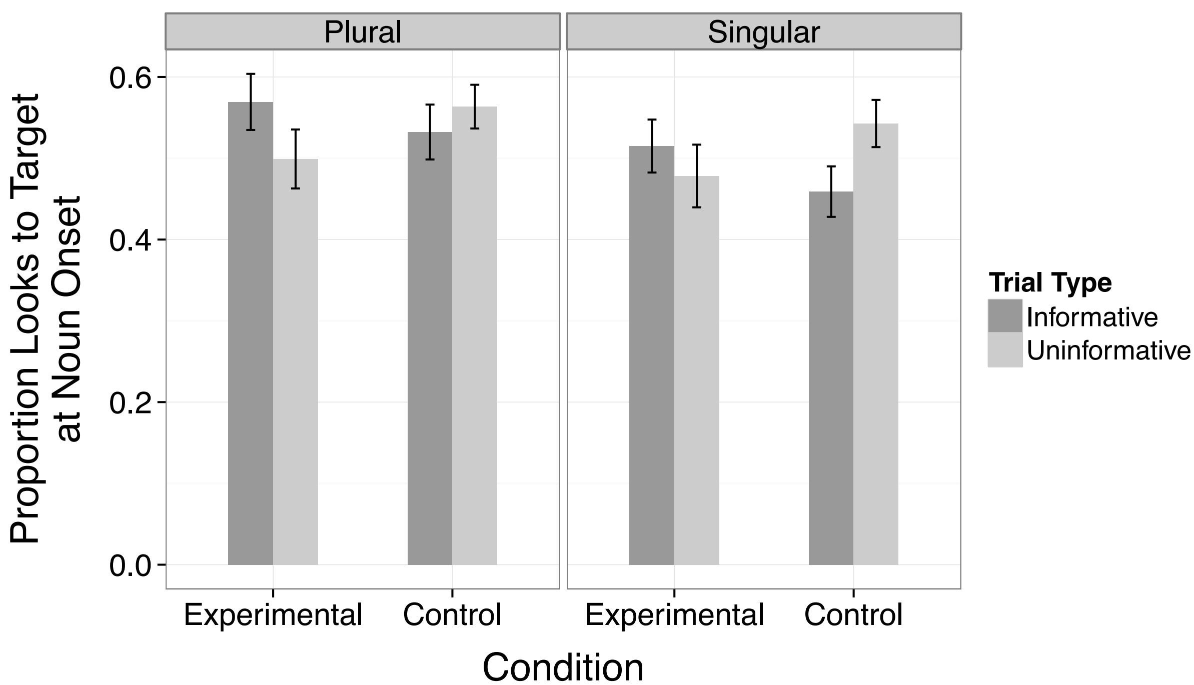

Figure 8. Mean (se) proportion looks to target at noun onset in Experiment 2 (30-month-olds), plotted by condition (experimental, control), trial type (informative, uninformative), and plurality (singular, plural).

Table 5. Estimates for binomial mixed effects model of fixations at noun onset, Experiment 2. Trial type and target plurality are within-participants factors, and condition is a between-participants factor. Statistically significant and marginal effects are shown in bold. (N=1800)

Fixed Effects

	Model Summary			Model Comparison		
	β	SE	z	χ^2	df	*p*
Intercept	0.08	0.10	0.82	0.64	1	.42
Condition	-0.02	0.10	-0.26	0.07	1	.79
Trial Type	-0.02	0.10	-0.23	0.05	1	.82
Plurality	-0.15	0.10	-1.54	2.35	1	.13
Condition × Trial Type	-0.44	0.19	-2.33	**5.43**	**1**	**.02**
Condition × Plurality	0.06	0.19	0.33	0.11	1	.74
Trial Type × Plurality	0.28	0.19	1.45	2.10	1	.15
Condition × Trial Type × Plurality	-0.00	0.38	-0.01	0	1	.99

Random Effects

	s^2
Participant (intercept)	0.00
Target (intercept)	0.07

This pattern was confirmed by a binomial mixed-effects model of looks to the target as opposed to the distractor picture at noun onset. As before, the predictors were condition, trial-type, and plurality. No random slopes met the inclusion criterion ($\alpha = .2$). Model results are in Table 5. This analysis revealed the critical significant interaction between trial-type and condition. Follow-up comparisons revealed that the simple main effect of trial-type (informative vs. uninformative) was marginally significant in the experimental condition ($\chi^2(1) = 3.3$, $p = .07$) and not significant in the control condition ($\chi^2(1) = 2.2$, $p = .14$). This pattern suggests that only the children in the experimental condition showed an informative-trial advantage.

This finding provides evidence of predictive processing in these young children. The presence of an agreeing verb, if it was informative in the referential context, directed children's gaze to the target picture before bottom-up information about the noun was available[7].

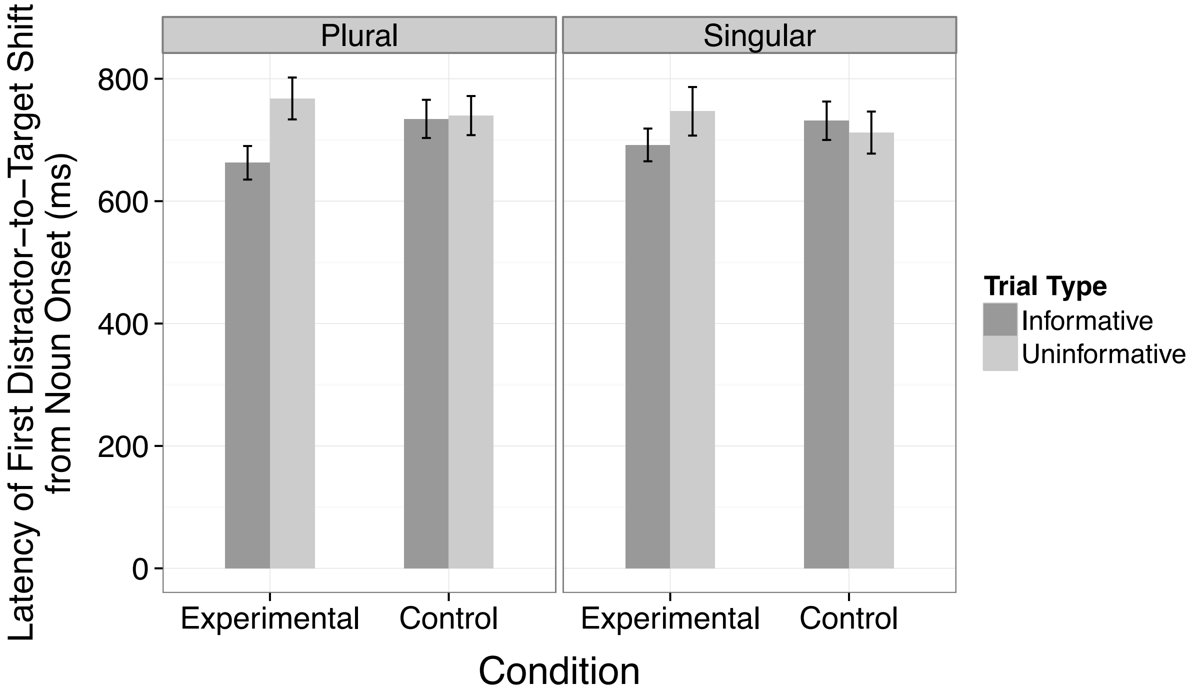

Figure 9. Mean (se) latency of first shift from distractor to target in Experiment 2 (30-month-olds) measured from noun onset, by condition (experimental vs. control), trial type (informative vs. uninformative) and plurality (singular, plural).

[7] An analysis of the probability of fixating the target at noun onset for the 3-year-olds and adults in Experiment 1 showed the same effects, including a significant interaction of condition by trial type. This analysis is reported in Appendix C.

Table 6. Estimates for mixed effects model of distractor-to-target shift latency measured from noun onset, Experiment 2. Trial type and target plurality are within-participants factors and condition is a between-participants factor. Statistically significant and marginal effects are shown in bold. (N=701).

Fixed Effects

	Model Summary			Model Comparison		
	β	SE	t	χ^2	df	*p*
Intercept	726.1	17.9	40.46	**48.70**	**1**	**<.0001**
Condition	-20.0	31.4	-0.64	0.40	1	.53
Trial Type	-32.7	27.9	-1.17	1.36	1	.24
Plurality	0.8	18.8	0.04	0.00	1	.95
Condition × Trial Type	-115.4	37.7	-3.06	**9.28**	**1**	**.002**
Condition × Plurality	-22.9	37.8	-0.61	0.37	1	.54
Trial Type × Plurality	-22.7	37.7	-0.60	0.36	1	.55
Condition × Trial Type × Plurality	-32.4	75.6	-0.43	0.18	1	.68

Random Effects

	s^2
Participant (intercept)	9892.7
Target (intercept)	601.4
Trial Type	3343.3

3.2.4. Latency of first distractor-to-target shift, measured from noun onset. As shown in Figure 9, upon hearing the noun, children were quicker to shift from the distractor to the target picture in informative than in uninformative trials, in the experimental condition only. This pattern was supported by a mixed-effects model of latency, with predictors and fitting procedures as described above. Only the random slope for trial-type by target noun met the inclusion criterion (α = .2). Model results are in Table 6. This analysis revealed the critical significant interaction between trial-type and condition, showing that the effect of an informative verb differed significantly across the experimental and control conditions. No other main effects or interactions were significant. The simple main effect of trial-type (informative vs. uninformative) was significant in the experimental condition ($\chi^2(1) = 6.2$, $p = .013$) but not the control condition ($\chi^2(1) = 0.6$, $p = .44$).

This result tells us that 2.5-year-olds used agreeing verbs to facilitate online noun comprehension: The children needed less information from the familiar noun itself if it was heralded by an agreeing verb that was informative in the referential context.

3.2.5. Correlations with vocabulary. The younger children in Experiment 2 were not at ceiling on the MCDI III (median = 55.5); thus we asked whether vocabulary scores predicted both speed of processing and the degree of predictive processing, as found in previous work (Lew-Williams & Fernald, 2007; Mani & Huettig, 2012). Collapsing across the experimental and control conditions, children with higher vocabularies had shorter shift latencies in uninformative trials, both for latency measured from the determiner ($r = -.40$, $p < .01$) and for latency measured from the noun ($r = -.33$, $p < .01$). Thus, consistent with prior work, higher-vocabulary children were quicker to identify target nouns in our task. However, inspection of the data revealed no clear evidence that the effect of the informative verb in the experimental condition depended on vocabulary for either latency measure. In fact, because lower-vocabulary children tended to have longer latencies overall, the mean difference between informative and uninformative trials on both latency measures was numerically larger among low-vocabulary children (determined by a median split). For example, in the experimental condition, the informative-trial advantage in distractor-to-target shift latency measured from the noun was 133 ms (SD = 154 ms) for low-vocabulary children, and 52 ms (SD = 150 ms) for high-vocabulary children.

3.3. Discussion

The results of Experiment 2 largely replicated the findings of Experiment 1 with younger children: 2.5-year-olds, at the early stages of including verbal agreement in their speech, used informative agreeing verbs to facilitate identification of a following subject noun phrase, and to pre-activate the number properties of the subject. An informative agreeing verb affected the latency with which children shifted their gaze to the target picture once they heard the target noun, but also the likelihood that they were already looking at the target at noun onset.

Although these patterns were similar for the 2.5- and 3-year-olds, aspects of our results suggest a timing difference between the age groups. Our two earliest measures—the latency of the first distractor-to-target shift after the verb, and the probability of a distractor-to-target shift in the pre-noun window—yielded statistically less robust effects for the 2.5-year-olds in Experiment 2. In contrast, our latest measure—the latency of the first distractor-to-target shift after the noun—yielded more robust effects for the younger children than for older children and adults. Because Experiments 1 and 2 used different recorded materials, it could be that this shift in the timing of our effect reflects unexpected differences in the clarity of the stimuli. However, it seems more likely that this timing difference is genuine and reflects well-known increases in the speed of language processing and in the command of verbal morphology across early childhood.

Our results therefore suggest both an early capacity for prediction that extends to the case of verb agreement, and developmental differences in the timing of correct predictions. These findings, in turn, have consequences for the contrasting advantages of prediction with which we began. First, when children make correct predictions, their predictions increase the speed and accuracy of sentence comprehension. To the extent that younger children are slower or less likely to generate correct predictions from the available linguistic cues, they will reap fewer of the benefits of prediction for fast and accurate comprehension. Accordingly, we should expect the balance of integration vs. (correct) prediction in online comprehension to change with development, as children's perception and use of linguistic evidence becomes faster and more reliable (e.g., Fernald, Perfors & Marchman, 2006; see Federmeier, 2007 for a similar argument regarding aging-related slowing in older adults' language processing). Second, when children's predictions are false, the tendency to predict could yield an error signal for learning. The consequences of developmental timing differences for error-based learning depend on what happens when the linguistic context or the listener's expertise do not permit strong predictions. Error-based learning accounts assume that language processing is inherently predictive, and that even when constraints are weak, learners make implicit predictions about the set of words that might come next (e.g., Chang et al., 2006). To illustrate, when children fail to use an agreeing verb to make a correct prediction in our task, they might use their knowledge of the context (i.e., *Where is/are the…* predicts a noun) to partially pre-activate the names of both pictures. Such errors would provide no benefit for comprehension during this trial, but would create opportunities to learn once the target picture is named; we return to this central question in the General Discussion.

In most of our measures we saw a numerical singular-plural asymmetry for children in both Experiments 1 and 2. The anticipatory effects of an agreeing verb persistently appeared strongest in plural trials. We attributed this numerical difference to the difference in referential informativeness of the singular and the plural: Where there are two cookies, there is also one. However, there are other potential explanations for this pattern. First, agreement can be neutralized in casual speech in many English dialects (including the one the authors speak), when ‘default’ singular agreement is followed by a plural form in inverted contexts like the ones in our stimulus materials (e.g., *There’s some cookies in the kitchen*; see Squires, 2014; Hay & Schreier, 2004). This pattern might render singular agreement less predictive of the form of the following noun than plural agreement (at least when the referential context makes the second person subject *you* unlikely). Similarly, psycholinguistic analyses of linguistic number often treat the plural as the marked case, and the singular as unmarked and interpreted as singular by default (e.g., Eberhard, 1997). Evidence for this difference in markedness comes from patterns of errors in adults’ production and comprehension (Bock & Miller, 1991; Pearlmutter, Garnsey & Bock, 1999; Wagers, Lau & Phillips, 2009). Future studies using tasks with different referential constraints or manipulating the plausibility of agreement neutralization will help to clarify whether this is an influence of dialect experience on children’s comprehension (e.g., Miller, 2012), an effect of markedness, or simply a difference in referential informativeness.

4. General Discussion

In two experiments, adults and children used the information carried by an agreeing verb to guide incremental sentence processing, and even to pre-activate features of an upcoming noun. The presence of a number-marked verb directed the listeners' gaze away from a number-mismatching distractor, and toward a suitable target picture. This was seen in shorter latencies to shift gaze from distractor to target after the verb, a higher probability of shifting from distractor to target before the onset of the noun, and a higher probability of fixating the target at noun onset; this last measure showed clear evidence of prediction based on an agreeing verb even among 2.5-year-olds. We also found evidence that the agreeing verb facilitated processing after noun information was available in the 2.5-year-olds, as shown by shorter latencies to shift from distractor to target after the noun. These findings add to the evidence for adults' and toddlers' anticipatory use of morphosyntactic cues in language comprehension in two important ways.

First, these data show that the incremental use of function words extends beyond the determiner-noun dependency. We argued in the Introduction that English subject-verb agreement, a famously minimal system, provides a sterner test of the child's predictive use of function words than the syntactically intimate and directionally consistent dependency between determiner and noun. An agreeing verb-form provides a pointer to the grammatical number of its subject, which can precede or follow it. The limited value of an agreeing verb as a direct cue to notional number can be seen in the difficulty English-speaking children have gleaning number meaning from verb agreement alone in judgment tasks, though they readily interpret number morphology within the noun phrase (e.g., Johnson et al., 2005; Miller, 2012; cf. Kouider et al., 2006; Nicolaci-da-Costa & Harris, 1983). Even so, the toddlers and adults in our experiments used agreeing verbs in inverted sentences to anticipate the number features of a noun.

Second, our experiments yielded new evidence for a strong form of prediction, the pre-activation of features of the upcoming target noun before that noun began. These data are the first evidence that children and adults can make such predictions based on an agreeing verb alone. Moreover, we noted in the Introduction that evidence for pre-activation based on function-word cues has been mixed. To our knowledge, only three previous reports yielded evidence that either adults or children could use function words alone to pre-activate features of upcoming nouns: Tsang and Chambers (2011) documented adults' predictive use of Cantonese classifiers, Melançon and Shi (in press) showed toddlers' predictive use of French gender-marked determiners, and Kouider et al. (2006) found that 2-year-olds could glean number meaning from the combination of an agreeing verb and a quantifier (*is a* vs. *are some*). We add to these positive findings and extend the evidence for prediction to the demanding case of subject-verb agreement.

Given the well-documented limitations of agreement production before age 3 or 4, it is striking that even the 2.5-year-olds in Experiment 2 used agreeing verbs predictively, albeit more slowly than 3-year-olds. As noted earlier, English-learning 2- to 3-year-olds often leave out markers of tense and agreement in their own speech, and they are particularly vulnerable to error with plural forms and inverted contexts (Rowland et al., 2005; Rubino & Pine, 1998; Theakston & Rowland, 2009). Our evidence that children just 2.5 years old can use an agreeing verb to anticipate the number of its subject is an impressive case of a familiar phenomenon: Children understand more than they say, and do so well enough to predict an upcoming noun.

These data provide new support for the proposal that language processing by toddlers and adults is inherently and broadly predictive. Predictions, when correct, facilitate quick, accurate language comprehension. Moreover, the ability to predict is a key requirement of error-based learning accounts. In order to learn language by learning to predict the future, one must be in the business of predicting the future. Our evidence tells us that children are, and that their ability to predict is not limited to the simplest cases.

4.1. The role of prediction in syntax learning

How might prediction guide syntax acquisition? Here we briefly sketch how an error-based learning procedure might work, drawing on theoretical proposals by Chang et al. (2006) and Jaeger and Snider (2013; see also Dell & Chang, 2014; DeLong, Troyer & Kutas, 2014). This sketch is intended to set the stage for future work and thus goes far beyond what we can conclude based on the present results.

Error-based learning accounts propose that the language processing system routinely predicts upcoming words and their features based on current knowledge. Given the infinite combinatory power of language (after all, we can speak of *cabbages and kings*), upcoming words are rarely individually predictable. For this reason, error-based learning accounts assume that predictions are distributed over large categories of words that are, in the learner's experience, compatible with the unfolding structure. Mismatches between predicted categories of words and observed outcomes yield error-signals that gradually refine the system, enabling it to make more accurate predictions in the future.

Error-based learning thus proposes a mechanism for distributional learning. Across many opportunities to predict, the system will implicitly identify categories of words based on the predictive dependencies in which they participate (e.g., Saffran, 2001; Reeder, Newport & Aslin, 2013). Over time, such a learner would come to use a determiner to predict a broad category of nouns (*the …*), and an agreeing verb to predict properties of its subject (*Where are the …*), because predictions of incongruent words lead to an error signal that changes the system.

Given this proposal, what is learned depends on what cues children use to make predictions, and what properties children use these cues to predict. Regarding sources of prediction, the current study provided new evidence for prediction from function words, as opposed to from the rich constraints of content words. Regarding targets of prediction, evidence mostly from the adult psycholinguistic literature suggests that the language processing system makes predictions about many properties of upcoming words, including semantic (e.g., Altmann & Kamide, 1999; see also, e.g., Mani & Huettig, 2012, for evidence from children), phonological (e.g., DeLong, Urbach & Kutas, 2005; Dikker, Rabagliati, & Pylkkänen, 2009), and grammatical features (e.g., Lau, Stroud, Plesch, & Phillips, 2006; van Berkum, Brown, Zwitserlood, Koojiman, & Hagoort, 2005; Wicha, Bates, Moreno, & Kutas, 2003). If we speculate that learners make similarly diverse predictions, then error-based learning would lead to multi-faceted knowledge about grammatical categories and dependencies. For instance, if the learner uses *the* to predict both a class of known nouns and semantic features typical of the noun category, then this would permit syntax-guided word learning, allowing the learner to infer that a new noun names an object (e.g., "Can you give me the blicket?"; Waxman & Booth, 2001).

This sketch is consistent with the proposal that early-learned function words can be used to *label the syntactic type* of the constituents in which they occur, such as noun or verb phrases (Bernal et al., 2010; Christophe, Millotte, Bernal, & Lidz, 2008; Morgan, Shi, & Allopenna, 1996; Shi, 2014). Our findings suggest an interesting extension of this proposal—that toddlers' knowledge of function words might help them *identify links between constituents* via the predictive dependency between their morphosyntactic features. Just as syntactic bootstrapping proposes that children use regularities in the number and type of arguments to identify verbs and infer their meanings (Fisher, Gertner, Scott, & Yuan, 2010; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Trueswell, Papafragou, & Choi, 2011), children might use the predictive dependency between an agreeing verb and properties of one of its arguments to identify diverse subject noun phrases as bearing the same syntactic relationship to the verb, regardless of variation in their sentence positions or their semantic roles.

Error-based learning requires children to predict the future, but also to learn from their mistakes. We sought evidence for prediction by putting children in a position to make correct predictions. But error-based learning predicts that we learn most when our predictions are wrong (e.g., Dell & Chang, 2014). Recent work on syntactic priming provides strong evidence for this pattern. Syntactic priming is the tendency for adults and children to re-use in production, or to anticipate in comprehension, recently experienced syntactic structures (e.g., Bencini & Valian, 2008; Bock, 1986; Thothathiri & Snedeker, 2008). Both preschoolers and adults show larger priming effects when the primed structure is more unexpected (Bernolet & Hartsuiker, 2010;

Fine & Jaeger, 2013; Jaeger & Snider, 2013; Peter, Chang, Pine, Blything, & Rowland, 2015). In a simple artificial grammar-learning task, even 9-month-olds learn most when patterns are unexpected (Gerken, Dawson, Chatila, & Tenenbaum, 2015). Such effects of (un)predictability on learning are consistent with the predictions of an error-based learning account.

This sketch highlights two key directions for future investigations. By exploring the sources and targets of prediction in early language processing (as in the present study), and by assessing how learning varies with what children might have predicted, we can probe the role of prediction in syntax acquisition.

4.2. Mechanisms of prediction

What features might listeners have predicted in our experiments, based on the form of an agreeing verb? Several possibilities are suggested by the nature of linguistic number-marking, and these possibilities have consequences for what could be learned via error-based learning.

We noted earlier that number could be computed in notional or grammatical terms. Children might use *are* to predict that the subject will refer to more than one thing, that it will be grammatically plural, or both. This contrast touches on long-standing questions about whether children represent early linguistic categories and dependencies in terms of meaning or form (e.g., Bowerman, 1973; Valian, 1986). English verb agreement best reflects the grammatical number of the subject noun phrase, not its notional number. But most object names, and therefore many early-acquired nouns (Samuelson & Smith, 1999), are count nouns, whose grammatical number varies in lockstep with notional number (e.g., *one cookie, many cookies*). Because our materials used familiar count nouns, children could succeed in our task by predicting either notional or grammatical number.

The present data do not permit us to choose among these possibilities. However, emerging evidence from other experiments in our lab suggests that a grammatical number contrast is sufficient to permit predictive use of use an agreeing verb (Lukyanenko & Fisher, 2012, 2013). In a task like the one used in the current experiments, 3-year-olds and adults heard sentences with an informative agreeing verb, but the target nouns included object mass nouns such as *toast* and *corn,* and *pluralia tantum* nouns (nouns that are always grammatically plural) such as *glasses* and *pants*. Combining these with count nouns yielded trials in which both pictures in view showed the same number of objects, but grammatical number was still informative. For example, children heard *Where are the pretty glasses?* while viewing one pair of eyeglasses and one phone. Interpreting *are* purely notionally, as a cue to seek a picture of two objects, would not lead to success in this context. Nonetheless, both 3-year-olds and adults used agreeing verb-forms to predict nouns: Upon hearing *Where are the…*, they rejected a single phone as a suitable referent, and shifted their eyes to the pair of glasses. This suggests that children and adults can derive *grammatical* predictions from an agreeing verb, without the support of a notional number contrast. This does not tell us that children or adults predicted *only* grammatical number, however. Experiments are underway to examine the interaction between grammatical and notional plurality, asking whether predictions are stronger when notional and grammatical number coincide than when they conflict. These investigations will help clarify the nature of the predictions listeners make on the basis of an agreeing verb.

4.3. The scope of prediction

For error-based learning to play a key role in early syntax acquisition, prediction must be widespread. We argued that young children's predictive use of agreeing verbs provides new evidence that it is. However, one might challenge this claim: We observed prediction on the basis of an agreeing verb, but in a task in which we increased the measurement window by interposing an adjective and decreased the referential options to the two pictures in view. This task showed a capacity for prediction, but in a constrained context. Most experimental evidence for prediction in language processing has this limitation, as researchers test whether a particular licensed prediction can be made (e.g., Federmeier, 2007; Kutas et al., 2011). One might reasonably ask how widespread such predictions would be in ordinary conversation.

Can children or adults generate the broad, category-wide predictions that are assumed on an error-based learning account? This is a matter of current controversy in the psycholinguistics literature, but evidence suggests that at least in adults, prediction is not limited to highly constrained contexts. For example, Szewczyk and Schriefers (2013) measured event-related brain potentials (ERPs) as Polish speakers read sentences allowing them to predict the animacy but not the identity of a sentence-final noun. They found a more negative N400 response (an ERP component reflecting lexical-semantic processing) to prenominal adjectives with morphological markings that conflicted with the predicted semantic feature. This suggests that the Polish speakers pre-activated the animacy of an upcoming noun—and its consequences for the morphology of a preceding adjective—in an otherwise unconstraining linguistic context.

Some of the strongest evidence for grammatical category-wide predictions comes from studies showing very early brain responses to syntactic category violations. For example, Dikker et al. (2009) presented adults with grammatically predictive sentence preambles presented word-by-word on a screen. Readers expect a noun to follow *The tasteless…*, but after *The tastelessly…* they expect a participle (as in *tastelessly marketed soda*). When a noun appeared instead of a participle (**The tastelessly soda*), the violation affected brain responses occurring only 100-130 ms later, measured via MEG (magnetoencephalography). These responses, which Dikker et al. argued were too early to reflect identification of the target word, depended on orthographic properties of the target nouns: Early effects were found for nouns that were orthographically typical of the noun category, by virtue of their affixes (e.g., *artist*, *princess*) or their sound patterns (e.g., *soda*; see Kelly, 1992, for evidence that grammatical categories tend to share typical sound properties), but not for nouns whose orthographic patterns were equally noun- or verb-like. Thus, at least in adults, prediction permits the pre-activation of broad classes of words, and includes information about the properties of the pre-activated categories—in this case their typical orthographic or morphological properties (see also DeLong et al., 2005; Lau et al., 2006). This is the kind of pre-activation assumed by error-based learning accounts: a broad swath of prediction, pre-activating the many words that might appropriately come next and their properties, each to a tiny degree.

Finally, although we assume a constrained task such as ours is helpful in measuring particular predictions, this does not imply that predictions are generated too slowly or weakly to be useful under ordinary circumstances. Agreement licenses abstract predictions; such diffuse category-level activation may need to accumulate over time in a relatively constrained task to be detected in our eye-gaze measures, but might nonetheless often be available for implicit comparison to the linguistic input (i.e., might often affect the probability distribution of predicted words). As outlined above, even diffuse predictions might have substantial effects given many opportunities to generate them and learn from their success or failure. In English, information about subject-verb agreement is available on every finite verb; other morphosyntactic dependencies are similarly ubiquitous (e.g., gender- or number-marked determiners). Even if children generate strong predictions (in time) in only a small proportion of these instances, they will still have many opportunities to learn. Error-based learning relies on the gradual accumulation of knowledge over many tiny predictions and adjustments in response to error (e.g., Chang et al., 2006).

4.4. Conclusions

In the present experiments, toddlers and adults used the information carried by an agreeing verb to anticipate the number features of an upcoming noun. Upon hearing *Where are the…,* they rejected a single object as a suitable referent, and therefore often shifted to the target picture before hearing the target noun. The predictive use of an agreeing verb could be seen in children as young as 2.5 years of age, who often omit function words in their own speech. This provides strong new evidence that even for toddlers, language comprehension is fundamentally predictive. Children's predictions may play a key role in early acquisition, both by making possible rapid language comprehension when predictions are correct, and by driving learning when they turn out to be wrong.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*, 247-264.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*(4), 502-518.

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

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition, 122,* 292-305.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience. 7*(1)*,* 66-80.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278.

Bates, D., Maechler, M., & Bolker, B. (2013). lme4.0: Linear mixed-effects models using s4 classes. Retrieved from http://lme4.r-forge.r-project.org/ (R package version 0.999999-3)

Bencini, G., & Valian, V. (2008). Abstract sentence representations in 3-year-olds: Evidence from language production and comprehension. *Journal of Memory and Language, 59,* 97-113.

Bernal, S., Dehaene-Lambertz, G., Millotte, S., & Christophe, A. (2010). Two-year-olds compute syntactic structure online. *Developmental Science, 13*(1), 69-76.

Bernal, S., Lidz, J., Millotte, S., & Christophe, A. (2007). Syntax constrains the acquisition of verb meaning. *Language Learning and Development, 3*(4), 325-341.

Bernolet, S., & Hartsuiker, R. (2010). Does verb bias modulate syntactic priming? *Cognition, 114*(3), 455-461.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology 18*, 355-387.

Bock, K., & Middleton, E. L. (2011). Reaching Agreement. *Natural Language and Linguistic Theory*, *29*, 1033-1069.

Bock , K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23*, 45-93.

Borovsky, A., & Creel, S. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology, 50*(5), 1600-1613.

Borovsky, A., Elman, J., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112*, 417-436.

Borovsky, A., Sweeney, K., Elman, J. L., & Fernald, A. (2014). Real-time interpretation of novel events across childhood. *Journal of Memory and Language, 73*, 1-14.

Bowerman, M. (1973). Structural relationships in children's utterances: Syntactic or semantic? In T. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press.

Brandt-Kobele, O.-C., & Höhle, B. (2010). What asymmetries within comprehension reveal about asymmetries between comprehension and production: The case of verb inflection in language acquisition. *Lingua, 120,* 1910-1925.

Broen, P. A. (1972). The verbal environment of the language-learning child. *Monograph of the American Speech and Hearing Association, 17*, 1-103.

Brown, R. (1973). *A first language: The early stages.* Harvard University Press.

Cazden, C. B. (1968). The acquisition of noun and verb inflections. *Child Development*, *39,* 433-448.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234-272.

Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and Speech, 51*(1-2), 61-75.

Cyr, M., & Shi, R. (2013). Development of abstract grammatical categorization in infants. *Child Development, 84*(2), 617-629.

Dahan, D., Swingley, D., Tanenhaus, M. K. & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition. *Journal of Memory and Language*, *42*, 465-480.

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634).

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass, 8*(12), 631-645.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121.

de Villiers, J. G., & de Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research, 2*(3), 267-278.

de Villiers, J. G., & Johnson, V. E. (2007). The information in third-person /s/: Acquisition across dialects of American English. *Journal of Child Language, 34*(1), 133-158.

Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition, 110,* 293-321.

Eberhard, K. (1997). The marked effect of number on subject-verb agreement. *Journal of Memory and Language*, *36*, 147-164.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology. 44,* 491-505.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User guide and technical manual* (2nd Ed.). Baltimore: Brookes.

Fernald, A., Perfors, A., & Marchman, V.A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42*, 98-116.

Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective-noun phrases. *Cognitive Psychology, 60*(3), 190-217.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In *Developmental Psycholinguistics: Online methods in children's language processing.* I.A. Sekerina, E.M. Fernandez & H. Clahsen, Eds. Amsterdam, Netherlands: John Benjamins Publishing Company.

Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science, 37*, 578-591.

Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science, 1,* 143-149.

Friedrich, M., & Friederici, A. D. (2005). Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *Neuroreport, 16*(6), 653-656.

Gerken, L., Dawson, C., Chatila, R., & Tenenbaum, J. (2015). Surprise! Infants consider possible bases of generalization for a single input example. *Developmental Science*, *18*(1), 80-89.

Gerken, L., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology, 26*(2), 204-216.

Gerken, L., & McIntosh, B. J. (1993). Interplay of function morphemes and prosody in early language. *Developmental Psychology, 29*(3), 488-457.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, *1*(1), 23-64.

Hay, J., & Schreier, D. (2004). Reversing the trajectory of language change: Subject-verb agreement with *be* in New Zealand English. *Language Variation and Change. 16*, 209-235.

Huang, Y. T., Zheng, X., Meng, X., & Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language, 69*(4), 589-606.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition, 127*, 57-83.

Johnson, V. E., de Villiers, J. G., & Seymour, H. N. (2005). Agreement without understanding?:

The case of the third person singular /s/. *First Language. 25*(3), 317-330.

Jolly, H. R., & Plunkett, K. (2008). Inflectional bootstrapping in 2-year-olds. *Language and Speech, 51*(1-2), 45-59.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass, 2*(4), 647-670.

Kedar, Y., Casasola, M., & Lust, B. (2006). Getting there faster: 18- and 24-month-old infants' use of function words to determine reference. *Child Development, 77*(2), 325-338.

Keeney, T. J., & Wolfe, J. (1972). The acquisition of agreement in English. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 698-705.

Kelly, M. H. (1992). Using sound to solve syntactic problems: the role of phonology in grammatical category assignments. *Psychological Review*, *99*(2), 349-364.

Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science, 14*(4), 925-934.

Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular-plural distinction. *Language Learning and Development. 2*(1), 1-25.

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time-course of anticipatory constraint integration. *Cognition, 119*, 23-42.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190-207). Oxford University Press.

Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language, 98*, 74-88.

Legendre, G., Barrière, I., Goyet, L., & Nazzi, T. (2010). Comprehension of infrequent subject-verb agreement forms: Evidence from French-learning children. *Child Development. 81*(6), 1859-1875.

Legendre, G., Culbertson, J., Zaroukian, E., Hsin, L., Barrière, I., & Nazzi T. (2014). Is children's comprehension of subject-verb agreement universally late? Comparative evidence from French, English, and Spanish. *Lingua, 144*, 21-39.

Leonard, L. B., Eyer, J. A., Bedore, L. M., & Grela, B. G. (1997). Three accounts of the grammatical morpheme difficulties of English-speaking children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 40*, 741-753.

Leonard, L. B., Miller, C. A., & Owen, A. J. (2000). The comprehension of verb agreement morphology by English-speaking children with Specific Language Impairment. *Clinical Linguistics & Phonetics, 14*(6), 465-481.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science. 18*(3), 193-198.

Lukyanenko, C., & Fisher, C. (2012). Grammatical or notional number? 3-year-olds' production and comprehension of verb agreement. Poster presented at the 25th annual CUNY Conference on Human Sentence Processing, New York, NY.

Lukyanenko, C., & Fisher, C. (2013) Here comes the subject! Listeners use number-marked verbs to predict subject number. Poster presented at the 26th annual CUNY Conference on Human Sentence Processing, Columbia, SC.

MacWhinney, B. (2000). The CHILDES Project: Tools for analyzing talk. Mahwah, NJ: Lawrence Erlbaum Associates.

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake – but only for skilled producers. *Journal of Experimental Psychology: Human Perception and*

Performance, 38(4), 843-847.

Maratsos, M. P., & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. E. Nelson (Ed.), *Children's Language* (Vol. 2). New York: Gardner.

Mariscal, S. (2009). Early acquisition of gender agreement in the Spanish noun phrase: Starting small. *Journal of Child Language, 36*(1), 143-171.

Melançon, A., & Shi, R. (in press). Representations of abstract grammatical feature agreement in young children. *Journal of Child Language*.

Miller, K. (2012). Not all children agree: Acquisition of agreement when the input is variable. *Language Learning and Development, 8*(3), 255-277.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science, 2*, 138-153.

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In J. L. Morgan and K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, (pp. 263-283). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nazzi, T., Barrière, I., Goyet, L., Kresh, S., & Legendre, G. (2011). Tracking irregular morphophonological dependencies in natural language: Evidence from the acquisition of subject-verb agreement in French. *Cognition*, *120*, 119-135.

Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow and C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition,* (pp. 109-150). Cambridge: Cambridge University Press.

Nicolaci-da-Costa, A., & Harris, M. (1983). Redundancy of syntactic information: An aid to young children's comprehension of sentential number. *British Journal of Psychology, 74*(3), 343-352.

Nicolaci-da-Costa, A., & Harris, M. (1984). Young children's comprehension of number markers. *British Journal of Developmental Psychology, 2*(2), 105-111.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language. 41*, 427-456.

Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language*, *81*, 1-15.

R Core Team. (2013) R: a language and environment for statistical computing (Version 3.0.2) [Software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology*, *66*(1), 30-54.

Rispoli, M., Hadley, P. A., & Holt, J. K. (2009). The growth of tense productivity. *Journal of Speech, Language, and Hearing Research, 52*, 930-944.

Rissman, L., Legendre, G., & Landau, B. (2013). Abstract morphosyntax in two- and three-year-old children: Evidence from priming. *Language Learning and Development, 9*(3), 278-292.

Roberston, E. K., Shi, R., & Melançon, A. (2012). Toddlers use the number feature in determiners during online noun comprehension. *Child Development*, *83*(6), 2007-2018.

Rowland, C. F., Pine, J. M., Lieven, E. V. M., & Theakston, A. L. (2005). The incidence of error in young children's wh-questions. *Journal of Speech, Language, and Hearing Research, 48*, 384-404.

Rubino, R. B., & Pine, J. M. (1998). Subject-verb agreement in Brazilian Portuguese: What low error rates hide. *Journal of Child Language, 25*(1), 35-59.

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition, 81*(2), 149-169.

Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition, 73*, 1-33.

Schütze, C. T. (2004). Why nonfinite *be* is not omitted when finite *be* is. In A. Brugos, L. Micciulla & C. E. Smith (Eds.), *Proceedings of the 28th annual Boston University Conference on Language Development*, (pp. 506-521). Somerville, MA: Cascadilla Press.

Shi, R. (2014). Functional morphemes and early language acquisition. *Child Development Perspectives*, 8(1), 6-11.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*(3), 302-319.

Smouse, M., Gxilishe, S., de Villiers, J. G., & de Villiers, P. A. (2012). Children's acquisition of subject markers in isiXhosa. In M. P. Larrañaga & P. Guijarro-Fuentes (Eds.), *Pronouns and clitics in early acquisition* (pp. 209-236). Berlin/New York: Mouton de Gruyter.

Soderstrom, M. (2002). *The acquisition of inflection morphology in early perceptual knowledge of syntax.* Unpublished doctoral dissertation, Johns Hopkins University.

Soderstrom, M., Wexler, K., & Jusczyk, P. W. (2002). English-learning toddlers' sensitivity to agreement morphology in receptive grammar. In B. Skarabela, S. Fish, & A. H.-J. Do (Eds.), *Proceedings of the 26th annual Boston University Conference on Language Development* (Vol. 2), pp. 643-652. Somerville, MA: Cascadilla Press.

Soderstrom, M., White, K. S., Conwell, E., & Morgan, J. L. (2007). Receptive grammatical knowledge of familiar content words and inflection in 16-month-olds. *Infancy, 12*(1), 1-29.

Song, H., & Fisher, C. (2007). Discourse prominence effects on 2.5-year-old children's interpretation of pronouns. *Lingua, 117*, 1959-1987.

Squires, L. (2014). Talker specificity and the perception of grammatical variation. *Language, Cognition and Neuroscience, 29*(7), 856-876.

Swingley, D., & Fernald, A. (2002). Recognition of words referring to present and absent objects by 24-month-olds. *Journal of Memory and Language. 46,* 39-56.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*. *71*, 73-108.

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language, 68*, 297-314.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2005). The acquisition of auxiliary syntax: BE and HAVE. *Cognitive Linguistics, 16*(1), 247-277.

Theakston, A. L., & Rowland, C. F. (2009). The acquisition of auxiliary syntax: A longitudinal elicitation study. Part 1: Auxiliary BE. *Journal of Speech Language and Hearing Research, 52,* 1449-1470.

Thorpe, K., & Fernald, A. (2006). Knowing what a novel word is not: Two-year-olds 'listen through' ambiguous adjectives in fluent speech. *Cognition. 100*(3), 389-433.

Thothathiri, M., & Snedeker, J. (2008). Syntactic priming during language comprehension in three-and four-year-old children. *Journal of Memory and Language*, *58*(2), 188-213.

Trueswell, J. C., Papafragou, A., & Choi, Y. (2011). Syntactic and referential processes: What develops? In E. Gibson & N. Perlmutter (Eds.), *The processing and acquisition of reference*. Cambridge, MA: MIT Press.

Tsang, C., & Chambers, C. G. (2011). Appearances aren't everything: Shape classifiers and referential processing in Cantonese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1065-1080.

Valian, V. (1986). Syntactic categories in the speech of young children. *Developmental Psychology, 22*(4), 562- 579.

van Berkum, J. J. (2008). Understanding sentences in context: What brain waves can tell us. *Current Directions in Psychological Science, 17*(6), 376-380.

van Berkum, J., Brown, C. M., Zwitserlood, P., Koojiman, V. & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*(3), 443-467.

van Heugten, M., & Johnson, E. K. (2011). Gender-marked determiners help young Dutch learners' word recognition when gender information itself does not. *Journal of Child Language, 38*(1), 87-100.

van Heugten, M., & Shi, R. (2009). French-learning toddlers use gender information on determiners during word recognition. *Developmental Science, 12*(3), 419-425.

van Heugten, M., & Shi, R. (2010). Infants' sensitivity to non-adjacent dependencies across phonological phrase boundaries. *Journal of the Acoustical Society of America, 128*, EL223-EL228.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language. 61,* 206-237.

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology, 43*, 217-242.

Wexler, K. (1994). Optional infinitives, head movement and the economy of derivations. In D. Lightfoot & N. Hornstein (Eds.), *Verb movement* (pp. 305-350). Cambridge University Press.

Wexler, K. (2011). Grammatical computation in the optional infinitive stage. In J. de Villiers & T. Roeper (Eds.) *Handbook of generative approaches to language acquisition* (pp. 53-118). Springer.

Wicha, N. Y., Bates, E. M., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters, 346*, 165-168.

Wood, J. N., Kouider, S., & Carey, S. (2009). Acquisition of singular-plural morphology. *Developmental Psychology, 45*(1), 202-206.

Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development, 83*(4), 1382-1399.

Zangl, R., & Fernald, A. (2007). Increasing flexibility in children's online processing of grammatical and nonce determiners in fluent speech. *Language Learning and Development, 3*(3), 199-231.

Appendix A: Corpus Analysis

We examined samples of maternal speech to Adam (Adam01-Adam55; aged 2;3-5;2; Brown, 1973), available in the CHILDES database (MacWhinney, 2000). This sample, comprising 20,344 maternal utterances and 92,318 words, was searched for uncontracted occurrences of the words *are* and *is*, yielding 745 tokens of *are* and 1623 tokens of *is*. We chose uncontracted occurrences to make possible an automatic search: contracted *is* is a homograph of the possessive (*Adam's cookie*), contracted *has* (*He's already done that*), and contracted *does* (*What's that do?*). The subject of each sentence was identified by hand and tagged as a preverbal (*Your hands are dry*) or postverbal subject (*Are your hands washed?*); this tagging was checked by both authors. WH-NP's were not coded as subjects, because in WH-questions, as in our informative stimulus sentences, the verb agrees with the postverbal noun phrase (e.g., *What is this?, What are these?*). The only exceptions to this rule were sentences in which the WH-phrase was the only noun phrase in the sentence that was not embedded within a prepositional phrase (e.g., *What else is in the box*? *What toys are in there? How many pencils are there?*). This resulted in the counts shown in Table A.1, which show that about two-thirds of the tokens of *is* and *are* preceded their subjects. The preponderance of postverbal subjects in this sample reflects the high frequency of questions and locative inversions in child-directed speech (e.g., Broen, 1972; Newport et al., 1977).

Table A.1. The positions of the subjects of *is* and *are*.

	Preverbal	Postverbal	Subject Absent[8]	Total	Proportion Postverbal
is	584	1057	4	1645	.64
are	235	504	6	745	.68
Total	819	1561	10	2390	.65

[8] The missing subjects were in cases such as *Are what?* and *Is checking who?*, perhaps used to prompt completion of a prior utterance.

In order to achieve a rough estimate of how often verb agreement is neutralized by the use of second-person subjects (*Are you in here? Where are you?*), we categorized the subjects of *are*. This was again done by hand, and checked by both authors. As shown in Table A.2, half of the tokens of *are* in our sample occurred with the subject NP *you.* In this sample of child-directed speech, *you* typically refers to Adam, and thus has a singular referent. 31% of the tokens of *are* occurred with plural proform subjects (the most frequent were *those, they, these, we*), and 18% with plural lexical NP subjects (e.g., *his ears*, *all the pennies, these two*). We did not undertake this systematic coding for the subjects of *is,* because we expected all of these subjects to be grammatically singular nouns, either proper names (*Mr. Bear*), singular count nouns (*a pencil*), mass nouns (*his food*), or singular pronouns (*he, she, this*). An inspection of the corpus revealed no cases in which *is* occurred with a plural noun phrase subject, although there were cases in which the notional number of the subject NP was ambiguous (e.g., <u>*What else*</u> *is in the box*?). Taken together, the data shown here reveal exactly what we would expect: *are* and *is* are both quite flexible in their position relative to their subjects, and *are* very frequently occurs in second-person contexts that neutralize subject-verb number agreement; *is* suffers no such neutralization.

Table A.2. Types of subject NPs with *are*, including both pre- and postverbal subjects.

Subject Type	Token Frequency	Proportion of Total
2nd-person (*you*)	367	0.49
Plural NP	136	0.18
Plural Proform	232	0.31
Other[9]	4	0.005
Subject Absent	6	0.008
Total	745	

[9] This category comprised three conjoined singular NP subjects (*Paul and Diandra, the saddle and bridle, you and Robin*) and one singular proform (*what are that?*), apparently produced (or transcribed) in error.

Appendix B: Mixed-effects Model Syntax

All models were run using R (version 3.0.2), and lme4.0 (version 0.999999-3).

Experiment 1

Latency from determiner onset:

```
lmer(DeterminerRT ~ AgeGroup * Condition * TrialType * Plurality + (1 |
participant) + (1 | target))
```

Analysis of distractor-to-target shift probability:

```
glmer(DTShift ~ AgeGroup * Condition * TrialType * Plurality + (1 |
participant) + (1 | target), family = "binomial")
```

Experiment 2

Latency from determiner onset:

```
lmer(DeterminerRT ~ Condition * TrialType * Plurality + (1 | participant) +
(1 | target))
```

Analysis of distractor-to-target shift probability:

```
glmer(DTShift ~ Condition * TrialType * Plurality + (1| participant) +
(1|target), family = "binomial")
```

Analyses of fixation at noun onset:

```
glmer(NounOnsetLook ~ Condition * TrialType * Plurality + (1 | participant) +
(1 | target), family = "binomial")
```

Latency from noun onset:

```
lmer(NounRT ~ Condition * TrialType * Plurality + (1 | participant) + (1 +
TrialType | target))
```

Appendix C: Additional Measures, Experiment 1

To understand how the effect of the informative verb developed over time in Experiment 1, and for comparison to Experiment 2, we examined two measures anchored at the target noun: (a) participants' likelihood of fixating the target as opposed to the distractor at target-noun onset, and (b) participants' latency to shift from the distractor to the target after noun onset.

Probability of fixating the target at noun onset. At the moment of target-noun onset, children have not yet received information about the noun itself, but have already heard a potentially informative agreeing verb. Figure C.1 shows the mean proportion of looks directed to the target picture at noun onset, by trial-type, condition, age group and target plurality. Both children and adults were more likely to be fixating the target at noun onset in informative than in uninformative trials in the experimental condition only. This tendency was more pronounced in plural trials, especially for the children. This pattern was supported by a binomial mixed-effects regression model of target looks, with fitting procedures and predictor variables as in the main analyses above. Only the random slope of age group by target met the inclusion criterion (α = .2). Model results are shown in Table C.1.

This analysis revealed significant main effects of trial-type and plurality, the crucial significant interaction of condition and trial-type, a significant interaction of age group, condition and plurality, but no significant interactions involving trial-type, condition and age group. We again saw a marginal 3-way interaction of trial-type, condition and plurality; this reflects the numerical singular-plural asymmetry visible in Figure C.1. The simple main effect of trial type (informative vs. uninformative) was significant in the experimental condition for the adults ($\chi^2(1) = 14.5$, $p = .0001$) but not for the children ($\chi^2(1) = 1.5$, $p = .23$), and not in the control condition for either age group (children: $\chi^2(1) = 0.2$, p = .62; adults: $\chi^2(1) = 0.2$, $p = .66$).

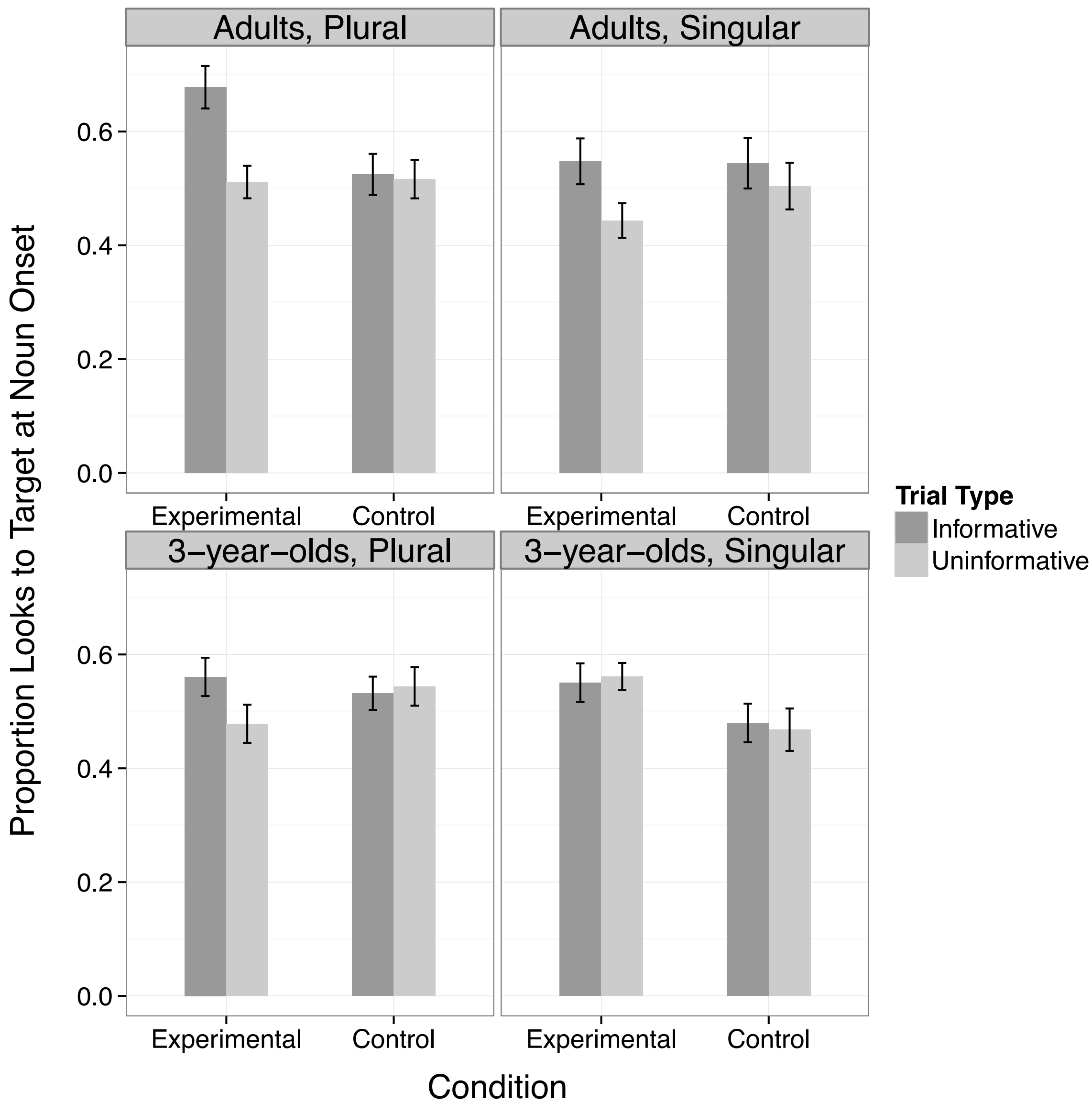

Figure C.1. Mean (se) proportion of trials in which participants were fixating the target at noun onset in Experiment 1, out of all trials in which they were fixating either the target or the distractor, by trial type (informative vs. uninformative), target plurality (singular vs. plural), age group (adults vs. children) and condition (experimental vs. control).

Table C.1. Estimates for binomial mixed effects model of the probability of fixating the target at noun onset, Experiment 1. Trial type and target plurality are within-participants factors, and condition and age group are between-participants factors. Statistically significant and marginal effects are shown in bold. (N=3237).

Fixed Effects

	Model Summary			Model Comparison		
	β	SE	z	χ^2	df	*p*
Intercept	0.109	0.084	1.30	1.52	1	0.22
Age Group	0.064	0.087	0.74	0.53	1	0.47
Condition	0.100	0.071	1.40	1.96	1	0.16
Trial Type	0.209	0.071	2.94	**8.62**	**1**	**0.003**
Plurality	0.141	0.071	1.98	**3.93**	**1**	**0.047**
Age Group × Condition	-0.025	0.143	-0.18	0.03	1	0.86
Age Group × Trial Type	0.201	0.143	1.40	1.97	1	0.16
Age Group × Plurality	0.127	0.143	0.89	0.79	1	0.37
Condition × Trial Type	0.282	0.142	1.98	**3.91**	**1**	**0.048**
Condition × Plurality	0.008	0.142	0.06	0.00	1	0.95
Trial Type × Plurality	0.133	0.142	0.94	0.87	1	0.35
Age Group × Condition × Trial Type	0.404	0.286	1.41	1.99	1	0.16
Age Group × Condition × Plurality	0.795	0.286	2.78	**7.73**	**1**	**0.005**
Age Group × Trial Type × Plurality	-0.036	0.286	-0.13	0.02	1	0.90
Condition × Trial Type × Plurality	0.490	0.285	1.72	**2.96**	**1**	**0.09**
Age Group × Condition × Trial Type × Plurality	-0.247	0.572	-0.43	0.18	1	0.67

Random Effects

	s^2
Participant (intercept)	0.000
Target (intercept)	0.047
Age group	0.020

The informative trial advantage observed in the experimental condition is additional evidence that participants used the number-marked verb predictively, to drive looks to a number-matching target before its label was heard.

Latency of first distractor-to-target shift from noun onset. Even in trials in which participants did not shift to the target by noun onset, a preceding agreeing verb might facilitate noun processing, leading to shorter shift latencies. Latency was measured in a 1500-ms window extending from 300 ms to 1800 ms after noun onset for children and from 200 ms to 1700 ms after noun onset for adults. 843 of children's trials (46% of 1852 included trials) and 689 of adults' trials (46% of 1504 included trials) were distractor-initial trials defined at noun onset. Of these, 671 of children's trials (80% of 843 trials) and 565 of adults' (82% of 689) included a direct distractor-to-target shift.

Figure C.2 shows the latency of the first distractor-to-target shift measured from noun onset. As shown in the Figure, latencies tended to be shorter in informative than uninformative trials in the experimental condition (adults: informative $M = 413$ ms, se = 31 ms; uninformative $M = 439$ ms, se = 31 ms; children: informative $M = 623$ ms, se = 23 ms; uninformative $M = 649$ ms, se = 21) as compared to the control condition (adults: informative $M = 454$ ms, se = 26 ms; uninformative $M = 402$ ms, se = 22 ms; children: informative $M = 593$ ms, se = 20; uninformative $M = 584$ ms, se = 22 ms). However, in contrast to the earlier measures, the differences were small, and the informative trial advantage appeared stronger in singular rather than plural trials.

To test this pattern, we fit a mixed-effects regression model of shift latency in R, using the *lmer()* function of the *lme4* package, with fitting procedures and predictor variables as in the main analyses. No random slopes met the inclusion criterion ($\alpha = .2$). Results are shown in Table

C.2. This analysis revealed significant main effects of age group and plurality, a marginal main effect of condition, and a significant interaction of trial type and plurality. The crucial interaction of trial type and condition was marginal. Follow-up comparisons using treatment coding to extract the simple main effect of trial type in each cell of the design were conducted for comparison to the other measures. The simple main effect of trial type was marginal in the adult control group ($\chi^2(1) = 3.7$, $p = .06$; note that this difference is in the wrong direction), and not significant elsewhere (all $\chi^2(1) < 1$).

Thus, as noted in the text, the numerical patterns were somewhat similar to those seen in our main measures – latencies tended to be shorter in informative trials –, but this pattern was not statistically robust. This suggests that while the effect of the agreeing verb can be extended in time, among adults and 3-year-olds the effect has mostly played out by the time information about the target noun is available. The pattern also differs in that it is numerically strongest in singular rather than plural trials. This suggests that singular agreement was not unhelpful, but that its effects may be slower to appear.

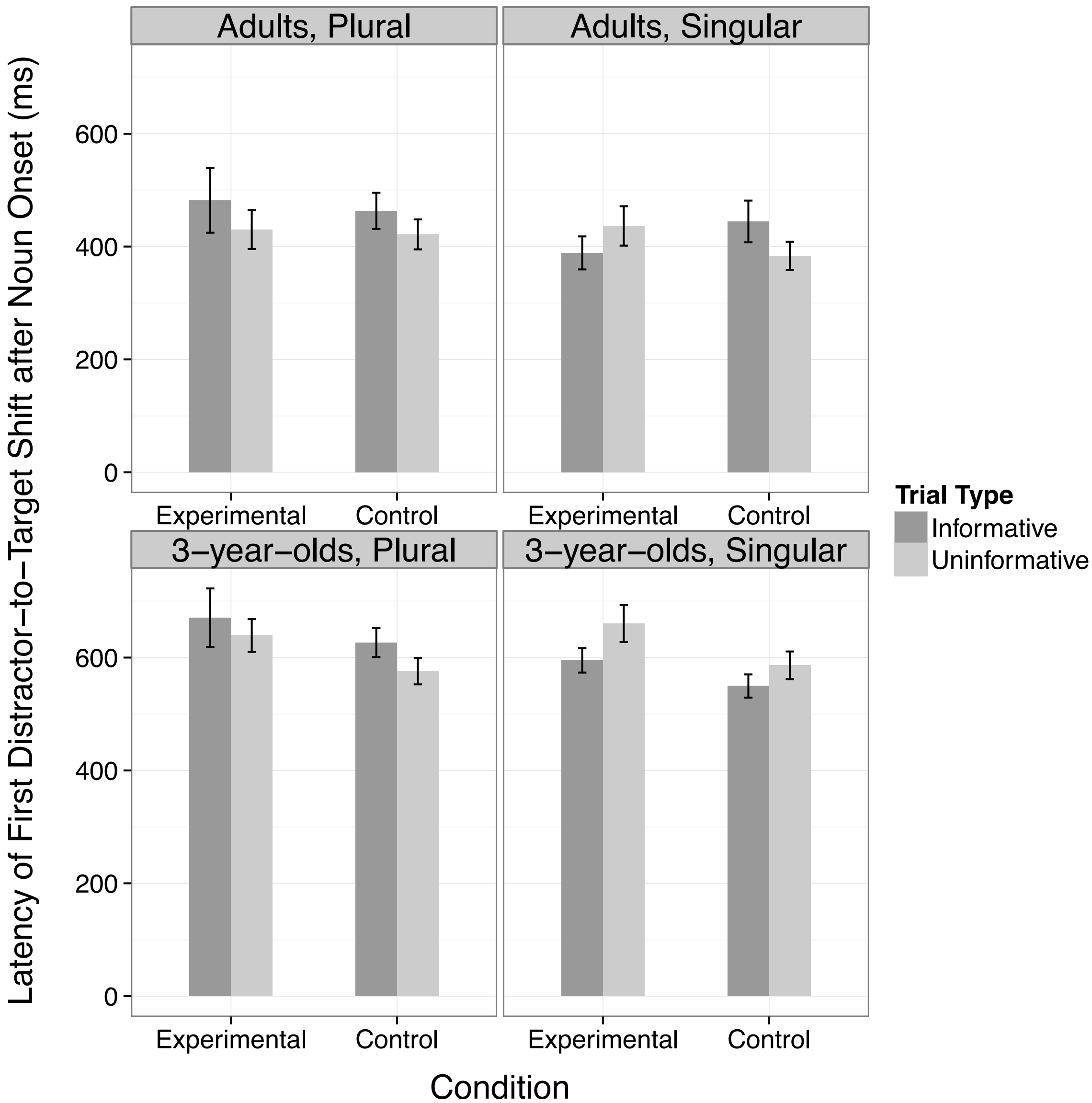

Figure C.2. Mean (se) latency of first shift from distractor to target in Experiment 1 measured from noun onset, by age group (3-year-olds, adults), condition (experimental vs. control), trial type (informative vs. uninformative) and plurality (singular, plural).

Table C.2. Estimates for binomial mixed effects model of latency of the first shift from distractor to target, measured from noun onset, in Experiment 1. Trial type and target plurality are within-participants factors, and condition and age group are between-participants factors. Statistically significant and marginal effects are shown in bold. (N=1236).

Fixed Effects

	Model Summary			Model Comparison		
	β	SE	t	χ^2	df	*p*
Intercept	528.9	12.9	40.96	**47.2**	**1**	**<.0001**
Age Group	-181.1	19.9	-9.09	**63.6**	**1**	**<.0001**
Condition	32.5	19.8	1.64	**2.7**	**1**	**0.097**
Trial Type	5.6	12.2	0.46	0.2	1	0.64
Plurality	28.6	12.2	2.34	**5.5**	**1**	**0.02**
Age Group × Condition	-53.4	39.9	-1.34	1.85	1	0.17
Age Group × Trial Type	29.4	24.6	1.20	1.43	1	0.23
Age Group × Plurality	19.0	24.5	0.77	0.61	1	0.44
Condition × Trial Type	-45.7	24.4	-1.87	**3.51**	**1**	**0.06**
Condition × Plurality	-15.5	24.4	-0.63	0.41	1	0.52
Trial Type × Plurality	55.8	24.4	2.28	**5.26**	**1**	**0.02**
Age Group × Condition × Trial Type	-18.1	49.1	-0.37	0.76	1	0.38
Age Group × Condition × Plurality	-4.4	49.1	-0.09	0.13	1	0.71
Age Group × Trial Type × Plurality	-32.0	49.1	-0.65	0.42	1	0.51
Condition × Trial Type × Plurality	26.1	49.0	0.53	0.28	1	0.60
Age Group × Condition × Trial Type × Plurality	78.5	98.3	0.80	0.64	1	0.42

Random Effects

	s^2
Participant (intercept)	6741.12
Target (intercept)	546.92