CAMBRIDGE
UNIVERSITY PRESS

Agreeing when to disagree: A corpus analysis of variable agreement in caregiver and child English

Cynthia Lukyanenko[1]* and Karen Miller[2]

[1]George Mason University, USA and [2]Penn State University, USA
*Corresponding author: Cynthia Lukyanenko. Email: clukyane@gmu.edu

Abstract

We characterized the patterns of agreement variation and consistency in three corpora of child and child-directed US English to better understand preschoolers' input and to compare preschoolers' own agreement production. We examined sentences with third-person subjects and tensed forms of BE in two large single-family corpora and one cross-sectional corpus collected during a Search-and-Find activity. Caregivers' agreement variation consistently reflected patterns previously found in adult-to-adult speech. Children's variation was conditioned by many of the same factors (e.g., sentence type, pronoun subject, and order of subject and verb) and clearly demonstrated acquisition of the categorical-variable split. However, some children showed substantially higher rates of nonagreeing forms (*There's the cherries*) than their caregivers and differed in their ranking of conditioning factors. We suggest that this reflects children's developing production processing abilities: shorter sentence-planning spans may make nonagreement a useful strategy for avoiding early number commitments in verb-first sentences.

Keywords: variable agreement; English; child language acquisition

Most English verbs agree with their subjects in person and number via a single affix marking the third-person singular form in the present tense. The verb BE has a particularly complex paradigm (for English), with agreement in both the present and past tense (e.g., *the baby is/was* versus *the kids are/were*) and a unique first-person singular present-tense form (*I am/was*). Variable agreement occurs in both prestige and vernacular Englishes around the world, particularly in plural existential sentences, as in (1).

(1) a. There were picnics   *agreement*   (Hay & Schreier, 2004:217)
     b. There's two doors   *nonagreement*

In these and other contexts, speakers may produce either an agreeing verb form, as in (1a), or a nonagreeing form, as in (1b).

Agreement variation has been extensively documented. Nonagreement occurs more often in speech than in writing (Crawford, 2005; Martinez Insua & Palacios Martinez, 2003), more often among less- than among more-educated speakers

CrossMark

(Britain & Sudbury, 2002; Hay & Schreier, 2004; Meechan & Foley, 1994; Tagliamonte, 1998), and is influenced by a variety of linguistic factors, including subject type, tense, and distance between subject and verb (Britain & Sudbury, 2002; Eisikovits, 1991; Hay & Schreier, 2004; Henry, 2016; Martinez Insua & Palacios Martinez, 2003; Meechan & Foley, 1994; Tagliamonte, 1998; Walker, 2021).

This combination of consistency and variability in English agreement creates a complex learning problem. Children must acquire a system in which the form of the verb *usually* depends on the properties of its subject, but in which certain combinations of social and linguistic factors *occasionally* lead to a different form. How do young children acquire both consistent agreement and the probabilistic patterns of agreement variation found in their communities? Does their developing model of this complex system align with their caregivers'?

English-learning children begin using agreeing verb forms in both production and comprehension around three years (Brown, 1973:271; Keeney & Wolfe, 1972; Lukyanenko & Fisher, 2016; Rissman, Legendre, & Landau, 2013; Theakston & Rowland, 2009), and reach high rates of accurate production around four years (Rice & Wexler, 2001:79). The comparatively few studies of children's acquisition of English agreement variation have found that Scottish and Northern Irish preschoolers produce variable agreement at rates that correlate with their input, and that their use, like their caregivers', is conditioned by sentence type (Belfast, Northern Ireland: Henry, 2016; Buckie, Scotland: Smith & Durham, 2019, Ch. 7). This is consistent with other evidence that the acquisition of variation proceeds in tandem with the acquisition of categorical patterns in language (see Smith & Durham, 2019, Ch. 1).

In the current paper, we explore agreement consistency and variation in child and caregiver English in the United States, focusing on children between two and six years old. Our first study explores agreement in two relatively large single-family corpora from the CHILDES database (MacWhinney, 2000), and the second explores a smaller corpus of caregiver and child speech during a semistructured "Search-and-Find" activity in the lab. In both studies, we take a broad approach to potential conditioning factors, including sentence types in our analyses that are rarely variable in adult English and others that nearly always are, to allow for the possibility that children's envelope of variation may differ from their caregivers'. Using this data, we characterize the envelope of variation, rates, and conditioning factors on agreement variation in caregiver and child US English.

Background

Acquisition of linguistic variation

Despite the fact that child-directed speech often differs substantially from adult-directed speech in speed, vocabulary, and sentence structure (e.g., Snow, 1972), rates and patterns of variation in child-directed speech closely resemble those in adult-to-adult speech (Miller, 2013a; Smith, Durham, & Fortune, 2007). When patterns differ, it can be as a direct result of other differences between child- and adult-directed speech (e.g., slower speech reducing deletion; Miller, 2013a), or as a result of social pressure (e.g., caregivers avoiding stigmatized variants such as English *ain't*; Foulkes, Docherty, & Watt, 2005; Smith et al., 2007).

Children produce variation as young as 2-3 years, but acquisition of the factors that constrain variation continues gradually through childhood (Chevrot, Nardy, & Barbu, 2011; Guy & Boyd, 1990; Kovac & Adamson, 1981; Miller, 2013a, 2013b, 2019; Shin, 2016; Shin & Miller, 2022; Smith, Durham, & Fortune, 2009; Smith et al., 2007). For instance, Guy and Boyd (1990) found that adults' -t/d omissions were conditioned by morphology, with more omissions on uninflected words (e.g., *mist*) than on regular (*missed*) or semiweak (*kept*) past-tense verbs. Children (ages 4-18), like adults, omitted -t/d more often on uninflected words, but also frequently omitted -t/d on semiweak verbs. Interestingly, when children fail to show sensitivity to a conditioning factor, they often appear to have regularized the input, using one variant near categorically (Shin & Miller, 2022, and references therein).

Variation can also affect the timing of acquisition. Children acquiring a morphological marker that is variably omitted often take longer to use the marker in comprehension (de Villiers & Johnson, 2007; Miller, 2007, Ch. 5; Miller & Schmitt, 2012), and children sometimes generalize variation to contexts where adults are categorical before arriving at adult patterns. Miller (2012) showed that children 3-5 years old who hear nonagreeing *don't* are more likely to produce nonagreeing *do* in yes-no questions ("Do your dad write with glitter glue?") than children without nonagreeing *don't* in their input, even though neither group hears nonagreeing *do* (see also Radford, 1992).

Children learn other aspects of variation readily. For instance, young children pick up on many categorical-variable splits, including in the agreement system. Smith and Durham (2019) found that children 2-4 years old in Buckie, Scotland, like their caregivers, never produced nonagreeing verb forms with *they* but did with NP subjects. Even in negation, where Buckie children acquired the variants sequentially, they respected categorical constraints: Like adults, children never used bare *na* with third-person singular subjects.

Agreement variation in English

Agreement variation in plural existentials as in (1) has been documented across centuries (Nevalainen, 2006) and in an extremely wide range of English varieties, from Belfast (Henry, 2016) to the American Rockies (Antieau, 2011) to the Falkland Islands (Britain & Sudbury, 2002). In mainstream Englishes, agreement variation occurs almost exclusively in existential and similar constructions (i.e., *here* and *where* sentences; Biber, Johansson, Leech, Conrad, & Finegan, 1999:186; Chambers, 2004; Krejci & Hilton, 2017). In varieties where variation occurs more broadly (e.g., in the past tense *They was/were very close*), nonagreement is more common in existentials than in other contexts (Antieau, 2011; Tagliamonte, 1998).

In existentials, variation typically involves a singular verb form occurring with a plural postverbal noun phrase, as in (1b). For instance, Crawford (2005) showed that in mainstream American and British English *there're* and *there are* were almost invariably followed by a plural noun (0.4% nonagreement), but rates of nonagreement were higher with *there's* and *there is* (15.6% and 4.4%, respectively). Variation occurs with full forms of BE but is most frequent with contracted *'s* (Crawford 2005; Krejci & Hilton, 2017; Martinez Insua & Palacios Martinez, 2003; Meechan & Foley, 1994).

Overview of the current study

In the current study, we examine agreement variation in US English, extracting all instances of third-person tensed BE and characterizing their subjects in two single-family corpora and one multifamily corpus. We examine and compare agreement in child and caregiver speech in a variety of contexts to explore whether children's envelope of variation matches their caregivers', and more broadly to ask how categorical and variable patterns are simultaneously acquired.

The single-family corpora let us characterize patterns of agreement variation in two families with different socioeconomic backgrounds and afford large enough samples to compare variation in a particular child's speech with their caregivers'. The cross-sectional corpus lets us estimate the prevalence of agreement variation across families. Examining both the input and children's production lays the groundwork for a better understanding of the learning processes that connect the two.

Methods

CHILDES corpora

We first examined two corpora from CHILDES (MacWhinney, 2000): Sarah (ages 2;3-5;1 [Brown, 1973]) and Nina (ages 1;11-3;3 [Suppes, 1974]). These relatively large, single-family corpora of US English allowed us to explore agreement variation in caregivers' child-directed speech, compare lower-SES and higher-SES caregivers' use of agreement variation, and compare individual children's patterns of agreement variation with their caregivers'.

Data for both corpora were collected in the late 1960s and early 1970s. Sarah's family was working-class and lived in the Northeast United States (Brown, 1973:51). Nina's family was middle-class (Miller, 2013b:307), and Nina lived with her mother in California, sometimes visiting her father in the Northeast. The children overlap in age, but Nina's recordings began when she was a few months younger and Sarah's continued longer. Nina's linguistic development was somewhat precocious, so despite the age difference, the children show substantial overlap in linguistic milestones during the time they were recorded (Miller, 2013b).

Extraction and exclusions

All sentences with tensed (potentially) third-person forms of BE (*is, are, was, were, 's, 're*) were extracted from transcripts of child and caregiver speech. In these sentences, the subject was identified. Tokens were then coded for subject person (second, third) and subject number (singular/plural/ambiguous).

Second-person sentences and third-person sentences in which the subject had ambiguous number (e.g., *mine, any*), could independently elicit variable agreement (e.g., conjoined singulars; Lorimor, 2007, Ch. 5), or was missing or unintelligible (e.g., *inside there is, I think xxx is a better idea*) were excluded, as were sentences in which *'s* was ambiguous between contracted *is, has*, or *does* (e.g., *What's he like?*), and sentences in which the verb had been added by the transcriber (e.g., *here ('s) horsie!, dere* [:there's] *a monkey.*). In the event of a self-correction, only the correction was analyzed.

After exclusions, there were a total of 25,566 tokens of BE with third-person singular or plural subjects for analysis (Sarah: $n = 2866$; Sarah's caregivers: $n = 6513$; Nina: $n = 4624$; Nina's caregivers: $n = 11{,}563$). Because there was essentially no variation in sentences with singular subjects, our analyses focused on the 3,247 sentences (13%) with plural subjects (Sarah: $n = 162$; Sarah's caregivers: $n = 453$; Nina: $n = 685$; Nina's caregivers: $n = 1947$). This includes both sentences where we expect adults' production to be variable and sentences where we would expect categorical production (see Table 1). By analyzing both, we address the question of whether children extend variation to nonvariable contexts or whether they show an adult-like categorical-variable split. We return briefly to sentences with singular subjects for an analysis of contractedness.

Coding and conditioning factors

Tokens were coded for a variety of potential conditioning factors. Representative examples are shown in Table 1.

Verb number. Each token was coded as *singular* or *plural*.

Speaker. Each utterance was tagged for speaker and categorized as *caregiver* (adult family member) or the target *child*. Speech from other adults (e.g., the researcher) or children (e.g., neighbors and friends) was not analyzed.

Sentence type. Previous studies of agreement variation single out there constructions[1] (Crawford, 2005; Martinez Insua & Palacios Martinez, 2003; Meechan & Foley, 1994), but other sources suggest that variation also occurs in superficially similar sentences with *here* and *where* (Biber et al., 1999; Chambers, 2004; Sparks, 1984). We therefore divided sentences into four categories: *where*, *here*, *there*, and *other* (see Table 1). Sentences that did not include *here*, *where*, or *there*, or in which the word was embedded inside another constituent (e.g., *That's the bin* [CP *where your toys go*], *Are the toys* [PP *over there*]?) were coded as *other*.

Subject type. Subject type often conditions agreement variation in non-there sentences (e.g., Hay & Schreier, 2004; Henry, 2016). In the current data, subjects were classified as either *pronouns* or *nonpronouns*. We defined the subject of the sentence as the (potential) agreement controller, regardless of its position in the sentence (e.g., *where are **your shoes**?, there's **his feet**, circles are here.*)

Order. The fact that the verb precedes the agreement controller in there sentences is cited as a potential reason for variation in these sentences (e.g., Chambers, 2004). However, order is rarely examined directly, and when it is, findings are mixed (e.g., Britain & Sudbury, 2002; Cheshire & Fox, 2009). We coded the order of subject and verb: *SV* versus *VS*.

Verb type. Existential constructions have a high frequency of copula BE. To ask if this plays a role in variation, we coded whether each instance of BE was a *copula* (e.g., *The cats are here*), or an *auxiliary* (e.g., *The cats are sleeping*). The few sentences in which the verb was ambiguous were grouped with auxiliaries (e.g., *Sure they are, Hot wheels are what?*).

We chose not to consider contractedness as a potential conditioning factor in our main analyses. We included both contracted and full-form verbs. Prior studies have found higher rates of nonagreement with *there's* than *there is* (e.g., Crawford, 2005),

Table 1. Sample sentences with plural subjects, divided by sentence type, order, and subject type

Subject Type	Sent. Type	Order	Sentence	Speaker	Child Age	File
non-pronoun	*here*	SV	*here the peas are for you*	Nina	2;5	nina28
		VS	*here's the candles*	Nina's mother	3;0	nina42
	there	SV	*what color blocks are there?*	Nina's mother	2;3	nina15
		VS	*is there more pieces of dirt?*	Nina's mother	2;4	nina20
	where	SV	*show me where the baby goats are*	Nina's mother	2;3	nina16
		VS	*where's the other crayons?*	Sarah's mother	4;5	sarah111
	other	SV	*my hands were sandy*	Nina	3;0	nina45
		VS	*are your hands sticky?*	Sarah's father	2;9	sarah029
pronoun	*here*	SV	*here they are*	Nina's grandmother	3;1	nina48
		VS	*are they here?*	Nina's mother	2;9	nina32
	there	SV	*hey (.) dere [:there] they are*	Sarah	3;4	sarah057
		VS	*what're they doing there?*	Nina	3;2	nina51
	where	SV	*I don't know where they are*	Sarah's mother	4;3	sarah102
		VS	*where are these?*	Sarah's mother	3;9	sarah077
	other	SV	*those are Mumma's dollies*	Sarah's mother	2;4	sarah007
		VS	*yeah (.) what're they on?*	Sarah's grandmother	2;6	sarah017

but for our analyses, we were concerned about the degree to which contractedness is confounded with sentence type, subject type, and order of subject and verb. It is rare to have a *there*, *where*, or *here* sentence with a pronoun subject and a contracted verb, because pronouns precede the verb in such sentences and contraction is impossible phrase-finally (e.g., *there they are*, **there they're*, **there they's*). Furthermore, the base rate of contraction is much higher for singular present tense verb forms than plural ones in these contexts. A search of the spoken subcorpus of the Corpus of Contemporary American English (Davies, 2008) showed that *there's* represents 66% of third-person singular present tense *there* + BE, while *there're* represents just 0.4% of third-person plural present tense *there* + BE (see also Westergren Axelsson, 1998). Thus, in sentences with plural subjects, base rates alone would result in contracted forms appearing to promote nonagreement. Other studies have noted these confounds and conducted descriptive analyses to explore contractedness patterns (e.g., Hay & Schreier, 2004; Meechan & Foley, 1994), and we do the same (see both sections below titled Contractedness).

We also excluded child age from our analyses. Both corpora are longitudinal, but the relative sparsity of morphosyntactic variables made formal analysis of change across development impractical. Despite the sparsity, visual inspection of the data broken out over six-month age-windows suggested that the patterns we describe below were stable for both caregivers and children.

Analysis approach

We take a three-step analysis approach. We first describe the data and then present two types of inferential statistical analyses: generalized linear models and best conditional inference trees (Tagliamonte & Baayen, 2012).[2] These analyses each provide different insights into the data: descriptive analyses present a general picture of how and where agreement variation occurs, generalized linear models estimate the contribution of each factor to the choice of a plural versus a singular verb, and conditional inference trees show which factors are most strongly predictive in which subsets of the data.

Results

Descriptive observations

Agreement production with plural subjects showed substantial variability between forms like those in (2a) and (2b) but was essentially categorical with singular subjects, as in (3).

(2) a. Where are your crayons?   Sarah's mother, 3;9
      b. Hey, where's your shoes?   Sarah's mother, 3;2
(3)   Where's the tree?   Nina's mother, 2;5

Figure 1 shows the distribution of plural verb forms across sentences with plural subjects, split by *corpus*, *speaker*, and *order* across the top, and by *subject type* and *sentence type* on the left. Nonagreement (i.e., lower percentages of plural verb-forms, darker cells) is common but not ubiquitous. It is, as expected, largely confined to

		Nina				Sarah			
		caregiver		child		caregiver		child	
		SV	VS	SV	VS	SV	VS	SV	VS
non-pronoun	here	100% (9/9)	62.2% (28/45)	100% (2/2)	13.3% (2/15)	100% (1/1)	0% (0/11)		8.3% (1/12)
	there	100% (10/10)	91.8% (123/134)	100% (2/2)	33.3% (10/30)	60% (3/5)	13.5% (5/37)		14.3% (2/14)
	where	100% (6/6)	97.5% (118/121)	100% (4/4)	4.2% (2/48)	100% (2/2)	7.4% (2/27)	100% (1/1)	0% (0/14)
	other	99.4% (158/159)	100% (147/147)	98.1% (102/104)	88.2% (15/17)	97.1% (68/70)	53.6% (15/28)	92.9% (26/28)	33.3% (3/9)
pronoun	here	100% (17/17)	100% (7/7)	100% (5/5)		100% (2/2)			
	there	100% (8/8)	100% (4/4)	100% (7/7)	100% (3/3)	100% (3/3)		100% (1/1)	
	where	100% (9/9)	100% (87/87)	100% (2/2)	100% (14/14)	100% (3/3)	100% (7/7)		100% (1/1)
	other	100% (646/646)	100% (538/538)	97.9% (379/387)	97.8% (44/45)	99.1% (220/222)	97.1% (34/35)	97% (64/66)	87.5% (14/16)

Figure 1. Percentage plural verb forms in sentences with plural subjects, split by corpus, speaker, and order of subject and verb on the horizontal axis, and by subject type and sentence type on the vertical axis. Number of plural verb tokens and total tokens are shown below percentages (plural/total). The darker the cell, the lower the rate of plural agreement. Blank cells indicate combinations of factors that did not occur.

sentences with postverbal, nonpronoun subjects. The proportion of nonagreement is particularly high for Sarah and her caregivers and for Nina. Nina's caregivers only occasionally produce nonagreement. Sarah and her caregivers produce nonagreement in *other* sentences, but Nina and her caregivers largely do not. Data is somewhat sparse, as indicated by the low token numbers in many cells and blank cells where factor combinations did not occur.

Generalized linear models

To more systematically explore the contribution of each conditioning factor, we fit two generalized linear models of verb form in sentences with plural subjects, one for each corpus. The models included the categorical predictors *speaker* (caregiver/child), *subject type* (pronoun/nonpronoun), *order* (SV/VS), *verb type* (auxiliary/copula), and *sentence type* (other/where/there/here), and no interactions. The first five predictors were binary and were entered into the model using effects coding, with the first-listed level coded as −0.5 and the second as 0.5. The last predictor had four levels and was entered using three treatment-coded contrasts, comparing *here*, *there*, and *where* sentences respectively to *other* sentences as a baseline. The dependent variable was verb form, with plural coded as 1 and singular as 0. Negative estimates therefore indicate more nonagreement in the second-listed level.

Nina's model revealed reliable effects of all contrasts except verb type, which was marginal, as shown on the left in Table 2. That is, Nina produced more nonagreement in sentences with plural subjects than her caregivers did, and sentences with nonpronoun subjects, sentences with VS order, and *there*, *where*, and *here* sentences had more nonagreement than sentences with pronoun subjects, SV order, and *other* sentences, respectively. The results of Sarah's model were similar, except that the effect of speaker is marginal, and strikingly smaller in magnitude (−0.79 versus −3.7 for Nina). This suggests that Sarah's production of plural verbs was more similar to her caregivers' than Nina's was.

Conditional inference trees

To explore the relationships among predictors, we fit a best conditional inference tree for each corpus. Best conditional inference trees are built in a series of binary splits. At each step, the data is split on the most strongly predictive factor: first the dataset as a whole, then each of the resulting subsets, until no further factors are reliably predictive. This results in a tree-like structure that reveals which factors are most strongly predictive in each subset of the data.

Figure 2 shows the best conditional inference tree for Nina's corpus. There are three key properties to notice. First, *where*, *there*, and *here* sentences are grouped together, opposite *other* sentences, suggesting that agreement is variable in all three structures. Second, speaker is a key factor on both main branches: Nina produces more singular verb forms than her caregivers for all sentence types. Third, splits below speaker differ. In *there*, *where*, and *here* sentences, Nina's caregivers differentiate between sentence types and, in *here* sentences, between SV and VS orders. In contrast, Nina differentiates primarily by subject type and order.

Figure 3 shows the best conditional inference tree for Sarah's corpus. In several respects, Sarah's data is like Nina's: *where, there*, and *here* sentences are grouped opposite *other* sentences and VS order favors singular verb forms. In contrast to Nina's data, Sarah does not differ strongly enough from her caregivers for speaker to appear as a predictor in the tree. *There, where*, and *here* sentences do not subdivide further, and there is substantial agreement variation even in *other* sentences, including effects of order and subject type.

Contractedness

Previous studies have found that plural subjects are more common with the reduced verb form *'s* than the full form *is* (e.g., Crawford, 2005; Hay & Schreier, 2004). Because of likely confounds, instead of including contractedness in our main analyses (see Coding and Conditioning Factors), we report an exploration of contractedness here. To avoid the baseline differences in contraction rates for *is* and *are*, we ask how often sentences with singular verb forms (*is, was, 's*) have singular or plural subjects.

First, we confirmed the previously observed pattern: Speakers produced more plural subjects with contracted singular verb forms than with full singular verb forms (Figure 4a). This was true for both caregivers and children, and despite the fact that our dataset includes a wider variety of sentence types than, for instance, Crawford's (2005) study of *there* + BE.

Table 2. Results from the generalized linear models of verb form in sentences with plural subjects for the Nina (*n* = 2632) and Sarah (*n* = 615) corpora

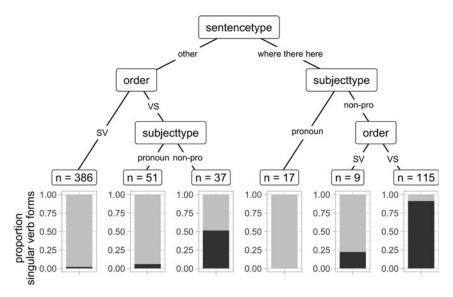| Factor	Nina					Sarah				
	Est.	95% CI	*p*-value	N	% Pl Verb	Est.	95% CI	*p*-value	N	% Pl Verb
(Intercept)	5.2	4.6, 5.9	<0.001			3.0	2.2, 4.0	<0.001		
Speaker	−3.7	−4.4, −3.2	<0.001			−0.77	−1.6, −0.00	0.052		
caregiver (−0.5)				1947	98.4				453	80.6
child (0.5)				685	86.6				162	69.8
Sentence Type										
(vs. other)				2043	99.3				474	93.7
here	−3.6	−4.6, −2.7	<0.001	100	70.0	−2.4	−4.2, −0.94	0.004	26	15.4
there	−1.7	−2.6, −0.81	<0.001	198	84.3	−1.8	−2.7, −0.88	<0.001	60	23.3
where	−1.8	−2.7, −1.0	<0.001	291	83.2	−1.8	−2.7, −0.88	<0.001	55	29.1
Order	−2.0	−2.9, −1.2	<0.001			−2.8	−3.6, −2.0	<0.001		
SV (−0.5)				1377	99.2				404	97.5
VS (0.5)				1255	91.0				211	39.8
Subject Type	−2.4	−3.2, −1.6	<0.001			−2.8	−3.8, −1.9	<0.001		
pronoun (−0.5)				1779	99.5				356	98.0
nonpronoun (0.5)				853	86.5				259	49.8
Verb Type	−0.88	−1.9, 0.02	0.072			−1.1	−3.0, 0.27	0.2		
auxiliary (−0.5)				993	99.4				121	98.3
copula (0.5)				1639	92.8				494	72.7

Figure 2. Best conditional inference tree for Nina's corpus. Node labels indicate the splitting factor and the edge labels indicate the levels of that factor in each branch. Each terminal node shows the number of tokens it contains (*n*) and the proportion of those tokens that have plural (light gray) and singular (dark gray) verb forms.
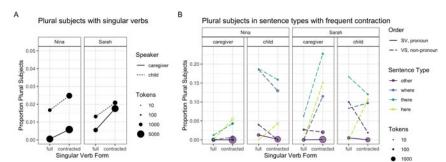
What drives this pattern? Is there something about contracted verb forms that permits variation, or are contracted forms and variation independently common in the same environments? To determine the likelihood of contraction independent of agreement variation, we examined contraction rates in sentences with both singular subjects and singular verb forms. The proportion of full-form verbs showed two notable patterns. First, contraction was present in almost all cells and common in many. Second, the four contexts in which children and caregivers in

Figure 3. Best conditional inference tree for Sarah's corpus. Node labels indicate the splitting factor, and the edge labels indicate the levels of that factor in each branch. Each terminal node shows the number of tokens it contains (*n*) and the proportion of those tokens that have plural (light gray) and singular (dark gray) verb forms.

Figure 4. The relationship between contractedness and agreement variation in sentences with singular verbs in the Sarah and Nina corpora. Panel A shows the overall rate of plural subjects with singular verb forms, split by contractedness, corpus, and speaker. Panel B shows rates of plural subjects with singular verb forms for cells with highest rates of contractedness, split by corpus, speaker, verb form, sentence type, subject type, and order. In both panels, point size indicates the number of contributing tokens.

both corpora were most likely to use contracted forms were VS *there*, *where*, and *here* sentences with nonpronoun subjects in VS order (*here* 4% full-form, *there* 22%, *where* 20%), and in *other* sentences with pronoun subjects in SV order (19% full-form). These contexts overlap substantially with the sentence types that promote nonagreement.

The independently high likelihood of contraction in key variable contexts might be enough to drive the pattern in Figure 4a. If so, once we control for sentence type, we would expect contracted and uncontracted verb forms to occur with plural subjects at similar rates. In contrast, if contraction uniformly promotes nonagreement, we would expect higher rates of plural subjects with contracted forms in all sentence types. The factors may also interact, with contraction occurring more often with plural subjects in some sentence types but not others.

Figure 4b shows the rate of plural subjects with contracted and full singular verb forms for five sentence types with high rates of contractedness: nonpronoun, VS *there, where, here*, and *other* sentences (e.g., there: *there's your ginger ale waiting for you*; where: *where's his legs?*; here: *here's the monkeys swinging*; other: *what is the owl sitting on?*), and pronoun, SV *other* sentences (e.g., *she's not afraid*). The plot shows that contractedness interacts with sentence type: caregivers produce more plural subjects with contracted verbs in VS *there, where*, and *here* sentences only. Rates are flat and low for SV *other* sentences and decrease for VS *other* sentences. Interestingly, the children are less consistent. Like their caregivers, they have more plural subjects with contracted singular verbs in VS *here* sentences, but they show flat or falling patterns for *there* and *where* sentences. Thus, while some of the effect in Figure 4a may come from the independent association of the same sentence types with contraction and variation, contractedness also promotes nonagreement in certain sentence types, particularly among caregivers.

Discussion

In our first study, we examined two corpora of caregiver and child speech and found substantial agreement variation. Variation occurred only in sentences with plural

subjects, and, like previous studies, we saw effects of subject type, order of subject and verb, and sentence type. We also observed substantial differences between caregiver rates of variation in Nina and Sarah's corpora, and between Nina's rate of variation and her caregivers'. In an additional descriptive analysis, we saw a small effect of contractedness on variation only in *there*, *where*, and *here* sentences in caregivers' speech.

These findings confirm that agreement variation is present in child-directed US English. The presence of variation, and the fact that the patterns echo those in previous studies, suggest that agreement variation is neither something that caregivers avoid nor something that independent properties of child-directed speech disfavor.

Another familiar pattern in these data is the difference between Sarah and Nina's caregivers. Previous studies have found that less-educated speakers tend to use higher rates of nonagreement (e.g., Meechan & Foley, 1994), and that variation in *other* sentences is largely absent in higher-prestige varieties of English (Chambers, 2004). Consistent with this, we see that rates of plural verb forms are much lower among Sarah's caregivers than among Nina's and that only Sarah's caregivers produce nonagreement in *other* sentences.

A new finding in the current data is the marked difference between Nina and her caregivers. While Sarah matches her caregivers relatively well, producing only slightly more nonagreement, Nina produces drastically more nonagreement than her caregivers. Why might this be? Looking at the descriptive analyses, Nina's nonagreement appears in the same cells as her caregivers': VS *where, there*, and *here* sentences. This suggests that Nina's over-production of nonagreement does not result from general confusion about agreement or about where variation is possible. However, in the conditional inference trees, the patterns that predict variation for Nina and her caregivers differ substantially. For her caregivers, sentence type is more predictive than order, and for Nina is the reverse. This mismatch suggests that Nina and her caregivers may be arriving at similar variation by different routes. This has interesting implications for acquisition: Children may be treating some potential conditioning factors as better bases for generalization than others, leading them to group *there*, *where*, and *here* sentences, even when they are differentiated in their input and despite the different linguistic analyses of their underlying structure. We return to this possibility in the General Discussion.

Search-and-Find corpus

Previous research suggests that agreement variation, particularly in existentials, is common not just across varieties but among individuals (e.g., Antieau, 2011; Hay & Schreier, 2004). To estimate its prevalence across families, we followed our analysis of Sarah and Nina's data by collecting a small corpus of caregiver and child speech using a Search-and-Find task. We first analyze data from all families together, providing information about patterns in the sample as a whole, and then we explore individual families' patterns.

In the Search-and-Find task, caregivers sat with their children and worked through a simple Search-and-Find book (see Figure 5) that we designed to elicit *there, where*, and *here* constructions, and to promote plural subjects. Though we recorded and transcribed both parent and child speech, child speech made up a

smaller proportion of the included sentences in this corpus (15% versus 31% for Sarah, 29% for Nina).

Methods

Participants

A total of one hundred English-speaking families participated in a Search-and-Find task over 105 sessions.[3] Sessions ranged from three to 18.5 minutes of recorded conversation (mean 7 minutes), for a total of about 12.5 hours of data. Children ranged in age from 1;7-6;0 (mean = 3;11, median = 3;11). Data was collected in 2016 and 2017 in central Pennsylvania. Participating caregivers grew up primarily in the Northeast and Mid-Atlantic US ($n = 77$) with smaller numbers from the Western US ($n = 8$), the Midwest ($n = 8$), and the US South ($n = 1$). Only one caregiver grew up outside the US (Toronto, Canada). The remaining caregivers specified broader regions (East/East Coast/Atlantic, $n = 5$; USA, $n = 1$). Caregivers' highest level of education ranged from high school ($n = 1$) to a PhD or MD ($n = 16$). The most common level was a Bachelor's degree ($n = 42$), and the median level was a Master's degree ($n = 35$). The remaining caregivers had some college or an Associate's degree ($n = 6$).

Two additional families participated in the task but were not included because the participating caregiver did not learn English in early childhood.[4]

Materials

Families were given a Search-and-Find book, in which each pair of pages included a moderately complex display of objects, and a smaller set of labeled objects (Figure 5). There were five pairs of pages (toys, beach, picnic, farm, bedroom), each with three singular and four plural items to locate.

Procedure

Sessions of the Search-and-Find task included one caregiver and their child or children. Caregivers and children were seated in adjacent chairs or with the child on the caregiver's lap in a corner of a quiet testing room with an audio recorder on a small

Figure 5. Sample pages from the Search-and-Find task.

table beside them. The researcher explained the task, then activated the recorder and moved behind a partition for the duration of the session. Caregivers were asked to work through the book with their child as they would at home.

Transcription, coding, and exclusions

Caregiver and child speech in each session was divided into turns and orthographically transcribed. Coding procedures and exclusions were identical to those described for the Sarah and Nina corpora above.

After exclusions, there were a total of 5,292 tokens of BE with third-person singular or plural subjects for analysis (children: 769 tokens, caregivers: 4,523 tokens). Of these, 26% ($n = 1383$) had plural subjects (children: $n = 170$, caregivers: $n = 1213$). Quantity of data varied between families. Children produced a median of five included sentences (range: 0-33), and caregivers produced a median of 39.5 (range: 3-139). As before, we focus on sentences with plural subjects and return to sentences with singular subjects only for an analysis of contractedness.

Results

Descriptive observations

Figure 6 shows the distribution of plural and singular verb forms in sentences with plural subjects in the Search-and-Find corpus, split by speaker, order, subject type, and sentence type. Patterns were very similar to those in Sarah and Nina's corpora: Singular verb forms (i.e., nonagreement) occurred primarily and frequently in sentences with postverbal, nonpronoun subjects, particularly *there*, *where*, and *here* sentences.

Generalized linear model

We fit a mixed-effects generalized linear model of verb form with the categorical predictors *speaker* (caregiver/child), *subject type* (pronoun/nonpronoun), *order* (SV/VS), *verb type* (auxiliary/copula), and *sentence type* (other/where/there/here), and random intercepts by family. No interaction terms were included. Coding and contrasts were identical to the models for Sarah and Nina (See above section titled Generalized Linear Model).

This model revealed reliable effects of speaker, order, subject type, and sentence types *there* and *here* versus *other*, as shown in Table 3. This means that children were less likely to provide plural verbs in sentences with plural subjects than their caregivers and that nonpronoun subjects, VS order, and *there* and *here* sentences all favored singular verb forms, as compared to pronoun subjects, SV order, and *other* sentences, respectively.

Conditional inference tree

As before, to better understand the relationships among these factors, we built the best conditional inference tree, shown in Figure 7. Three patterns stand out. First, this tree groups *where, there*, and *here* sentences opposite *other* sentences, even
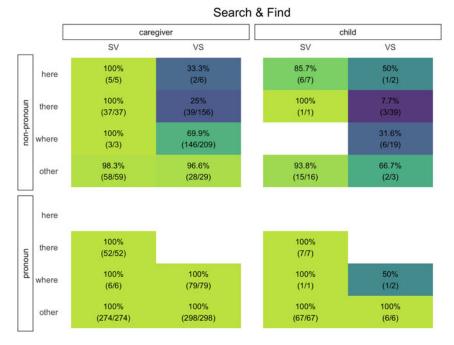
Figure 6. Percentage plural verb forms in sentences with plural subjects, split on the horizontal axis by speaker and order and on the vertical axis by subject type and sentence type. Number of plural verb tokens and total tokens (i.e., plural + singular verb forms) are shown below percentages (plural/total). The darker the cell, the lower the rate of plural agreement. Blank cells indicate combinations of factors that did not occur.

when the mixed effects model did not flag the *where-other* contrast as reliable. Even if the pattern is somewhat less reliable for *where* sentences in this data, there are still important similarities between patterns in *where* sentences and those in *there* and *here* sentences. Second, subject type and the order are again key predictors. Finally, children's rate of nonagreement in nonpronoun, VS, and *where* sentences is similar to the whole group's rates in nonpronoun, VS, *there*, and *here* sentences, although caregivers' rates are lower.

Contractedness

The effect of contractedness in sentences with singular verbs in the Search-and-Find corpus strongly resembled those for Sarah and Nina. First, we saw that both caregivers and children produced more plural subjects with contracted singular than with full singular verbs (Figure 8a).

As for Sarah and Nina, contracted singular verbs were common and were particularly likely in VS *there*, *where*, and *here* sentences, even when singular was the expected form. Figure 8b shows the rate of plural subjects with contracted and full-form singular verbs for nonpronoun, VS *there, where, here*, and *other* sentences, and pronoun and SV *other* sentences. As before, we found an interaction: SV *other* sentences with pronoun subjects and singular verbs never had plural subjects despite

Table 3. Results from the generalized linear models of verb form in sentences with plural subjects for Search-and-Find corpus (*n* = 1383)

Factor	Estimate	95% CI	*p*-value	*N*	% Pl Verb
		Search & Find			
(Intercept)	8.3	5.4, 11	<0.001		
Speaker	−2.1	−3.2, −1.1	<0.001		
caregiver (−0.5)				1213	84.7
child (0.5)				170	68.2
Sentence Type					
(vs. other)				752	99.5
here	−3.3	−5.7, −0.90	0.007	20	70.0
there	−4.5	−6.1, −2.9	<0.001	292	47.6
where	−1.2	−2.7, 0.29	0.11	319	75.9
Order	−5.2	−6.9, −3.6	<0.001		
SV (−0.5)				535	99.4
VS (0.5)				848	72.1
Subject Type	−5.2	−7.4, −3.1	<0.001		
pronoun (−0.5)				792	99.9
non-pronoun (0.5)				591	59.6
Verb Type	−2.5	−7.3, 2.2	0.3		
auxiliary (−0.5)				63	98.4
copula (0.5)				1320	81.9

Figure 7. Best conditional inference tree for the Search-and-Find corpus. Node labels indicate the splitting factor, and the edge labels indicate the levels of that factor in each branch. Each terminal node shows the number of tokens it contains (*n*) and the proportion of those tokens that have plural (light gray) and singular (dark gray) verb forms.
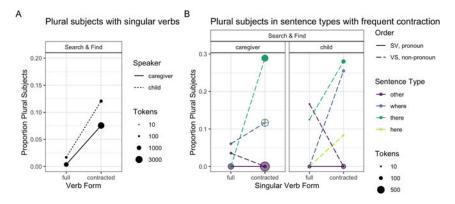
Figure 8. The relationship between contractedness and agreement variation in sentences with singular verbs in the Search-and-Find corpus. Panel A shows the overall rate of plural subjects with singular verb forms, split by contractedness and speaker. Panel B shows rates of plural subjects with singular verb forms for cells with highest rates of contractedness, split by speaker, verb form, sentence type, subject type, and order. In both panels, point size indicates the number of contributing tokens.

high rates of contraction. When they appeared with contracted verbs, VS *other* sentences with nonpronoun subjects were *less* likely to occur with plural subjects, and VS *there*, *where*, and *here* sentences were more likely to do so. In contrast to Sarah and Nina, and like their caregivers, the children in the Search-and-Find corpus produced more plural subjects with contracted than with full-form singular verbs in nonpronoun VS *here, there*, and *where* sentences.

Prevalence of variation across families

The cross-sectional sample allows us to characterize the prevalence of variation across families. Most caregivers produced at least one singular verb form with a plural subject (58/99,[5] 59%), though some categorically produced plural agreement (41/99, 41%). If we look only at those families who produced at least one likely context for variation (i.e., a VS *there*, *where*, or *here* sentence with a plural, nonpronoun subject), the proportion who produce at least one instance of nonagreement is even higher (58/78, 74%).

Figure 9a shows that as the number of likely contexts for variation increases, the likelihood of producing at least one instance of nonagreement does also. In contrast, the average rate of plural verb forms in likely contexts for variation remains stable regardless of the number of contexts produced (Figure 9b). Extreme values are most common among caregivers who produce few likely contexts for variation and occur in both directions (all plural or all singular). This suggests that extreme values may be the result of sampling error, and that it is likely that all families vary at least occasionally.

Discussion

Patterns in the Search-and-Find corpus echo those in Sarah and Nina's data: Agreement variation is present in caregivers' child-directed speech, it occurs only
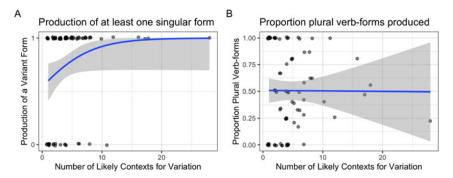
Figure 9. Patterns of variability across caregivers in the Search-and-Find task. Panel A shows the relationship between the number of likely contexts for variation (*there*, *where*, and *here* VS sentences with plural nonpronoun subjects) and caregivers' production of at least one instance of nonagreement, with a logistic fit. Panel B shows the relationship between the number of likely contexts for variation and the proportion of plural verb forms each caregiver produced, with a linear fit.

with plural subjects, and there are familiar effects of subject type, order, and sentence type. Like in Nina's corpus, we saw a reliable difference between caregivers and children, with children producing more nonagreement than adults. We saw a small effect of contractedness on caregivers' and children's speech in VS, pronoun subject, and in *there*, *where*, and *here* sentences.

In this corpus we were also able to explore how widespread agreement variation is across families. We argue that the patterns are consistent with widespread or universal variation: as families produced more likely contexts for variation, the probability of observing at least one instance of nonagreement rapidly approached certainty.

General discussion

In two studies of caregiver and child US English, we found substantial agreement variation. Adults' use of variable agreement in child-directed speech followed patterns familiar from studies of adult-to-adult speech. It occurred only with plural subjects, and nonagreement was more common in sentences with postverbal, nonpronoun subjects, particularly *there*, *where*, and *here* sentences.

Comparing children's variation to their caregivers' resulted in more complicated patterns. In all three corpora, children's production of agreement was conditioned by many of the same key factors: Like adults, children consistently produced singular verb forms with singular subjects, near-categorically produced plural verb forms with plural pronoun subjects and plural nonpronoun subjects that preceded the verb, and variably produced singular and plural verb forms elsewhere. However, Nina and the children in the Search-and-Find corpus differed from their caregivers in rate of agreement and ranking of conditioning factors, producing patterns that looked much more like Sarah's, who in turn matched her caregivers well.

These findings (a) provide crucial background information about agreement in children's linguistic input, (b) demonstrate that children's agreement production reflects sophisticated knowledge of linguistic variation and enrich our understanding

of how children learn language, and (c) inform analyses of the mechanism and sources of agreement variation. We briefly expand on each in turn.

Agreement variation in the input

Prior research on the acquisition of English verb agreement frequently assumes that it is categorical (e.g., Lukyanenko & Fisher, 2016; Theakston & Rowland, 2009). This is a reasonable simplifying assumption for studies using sentence types in which adults produce agreeing verb forms categorically (e.g., subject-first declaratives: *the tigers are holding the pen*; Theakston & Rowland, 2009:1454), and in the absence of detailed information about the presence and patterns of agreement variation in the appropriate variety of child-directed English. However, studying children's knowledge and use of variation stands to tell us at least as much about how children learn, categorize, and generalize as studying the places where behavior is categorical. Close examination of patterns of variability and consistency in the input is a crucial prerequisite to this work.

The current study demonstrates that agreement variation is common and widespread in child-directed US English, and that it patterns with adult-directed English (e.g., Crawford, 2005; Meechan & Foley, 1994; Walker, 2021). This provides important background information for future studies of English agreement acquisition. As a concrete example, it provides support for the speculation that variation may be a reason for the widely observed asymmetry between singular and plural verb forms in comprehension. In several studies, researchers have found that children are less likely to treat singular than plural forms of BE as informative cues to subject number (Davies, Rattanasone, & Demuth, 2020; Lukyanenko & Fisher, 2016; Lukyanenko & Miller, 2018). This asymmetry does not appear to extend to other cues to number (e.g., nominal plural: Davies, Rattanasone, Schembri, & Demuth, 2019; demonstratives: Reuter, Sullivan, & Lew-Williams, 2022), or even to contexts where variation with BE is less likely (yes-no questions: Deevy, Leonard, & Marchman, 2017, Figure 1), making variation a likely explanation. Children may be rightly treating *is* and *'s* as uninformative cues since, in their experience, singular forms of BE aren't picky about their subjects. The current study opens the way for further investigation of this phenomenon and for other studies of how children's real-time use of agreement during comprehension is influenced by patterns of variability and consistency in their input (Lukyanenko & Miller, 2018).

Children's use of categorical and variable patterns and implications for acquisition

Agreement variation presents an interesting challenge for learners, since the same verb forms are used categorically in some contexts and variably in others. Our data showed no tendency for children to impose categorical structure on adults' variability nor to vary where adults were categorical. Instead, children's production of English agreement respected an adult-like categorical-variable split from early childhood.

Children consistently produced singular verb forms with singular subjects, and plural forms with plural preverbal and pronoun subjects. This is consistent with

research demonstrating that young children respect categorical-variable splits (Johnson & White, 2019; Smith & Durham, 2019) and with findings from the acquisition literature that when children begin to produce agreement, they typically produce the expected form of agreeing verbs (e.g., Wexler, 2011).

Studies of artificial language learning suggest that children have a stronger tendency than adults to impose categorical structure, even on conditioned variation (Hudson Kam, 2015; Samara, Smith, Brown, & Wonnacott, 2017; Schwab, Lew-Williams, & Goldberg, 2018; Sneller & Newport, 2020). In the current study we saw no tendency for children to be more categorical than adults. Where children and caregivers differed, the children tended to be more variable. For instance, Nina's caregivers produced high rates of agreement with postverbal plural subjects in *there* sentences (123/134, 91.8%), but, rather than rounding up and producing only plurals, Nina regularly used both forms (10/30, 33.3% plural).

Our data show that children's agreement production reflects sensitivity to the patterns of variation in their caregivers' speech. Children show substantial alternation among the available forms in appropriate sentence types, and the same factors promote verb nonagreement for adults and children. This adds to evidence that young children use variable English agreement in relatively adult-like ways (Henry, 2016; Smith & Durham, 2019:148-160) and contrasts with studies of other variables that find categorical production first, followed by variation (e.g., *isn't* versus *ain't*: Miller 2015; negation: Smith & Durham, 2019:133-148; see Shin & Miller, 2022 for a review). It also suggests an alternative explanation for certain seemingly nonadult-like patterns observed in prior acquisition studies. It may be that children's more frequent "errors" with postverbal plural subjects (e.g., Theakston & Rowland, 2009) are not errors sparked by the high frequency of singular verb forms, but evidence of sensitivity to variation in the input.

Children were not entirely adult-like: Conditional inference trees indicated that both Nina and the children in the Search-and-Find corpus treat *there*, *where*, and *here* sentences more uniformly than their caregivers, and that Nina seems to place more weight on subject type and order than her caregivers. Differences of this kind represent an important area for future research. One possibility is that children track and represent different features than adults do. Preliminary evidence for this comes from a study demonstrating that children and adults differ in their use of singular forms in comprehension (Lukyanenko & Miller, 2018). In an eye-tracking task, adults treated full-form singular *is* as an informative cue to subject number, but children did not. Neither age group treated contracted *'s* as an informative cue. This suggests that adults track the contractedness of the verb, but that children do not. Consistent with this, in the current study, Sarah's and Nina's caregivers produced more plural subjects with contracted than with full-form singular verbs, but Sarah and Nina did not.

Analyses of English agreement variation

The current data have implications for linguistic analyses of English agreement variation. In particular, the commonalities between *there*, *where*, and *here* sentences that we quantify present a challenge for theoretical approaches that hinge on the existential structure of variable contexts (e.g., Meechan & Foley, 1994). Such analyses may

successfully extend to presentational *here* sentences but will likely struggle to explain variation in *where* sentences, given the different relationship between the subject and the verb.

Another class of explanations hinges on the nature of *'s*. Some proposals suggest that speakers treat *there's* as a single unit that does not participate in agreement dependencies (e.g., Rupp & Britain, 2019; Smith & Durham, 2019). Other proposals suggest that *'s* is or is becoming a nonagreeing clitic (e.g., Krejci & Hilton, 2017). As sole explanations, our data suggest that neither is sufficient. Variation in *there*, *where*, and *here* sentences requires the first group of proposals to posit many fused forms across a wider variety of sentence structures. Similarly, speakers' categorical use of agreement with pronoun subjects despite high rates of contractedness likely requires the second group to posit two versions of *'s*, one that occurs with pronouns that is retaining its agreement features, and one that occurs in *there*, *where*, and *here* sentences that is losing them.

In our view, the patterns of variation we describe are most consistent with a processing explanation or some combination of processing and the explanations above. Processing explanations suggest that it is more effortful to select an agreeing form when one must look ahead for the agreement controller than when the controller has already been produced (e.g., Chambers, 2004; Cheshire & Fox, 2009). Such explanations are consistent with children's higher rates of nonagreement, as well as the higher rates of nonagreement in the youngest Buckie children noted by Smith and Durham (2019:158). Because children plan their sentences in shorter chunks than adults (McDaniel, McKee, & Garrett, 2010; Redford, 2013), they might use singular verb forms as a strategy for avoiding early number commitments in VS sentences.

Conclusion

Our analyses demonstrate the widespread presence of agreement variation in child-directed US English, children's accurate production of categorical agreement, sensitivity to the same conditioning factors in adults' and children's production of variation, remarkable consistency in how variation patterns across *there*, *where*, and *here* sentences, and a tendency for some children to be more variable than their caregivers, but only in variable contexts.

The language acquisition literature and the sociolinguistics literature have a history of approaching children's language in very different ways, with acquisitionists focusing on categorical patterns (e.g., Keeney & Wolfe, 1972; Lukyanenko & Fisher, 2016; Theakston & Rowland, 2009), and the sociolinguists focusing on variation (e.g., Henry, 2016; Smith & Durham, 2019). We are not the first to notice that there is progress to be made by exploring consistent patterns, variability, and their interaction in acquisition (e.g., Johnson & White, 2019; Roberts, 1997). We hope that this study and others like it will inform future research in both fields.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10. 1017/S0954394523000054.

Data availability statement. Data and *R* code for all analyses in this paper may be accessed at https://osf. io/fwyvd/.

Competing interests. The authors declare none.

Notes

1. The presence of variation in nonexistential *where* and *here* sentences, the difficulty of discriminating true existential uses of *there* from locative ones in transcripts (locative: *There's a cat! Now where's the other one?* versus existential: *There's a cat! I didn't know you had a cat!*), and the fact that previous studies of variation have largely defined existentials as any sentence in which *there* immediately precedes BE (e.g., Cheshire & Fox, 2009; Eisikovits, 1991; Tagliamonte, 1998), led us to include locative and adverbial uses of *here*, *there*, and *where* (e.g., *there's the cat, the cat's there, is the cat there?, is there a cat?*). First, we reasoned that if these sentences were difficult for linguists to distinguish, young children might also group them together. Second, variation has been observed to be possible with both locative and existential *there* (Sparks, 1984:182). Further, taken together, the factors we coded successfully single out the same *there* sentences that have been the focus of previous work: *there* sentences with nonpronoun subjects and verb-subject order. See Table 1 and online Appendix 1 for more detail.

2. Analyses were conducted in *R* (R Core Team, 2021, version 4.1.0). Generalized linear models were fit using *lme4* (Bates, Maechler, Bolker, & Walker, 2015, version 1.1.27), and conditional inference trees using *partykit* (Hothorn & Zeileis, 2015; Hothorn, Hornik, & Zeileis, 2006, version 1.2.13). Graphs were built using *ggplot2* (Wickham, 2016) and *ggparty* (Borkovec & Madin, 2019), and the linear model summary tables were adapted from those output by *gtsummary* (Sjoberg, Curry, Hannum, Larmarange, Whiting, & Zabor, 2021).

3. Families who had more than one child in the target age range were given the option to participate in one session with both children or with each child separately.

4. This does not make agreement variation in their child-directed speech any less interesting than that of caregivers who grew up speaking English. However, because agreement variation in the speech of later learners of English is likely governed by different conditioning factors, we chose not to group them with early learners.

5. One of the one hundred families never produced BE with a plural subject.

References

Antieau, Lamont. (2011). Was/were variation in the middle Rocky Mountains. *Kansas Working Papers in Linguistics* 32:48–66.

Bates, Douglas, Maechler, Martin, Bolker, Ben, & Walker, Steve. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1):1–48.

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan, & Finegan, Edward. (1999). *Longman Grammar of Spoken and Written English*. New York: Longman.

Borkovec, Martin, & Madin, Niyaz. (2019). ggparty: 'ggplot' Visualizations for the 'partykit' Package. R package version 1.0.0. https://CRAN.R-project.org/package=ggparty

Breivik, Leiv Egil. (1981). On the interpretation of existential *there*. *Language* 57(1):1–25.

Britain, David, & Sudbury, Andrea. (2002). There's sheep and there's penguins: 'Drift', 'slant' and singular verb forms following existentials in New Zealand and Falkland Island English. In M. Jones & E. Esch (eds.), *Language change: The interplay of internal, external and extra-linguistic factors*. Berlin: De Gruyter Mouton. 209–42.

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

Chambers, J. K. (2004). Dynamic typology and vernacular universals. In B. Kortman (ed.), *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective*. Berlin: De Gruyter Mouton. 127–46.

Cheshire, Jenny, & Fox, Sue. (2009). Was/were variation: A perspective from London. *Language Variation and Change* 21(1):1–38.

Chevrot, Jean-Pierre, Nardy, Aurelie, & Barbu, Stéphanie. (2011). Developmental dynamics of SES-related differences in children's production of obligatory and variable phonological alternations. *Language Sciences* 33:180–91.

Crawford, William. (2005). Verb agreement and disagreement: A corpus investigation of concord variation in existential there + be constructions. *Journal of English Linguistics* 33:35–61.

Davies, Mark. (2008). *The corpus of contemporary American English (COCA)*. Available Online at https://www.english-corpora.org/coca/.

Davies, Benjamin, Rattanasone, Nan Xu, & Demuth, Katherine. (2020). Comprehension of the copula: Preschoolers (and sometimes adults) ignore subject–verb agreement during sentence processing. *Journal of Child Language* 47(3):695–708.

Davies, Benjamin, Rattanasone, Nan Xu, Schembri, Tamara, & Demuth, Katherine. (2019). Preschoolers' developing comprehension of the plural: The effects of number and allomorphic variation. *Journal of Experimental Child Psychology* 185:95–108.

Deevy, Patricia, Leonard, Laurence, & Marchman, Virginia. (2017). Sensitivity to morphosyntactic information in 3-year-old children with typical language development: A feasibility study. *Journal of Speech Language and Hearing Research* 60:668–74.

de Villiers, Jill G., & Johnson, Valerie E. (2007). The information in third-person /s/: Acquisition across dialects of American English. *Journal of Child Language* 34:133–58.

Eisikovits, Edina. (1991). Variation in subject–verb agreement in Inner Sydney English. In J. Cheshire (ed.), *English around the World*. Cambridge: Cambridge University Press. 235–55.

Foulkes, Paul, Docherty, Gerard, & Watt, Dominic. (2005). Phonological variation in child-directed speech. *Language* 81:177–206.

Guy, Gregory, & Boyd, Sally. (1990). The development of a morphological class. *Language Variation and Change* 2(1):1–18.

Hay, Jennifer, & Schreier, Daniel. (2004). Reversing the trajectory of language change: Subject-verb agreement with be in New Zealand. *Language Variation and Change* 16(3):209–35.

Henry, Alison. (2016). Acquiring language from variable input subject-verb agreement and negative concord in Belfast English. *Linguistic Variation* 16(1):131–50.

Hothorn, Torsten, Hornik, Kurt, & Zeileis, Achim (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3):651–74.

Hothorn, Torsten, & Zeileis, Achim. (2015). partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research* 16:3905–09.

Hudson Kam, Carla. (2015). The impact of conditioning variables on the acquisition of variation in adult and child learners. *Language* 91(4):906–37.

Johnson, Elizabeth K., & White, Katherine S. (2019). Developmental sociolinguistics: Children's acquisition of language variation. *WIREs Cognitive Science* 1515:1–15.

Keeney, Terrence, & Wolfe, Jean. (1972). The acquisition of agreement in English. *Journal of Verbal Learning & Verbal Behavior* 11(6):698–705.

Kovac, Ceil, & Adamson, H. D. (1981). Variation theory and first language acquisition. In D. Sankoff & H. Cedergren (eds.), *Variation omnibus: Current inquiry into language and linguistics*. Alberta: Linguistics Research Inc. 403–10.

Krejci, Bonnie, & Hilton, Katherine. (2017). There's three variants: Agreement variation in existential *there* constructions. *Language Variation and Change* 29(2):187–204.

Lorimor, Heidi. (2007). *Conjunctions and grammatical agreement*. Doctoral dissertation, University of Illinois at Urbana-Champaign.

Lukyanenko, Cynthia, & Fisher, Cynthia. (2016). Where are the cookies: Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition* 146:349–70.

Lukyanenko, Cynthia, & Miller, Karen. (2018). Children's and adults' processing of variable agreement patterns: Agreement neutralization in English. In A. B. Bertolini & M. J. Kaplan (eds.), *Proceedings of the 42nd Annual Boston University Conference on Language Development*. Somerville: Cascadilla Press. 479–92.

MacWhinney, Brian. (2000). *The CHILDES project: Tools for analyzing talk, 3rd Edition*. Mahwah, NJ: Lawrence Erlbaum.

Martinez Insua, Ana, & Palacios Martinez, Ignacio. (2003). A corpus-based approach to non-concord in present-day English existential *there*-constructions. *English Studies* 84(3):262–83.

McDaniel, Dana, McKee, Cecile, & Garrett, Merrill. (2010). Children's sentence planning: Syntactic correlates of fluency variations. *Journal of Child Language* 37:59–94.

Meechan, Marjory, & Foley, Michele. (1994). On resolving disagreement: Linguistic theory and variation—there's bridges. *Language Variation and Change* 6(1):63–85.

Miller, Karen. (2007). *Variable input and the acquisition of plurality in two varieties of Spanish.* Doctoral dissertation, Michigan State University.

Miller, Karen. (2012). Not all children agree: Acquisition of agreement when the input is variable. *Language Learning and Development* 8:255–77.

Miller, Karen. (2013a). Acquisition of variable rules: /s/-lenition in the speech of Chilean Spanish-speaking children and their caregivers. *Language Variation and Change* 25(3):311–40.

Miller, Karen. (2013b). Variable input: What Sarah reveals about non-agreeing don't and theories of root infinitives. *Language Acquisition: A Journal of Developmental Linguistics* 20(4):305–24.

Miller, Karen. (2015). Children's production of ain't. In P. Donaher & S. Katz (eds.), *Ain'thology: The history and life of the taboo word ain't.* Newcastle: Cambridge Scholars Publishing. 96–112.

Miller, Karen. (2019). Children's acquisition of sociolinguistic variation. In T. Ionin & M. Rispoli (eds.), *Three streams of generative language acquisition research: Selected papers from the 7th meeting of Generative Approaches to Language Acquisition – North America.* Amsterdam: John Benjamins. 35–58.

Miller, Karen, & Schmitt, Cristina. (2012). Variable input and the acquisition of plural morphology. *Language Acquisition: A Journal of Developmental Linguistics* 19(3):223–61.

Nevalainen, Terttu. (2006). Negative concord as an English 'vernacular universal': Social history and linguistic typology. *Journal of English Linguistics* 34(3):257–78.

Radford, Andrew. (1992). The acquisition of the morphosyntax of finite verbs in English. In J. Meisel (ed.), *The acquisition of verb placement: Functional categories and V2 phenomena in language acquisition.* Dordrecht: Kluwer Academic Publishers. 23–62.

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Redford, Melissa A. (2013). A comparative analysis of pausing in child and adult storytelling. *Applied Psycholinguistics* 34(3):569–89.

Reuter, Tracy, Sullivan, Mia, & Lew-Williams, Casey. (2022). Look at that: Spatial deixis reveals experience-related differences in prediction. *Language Acquisition: A Journal of Developmental Linguistics* 29(1):1–26.

Rice, Mabel, & Wexler, Kenneth. (2001). *Test of early grammatical impairment (TEGI).* New York: Pearson Publishing.

Rissman, Lilia, Legendre, Geraldine, & Landau, Barbara. (2013). Abstract morphosyntax in two-and three-year-old children: Evidence from priming. *Language Learning and Development* 9(3):278–92.

Roberts, Julie. (1997). Acquisition of variable rules: (-t,d) Deletion and (ing) production in preschool children. *Journal of Child Language* 24:351–72.

Rupp, Laura, & Britain, David. (2019). *Linguistic perspectives on a variable English morpheme.* London: Palgrave Macmillan.

Samara, Anna, Smith, Kenny, Brown, Helen, & Wonnacott, Elizabeth. (2017). Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cognitive Psychology* 94:85–114.

Schwab, Jessica F., Lew-Williams, Casey, & Goldberg, Adele E. (2018). When regularization gets it wrong: Children over-simplify language input only in production. *Journal of Child Language* 45(5):1054–72.

Shin, Naomi. (2016). Acquiring patterns of morphosyntactic variation: Children's Spanish subject pronoun expression. *Journal of Child Language* 43(4):914–47.

Shin, Naomi, & Miller, Karen. (2022). Children's acquisition of morphosyntactic variation. *Language Learning and Development* 18(2):125–50.

Sjoberg, Daniel D., Curry, Michael, Hannum, Margie, Larmarange, Joseph, Whiting, Karissa & Zabor, Emily C. (2021). *gtsummary: Presentation-ready data summary and analytic result tables.* R package version 1.4.1. https://CRAN.R-project.org/package=gtsummary

Smith, Jennifer, & Durham, Mercedes. (2019). *Sociolinguistic variation in children's language: Acquiring community norms.* Cambridge: Cambridge University Press.

Smith, Jennifer, Durham, Mercedes, & Fortune, Liane. (2007). Mam, my trousers is fa'in doon!: Community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change* 19(1):63–99.

Smith, Jennifer, Durham, Mercedes, & Fortune, Liane. (2009). Universal and dialect-specific pathways of acquisition: Caregivers, children, and t/d deletion. *Language Variation and Change* 21(1):69–95.

Sneller, Betsy, & Newport, Elisa. (2020). Acquisition of phonological variation: Evidence from artificial language learning. Poster presented at The Boston University Conference on Language Development.

Snow, Catherine E. (1972). Mothers' speech to children learning language. *Child Development* 43(2):549–65.

Sparks, Randall B. (1984). Here's a few more facts. *Linguistic Inquiry* 15:179–83.

Suppes, Patrick. (1974). The semantics of children's language. *American Psychologist* 29:103–14.

Tagliamonte, Sali. (1998). Was/were variation across the generations: View from the city of York. *Language Variation and Change* 10(2):153–92.

Tagliamonte, Sali, & Baayen, Harald. (2012). Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24(2): 135–78.

Theakston, Anna, & Rowland, Caroline. (2009). The acquisition of auxiliary syntax: A longitudinal elicitation study. Part 1: Auxiliary be. *Journal of Speech, Language, and Hearing Research* 52:1449–70.

Walker, James A. (2021). Modelling socio-grammatical variation: Plural existentials in Toronto English. In K. V. Beaman, I. Buchstaller, S. Fox & J. A. Walker (eds.), *Advancing socio-grammatical variation and change: In honour of Jenny Cheshire*. New York: Routledge. 165–84.

Westergren Axelson, Margareta. (1998). Contraction in British newspapers in the late 20th century. *Acta Universitatis Upsaliensis: Studia Anglistica Upsaleiensia* 102.

Wexler, Kenneth. (2011). Grammatical computation in the optional infinitive stage. In J. de Villiers & T. Roeper (eds.), *Handbook of generative approaches to language acquisition*. Dordrecht: Springer. 53–118.

Wickham, Hadley. (2016), *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.