Receptive and expressive language ability differentially support symbolic understanding over time: Picture comprehension in late talking and typically developing children

Rachael W. Cheung [a,b,*], Calum Hartley [a], Padraic Monaghan [a]

[a] Department of Psychology, Lancaster University, Bailrigg, Lancaster LA1 4YF, UK
[b] Psychology in Education Research Centre, Department of Education, University of York, York YO10 5DD, UK

ARTICLE INFO

ABSTRACT

Symbols are a hallmark of human communication, and a key question is how children's emerging language skills relate to their ability to comprehend symbols. In particular, receptive and expressive vocabulary may have related, but distinct, roles across early development. In a longitudinal study of late talking (LT) and typically developing (TD) children, we differentiated the extent to which expressive and receptive language skills predicted symbolic understanding as reflected in picture comprehension and how language skills inter-related with social skills. LT and TD children were tested on a picture comprehension task that manipulated the availability of verbal labels at 2.0–2.4 years and 3.5–3.9 years of age. Although all children improved in accuracy over time as expected, TD children exhibited an advantage over LT children, despite both groups using verbal labels to inform their mapping of picture–object relationships. Receptive and expressive vocabulary also differed in their contribution at different ages; receptive vocabulary predicted performance at ~2.0 years of age, and expressive vocabulary predicted performance at ~3.5 years of age. Task performance at 3.5 years was predicted by earlier receptive vocabulary, but this effect was largely mediated by concurrent expressive vocabulary. Social ability across the whole sample at ~2.0 years also predicted and mediated the effect of receptive vocabulary on concurrent task performance. These findings suggest that LT

* Corresponding author at: Psychology in Education Research Centre, Department of Education, University of York, York YO10 5DD, UK.

E-mail address: rachael.cheung@york.ac.uk (R.W. Cheung).

children may have delays in developing picture comprehension over time and also that social ability and language skills may differentially relate to symbolic understanding at key moments across development.

Introduction

The use of symbols is a uniquely human cognitive hallmark and is vital to communication (DeLoache, 1995; Tomasello, Carpenter, Call, Behne, & Moll, 2005). A symbol is something that someone intends to represent something else and can take many forms, including gestures, graphics, text, words, and maps (DeLoache, 2004). Children are immersed in a symbolic world from infancy, and the types of symbols children understand are subject to both cultural context and social scaffolding (Callaghan et al., 2011; Rakoczy, Tomasello, & Striano, 2005).

Children in Western societies are exposed to pictures from an early age. Children use linguistic labels to scaffold their understanding of pictures (Callaghan, 2000), and the development of language and other symbolic domains, such as symbolic play, are closely related (Quinn, Donnelly, & Kidd, 2018). This means that early language impairments have the potential to also affect children's understanding of nonlinguistic symbol systems. Although the literature has established that symbolic understanding, language ability, and social context interact in typical development (Callaghan & Corbit, 2015), we do not fully understand how these domains affect each other over time. Furthermore, their trajectory in atypical development is not well defined, and the effect of language delay on how children understand pictures remains under-investigated. Examining the effect of language delay on picture comprehension is crucial to understanding whether children with these difficulties have functional impairments in additional symbolic domains and also offers an opportunity to elucidate how language scaffolds symbolic understanding during development.

Language and picture comprehension in typical development

For children to understand pictures as symbols, they need to acquire *dual representation*—the understanding that a symbol is not just an object, but also a representation of something else (DeLoache, 2004). At 9 months of age, infants manually investigate pictures as if they were real objects, grasping and plucking at depicted items (DeLoache, Pierroutsakos, Uttal, Rosengren, & Gottlieb, 1998). By 18 months of age, they begin pointing and talking about pictures rather than handling them, suggesting that they have begun to treat pictures as symbols rather than as objects in themselves (Pierroutsakos & DeLoache, 2003).

Language can aid children in understanding the representative nature of pictures because verbal labels provide clues about how two-dimensional visual symbols relate to referents in the world (Callaghan, 2000; Ganea, Allen, Butler, Carey, & DeLoache, 2009). When testing 2-year-olds, Preissler and Bloom (2007) demonstrated that labeling a picture of an unfamiliar object ("This is a *wug*. Can you show me another one?") directed children to identify the symbolized object 90% of the time, whereas children identified symbolized referents only 30% of the time when pictures were not labeled ("Look at this. Can you show me another one?"). As children quickly learn that verbal labels refer to objects in the world, the act of labeling cues children to view pictures as symbolic representations rather than as objects. Children aged 15, 18, and 24 months will spontaneously extend a novel label (e.g., "whisk") taught using a picture to its corresponding three-dimensional referent (e.g., an actual whisk) (Ganea et al., 2009; Preissler & Carey, 2004). These findings show that young children understand that verbal labels paired with pictures refer to independently existing referents and also that the pictures themselves are representational and not the exclusive referents for their associated labels.

However, language itself is a symbol system in which caregivers heavily invest, going to considerable lengths to teach their children words. Thus, children may learn verbal representations for concepts (e.g., understanding how the label "dog" relates to the world) before they learn how pictures or other symbols relate to the same concept. Callaghan (2000) explicitly demonstrated that children use verbal labels to scaffold their understanding of pictures and objects but also that this differs according to age. Children were shown a series of line drawings and were asked to identify the referent of each drawing from a pair of three-dimensional objects (with the picture removed from view). In some trials, linguistic scaffolding was unavailable, with the two paired objects having the same category label (e.g., two types of dog). In other trials, linguistic scaffolding could be used, with the two paired objects having distinct category labels (e.g., dog and cat). The study demonstrated that 2.5-year-olds performed above chance only when pictures could be unambiguously matched to objects using distinct verbal labels, whereas 3-year-olds performed above chance even without linguistic scaffolding. For younger children, whose understanding of the pictorial symbol system was relatively fragile, verbal labels were valuable in bridging the gap between images and their depicted referents. Older children, however, were able to rely on the perceptual similarities between images and their referents to accurately identify picture–object relationships in the absence of linguistic scaffolding.

More broadly, language may provide a basis for other symbol systems during development (Callaghan, 2020; Nelson, 2007; Tomasello, 2003). A meta-analysis of symbolic play studies found a significant interaction for symbolic play between age and whether expressive or receptive language measures were used (35 studies, p = .006; Quinn et al., 2018). This demonstrated that symbolic play was related to concurrent receptive measures in children under 3 years of age (r = .41), whereas concurrent expressive measures better predicted symbolic play in studies of children over 3 years of age (r = .36). However, this interaction was driven by a difference in effect sizes for receptive vocabulary, rather than expressive vocabulary, given that the expressive effect size remained stable across ages, making any differential effects at different ages hard to clearly identify. Because picture comprehension and symbolic play skills appear to be closely related (Rochat & Callaghan, 2005), any differential effects of emerging language ability on symbolic play may also affect picture comprehension at different ages.

Few studies have assessed how picture comprehension and language skills inter-relate during early development. Of these studies, some have found different effects of receptive and expressive language ability on pictorial understanding and wider symbolic ability. Callaghan and Rankin (2002) assessed picture comprehension and production at 28, 36, and 42 months of age. The authors found that picture comprehension scores positively correlated with receptive language and that picture production scores positively correlated with expressive language. Kirkham, Stewart, and Kidd (2013) also assessed the relationship among language, pictorial understanding, and symbolic play. They found that the mean length of five longest utterances (MLU5) at 4 years of age predicted symbolic play and representational drawing ability at 5 years of age, and that receptive and expressive language score combined at 4 years predicted symbolic play at 5 years. Receptive vocabulary has also been found to correlate with performance on scale model and picture search tasks (finding a hidden object in a room based on the location of a miniature object positioned in a scale model of the same room or searching a location represented by a picture; Hartley & Allen, 2015a; Homer & Nelson, 2009).

In summary, cross-sectional and longitudinal evidence shows that linguistic and nonlinguistic symbolic domains are developmentally inter-related. Verbal labeling scaffolds symbolic understanding of picture–object relationships (Callaghan, 2000; Ganea et al., 2009), and expressive and receptive language abilities correlate with pictorial tasks but may exhibit different effects at different ages (Callaghan & Rankin, 2002; Kirkham et al., 2013). However, we do not know whether early language delays cause deficits in picture comprehension over time. Furthermore, in typical development, differential effects of expressive and receptive language on symbolic ability have proved to be difficult to identify (Quinn et al., 2018). Studying productive language impairments provides a unique opportunity to explore how receptive and expressive language skills interact differentially with pictorial understanding over time.

Language and picture comprehension in atypical development

Late talking (LT) children are typically at or below the 10th percentile of expressive vocabulary compared with other children their age between 18- to 30-months-old, without neurodevelopmental or sensory deficits (Fisher, 2017). Although the majority of LT children reach typically developing expressive range by approximately 5 years of age (Rescorla, 2011), a minority (∼12–25%; Collisson et al., 2016; Henrichs et al., 2011; Reilly et al., 2010; Roulstone, Loader, Northstone, & Beveridge, 2002; Zubrick et al., 2007) develop developmental language disorder (DLD). Furthermore, despite many LT children reaching the neurotypical range for expressive vocabulary by school age, they consistently score on the lower end of this range across a variety of language measures (Domsch et al., 2012; Rescorla, 2002, 2005; Rescorla, Dahlsgaard, & Roberts, 2000; Rice, Taylor, & Zubrick, 2008).

LT children are characterized by expressive vocabulary deficits yet can have varying receptive vocabulary skills (Fisher, 2017), whereas expressive and receptive vocabulary are more tightly intertwined in typically developing (TD) children. Evidence that expressive and receptive vocabulary might exert differential effects on pictorial understanding can be found in autism spectrum disorder (ASD) studies, where children with ASD often exhibit a range of language difficulties (Eigsti, de Marchena, Schuh, & Kelley, 2011). Studies that test the extension of words from pictures to symbolized referents in minimally verbal children with ASD, who are matched with TD children on receptive vocabulary, have found deficits in the ASD sample (mean receptive age ∼3.5 years; Hartley & Allen, 2015b; Preissler, 2008). However, when adapting Callaghan's (2000) linguistic scaffolding task for TD and ASD samples that were matched on both expressive *and* receptive language (mean ∼4.5 years of age), Hartley, Trainer, and Allen (2019) found that children with ASD and TD children performed identically across all trial types. Both samples showed lower accuracy on trials where they could not use verbal labels relative to trials where they could. In the ASD sample, both receptive and expressive language predicted task performance; in the TD sample, only receptive language was predictive. These studies suggest that children with expressive deficits, but not receptive deficits, might struggle with using verbal scaffolding in pictorial understanding tasks.

One possible explanation for differential effects of receptive and expressive language on pictorial understanding is simply that children who say less experience fewer opportunities to participate in social situations where pictures are used. Many accounts of symbolic understanding rely on a foundation of sociocognitive skills such as imitation and intention reading (Nelson, 2007; Rakoczy et al., 2005; Rochat & Callaghan, 2005; Tomasello, 2003, 2010; Vygotsky, 1980). For example, Rochat and Callaghan (2005) argued that pictures are inherently communicative, and understanding them is driven by a "basic affiliative need" (p.3) to communicate and identify with other humans. They described pictorial understanding development in stages that are built on social factors, beginning from infants (12 months of age) who imitate the actions of adults when given pictorial symbols, to toddlers (2–4 years) who use social scaffolding through language and imitation to understand pictures, and finally to school-aged children (4–5 years) who begin to understand not only symbol–referent relations but also the intentions of the symbol creator.

Differences in sociocognitive ability may contribute to some of the differences in pictorial understanding found in ASD (Hartley & Allen, 2014) and may also be affected by a delay in expressive vocabulary (although directionality in LT is difficult to specify). Expressive delay could potentially reduce opportunities to learn from caregivers that verbal labels are used to scaffold picture comprehension and result in LT children having less practice in applying a linguistic strategy. Caregivers of children with expressive language delay have been found to provide less complex recasts (Conti-Ramsden, 1990), provide less lexical and prosodic information (D'Odorico & Jacob, 2006), and produce fewer expansions, less self-directed speech, and less general responses (Vigil, Hodges, & Klee, 2005). Others have found no difference in maternal input but rather that because LT children simply say less, caregivers have less to expand on (Paul & Elwood, 1991). Outcome studies also suggest that there may be social impairments associated with expressive language delay, with some finding lower social competency in LT children (Horwitz et al., 2003; Longobardi, Spataro, Frigerio, & Rescorla, 2016).

Overall, despite sociocognitive skills forming the basis of theoretical accounts of symbolic development, we do not know how individual differences in social ability in LT children might interact with pictorial understanding. It is possible that the impact of expressive language delay on the availability

of social scaffolding, or vice versa, may affect pictorial understanding. Equally, social ability may well compensate for deficits in expressive vocabulary; LT children who are more socially orientated may invite more social scaffolding behavior than those who are not.

The current study

There are three distinct areas in which further research is necessary. First, although symbols form a key part of communication throughout life, and TD children use language before 3 to 4 years of age to scaffold their understanding of pictorial symbols, we do not know how early language delay affects picture comprehension in the absence of ASD. No studies to date have investigated how linguistic scaffolding of pictorial understanding might be affected in LT children, and other research suggests that language delay might be related to differences in symbolic play (Lyytinen, Poikkeus, Laakso, Eklund, & Lyytinen, 2001; Rescorla & Goossens, 1992). Because symbolic play, pictorial understanding, and language are developmentally inter-related (Callaghan & Rankin, 2002; Kirkham et al., 2013), LT children may also exhibit deficits in pictorial understanding.

Second, despite evidence from typical and atypical populations that expressive and receptive vocabulary might have different effects as pictorial understanding develops, very few studies have probed this relation directly. This means that we do not know how emerging language skills interact with pictorial understanding at different ages.

Third, regardless of theoretical literature maintaining that social scaffolding and language are crucial to pictorial understanding, the relationships among individual social ability, language delay, and pictorial understanding have not been directly investigated in populations with isolated expressive language delay.

We addressed these issues by adapting Callaghan's (2000) verbal scaffolding picture comprehension task in a longitudinal study of LT and TD children. We manipulated the availability of verbal labels when asking children to match pictures to real objects and assessed their concurrent language skills at 2.0–2.4 years of age (Timepoint 1 [T1]) and 3.5–3.9 years of age (Timepoint 2 [T2]). We also considered the effect of social ability measured at 2.0–2.4 years.

We hypothesized that LT children would respond less accurately than TD children when linguistic scaffolding is available and on par with TD children in conditions when linguistic scaffolding is inaccessible. We also hypothesized that expressive vocabulary at both T1 and T2 would positively predict picture comprehension accuracy and that receptive vocabulary would be positively correlated with expressive vocabulary. We originally hypothesized that LT children who did not recover (i.e., they failed to reach the TD expressive vocabulary range by the second timepoint at 3.5–3.9 years of age) would perform less accurately with linguistic scaffolding and that LT children who did recover (i.e. they reached the TD expressive range by T2) would perform on par with TD children (see pre-registrations). However, because we could test only half the original sample due to COVID-19, resulting in small subgroups, we used concurrent expressive vocabulary as a continuous variable across all participants at T2.

As an exploratory analysis, we also hypothesized that children with less sophisticated social ability would score lower on picture comprehension accuracy.

Method

Participants

Participants were part of a longitudinal project intended to capture differences between LT and TD children for 18 months between 2.0–2.4 years and 3.5–3.9 years of age. The picture comprehension task was administered at the first and second timepoints.

Participants were recruited using flyers from Lancaster University Babylab, via health visitors in the Lancashire local authority, and from nurseries in the local area. Once consent to contact was obtained, parents completed the Oxford Communicative Development Inventories (CDI; Hamilton, Plunkett, & Schafer, 2000), a parent-reported checklist of words the child says and understands. Children were

included in the study if they met one of the following criteria: TD with productive vocabulary score ≥ 25th percentile or LT with productive vocabulary score ≤ 10th percentile. These criteria were chosen to ensure two distinct groups, with the LT criterion being consistent with prior literature (Fisher, 2017). Inclusion criteria also included monolingual British English with no history of developmental or sensory delays or disorders.

A total of 85 families completed the CDI; of these, 24 were excluded due to the aforementioned criteria. A total of 61 children (40 TD and 21 LT) took part in the study at the first timepoint aged 2.0– 2.4 years (T1); however, 2 TD children did not complete the pictorial understanding task due to fussiness and so were excluded, resulting in a final sample of 59 children (38 TD and 21 LT). At 3.5–3.9 years of age (18 months from baseline; T2), a total of 29 children (20 TD and 9 LT) were tested before the COVID-19 pandemic halted all face-to-face testing.

Questionnaires

Participants completed the Oxford CDI (Hamilton et al., 2000) at consent to contact. Caregivers completed a demographics questionnaire and the Preschool Social Responsiveness Scale-2 (SRS-2; Constantino & Gruber, 2012) at the T1 test visit. The SRS-2 is a caregiver-reported list of behaviors rated by frequency of occurrence that is often used to identify children at risk for ASD. However, because it is designed to discriminate typical and atypical behaviors at a young age and was normed on a well-distributed sample of children aged 30 to 54 months (albeit in the United States), it was used as a measure of individual social proficiency (raw scores) within our sample. The experimenter conducted the Receptive and Expressive One-Word Picture Vocabulary Tests (ROWPVT-4 and EOWPVT-4, respectively; Martin & Brownell, 2011) and the Leiter-3 Non-Verbal IQ 4 subscore scale (Roid, Miller, Pomplun, & Koch, 2013) at the T2 test visit. The ROWPVT-4 and EOWPVT-4 are picture-based vocabulary tests; the ROWPVT-4 is administered by showing four pictures to children and asking them to identify a named picture (e.g., "Which is the rabbit?"), whereas the EOWPVT-4 is administered by showing a single picture to children and asking them to name the picture (e.g., "What's this?"). The Leiter-3 Non-Verbal IQ 4 subscale involves a series of core subtests that measure reasoning, visualization, and problem solving without the need to use verbal instructions, allowing assessment of nonverbal IQ in children with low language ability.

Picture comprehension task

Objects

We adapted Callaghan's (2000) picture comprehension task, using the same criteria for selecting relevant stimuli. There were 32 different objects in total, split into 16 pairs. For each condition, there were 4 trials, 1 trial from each of four groups: animals, natural, household/indoor artifacts, and vehicles (16 trials in total). For the Familiar–Matched Label condition, pairs of familiar objects had the same basic label (e.g., dog) but different subordinate labels (e.g., German Shepherd and Burmese Mountain). For the Unfamiliar–Matched Label condition, pairs of unfamiliar objects had the same basic label (e.g., coral) but different subordinate labels (e.g., elkhorn coral and encrusting coral). For the Familiar–Distinct Label condition, pairs of familiar objects had the same global label (e.g., animal) but different basic labels (e.g., cat and rabbit). For the Unfamiliar–Distinct Label condition, pairs of unfamiliar objects had the same global label (e.g., vehicle) but different basic labels (e.g., quadbike and jet ski).

We ensured that perceptual discriminability of paired objects was similar across trial types and stimuli groups. For sets of animals, different fur colors and poses were chosen (e.g., sitting vs. standing dogs); for artifacts, different colors, materials, and shapes were chosen, and so on. All objects were roughly the same size. Caregivers were consulted prior to participation on their children's familiarity with the test objects, and the age norms for objects were checked using Fenson et al. (1994; familiar objects: M_{age} = 13.92 months, range = 10–16). Example stimuli can be seen in Fig. 1 (see Appendix for all stimuli used).
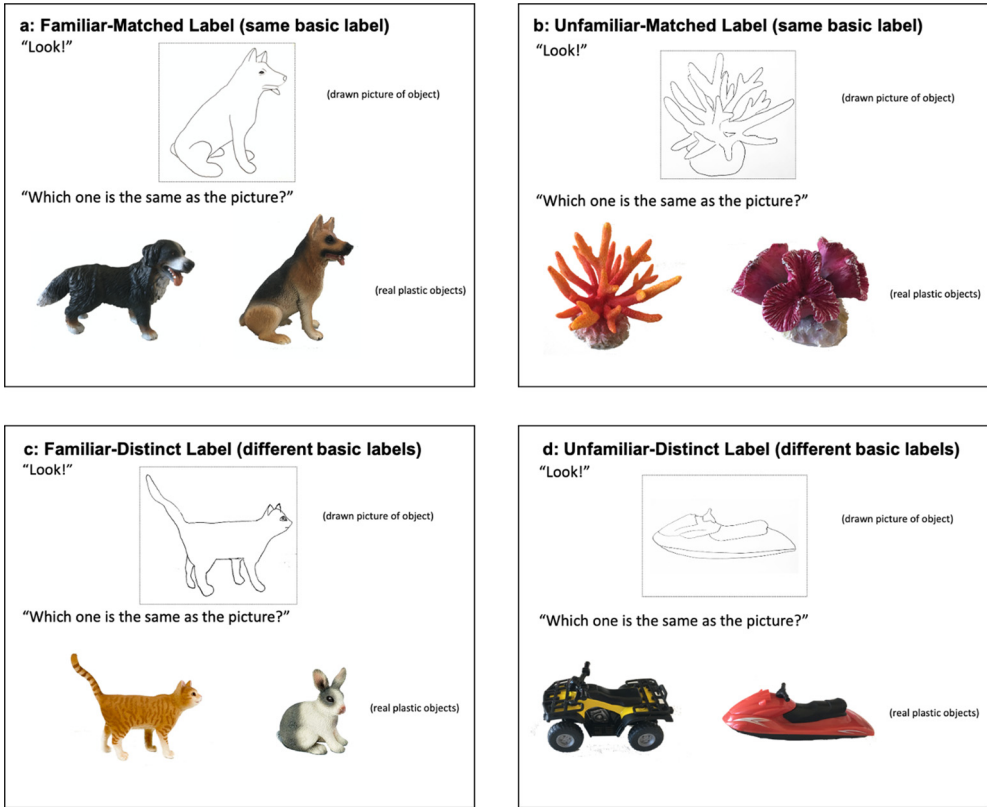
Fig. 1. Example of stimuli and trial types used: (A) Familiar–Matched Label; (B) Unfamiliar–Matched Label; (C) Familiar–Distinct Label; (D) Unfamiliar–Distinct Label.

Pictures

A total of 16 black and white laminated cards were used that had a simple black pen drawing of one familiar or unfamiliar object.

Display

Objects were placed on a tray with a deep lid that had a handle and a cutout at the back that allowed the experimenter to rearrange objects out of sight of children. The objects remained hidden until the experimenter lifted the lid to reveal the two objects sitting on the tray.

Procedure

We adapted Callaghan's (2000) picture comprehension task that manipulates the availability of linguistic scaffolding. We manipulated the label for choice objects across conditions where it could not be used (Matched Label trials; two objects with the same basic label, e.g., two types of dog) and conditions where it could be used (Distinct Label trials; two objects with different basic labels, e.g., rabbit and cat). We also manipulated the familiarity of objects depending on children's knowledge of the labels and objects (Familiar and Unfamiliar) within the Matched Label and Distinct Label trials. The order of trial types was randomized per participant, with no more than 2 trials of the same type presented consecutively.

In Familiar–Distinct Label trials, linguistic scaffolding is possible; if participants generate a label when the target is cued, they can achieve a correct response by identifying the target object based

on its matching label rather than responding based on perceptual similarity alone. However, this strategy is unavailable in the other conditions when both referent objects share the same known label as the picture (Familiar–Matched Label) or labels are unknown (Unfamiliar trials).

The task was administered at T1 and T2. Participants were tested with the same mobile setup for the task either at participants' homes or in a designated room at the Babylab, depending on families' preference. Where visits took place at a home, care was taken to ensure a clear space and a quiet environment with just the experimenter, child, and caregiver present.

During the task, the child and experimenter were sitting on opposite sides of a 1-m-wide, low foldout table. The experimenter held up the relevant trial picture card (e.g., cat) and said "Look!" The picture was presented for 4 s before being removed from view. The experimenter then lifted the lid of the box to reveal the two relevant trial objects, one of which resembled the picture (e.g., cat), and the other of which was a paired foil object (e.g., rabbit). On displaying the objects, the experimenter asked "Which one is the same as the picture?" The trial ended when the child made a response (either by pointing with fingers or palm or picking up the relevant object).

Results

All data and code can be found on the Open Science Framework (OSF) at https://osf.io/ywmx5/?view_only=14f51c730c4c47758893bc684d7cebf5 alongside preregistrations with a document that explains deviations due to the COVID-19 pandemic.

Overview of analyses

We first describe the sample and report descriptive analyses of the data using Welch one-sample *t* tests identifying how TD and LT children performed against chance. We then conducted three analyses to assess our research hypotheses. The first analysis tested the longitudinal predictive effect of LT status over time using generalized linear mixed-effects (GLME) modeling. The second tested whether receptive and expressive vocabulary measures could predict performance cross-sectionally at different ages (T1 and T2) using GLME analyses and a post hoc mediation analysis. The third assessed whether social ability at T1 had any additive predictive value on accuracy at T1 or T2 by comparing GLME model fits with the data, with and without social ability, and by using a post hoc mediation analysis.

Sample

Table 1 contains T1 and T2 final sample demographics, questionnaire, and vocabulary scores. Nearly all families identified as White British, and the majority (92%) had at least one parent with an education level corresponding to an undergraduate university degree or above. Due to the COVID-19 pandemic halting all face-to-face testing, only 29 of the original 59 children were tested at T2. TD and LT children did not differ in SRS-2 scores, $t(46.39) = 1.35$, $p = .183$, or Leiter-3 scores, $t(10.02) = -1.45$, $p = .178$. To be included in the study, children needed to exceed a Leiter-3 cutoff score of 79, which corresponds to the upper boundary of "low" performance associated with the test (Roid et al., 2013); no children were excluded on this basis.

Descriptive analyses

We used Welch's one-sample *t* tests to compare each population's overall picture comprehension accuracy, and accuracy on each trial type, against chance (50%). The data were checked for outliers using the *rstatix* package in R and were normally distributed using the Shapiro–Wilk test.

At the first timepoint (T1), when participants were 2.0 to 2.4 years of age, TD children performed significantly above chance overall ($M = 0.60$, $t(37) = 4.64$, $p < .001$). TD children performed below chance on the Familiar–Matched Label trials but above chance on all other trial types: Familiar–Dis-

Table 1
Means (and standard deviations) for demographic, questionnaire, and vocabulary scores for samples at T1 and T2.

| | T1: 2.0–2.4 years of age (*N* = 59) | | T2: 3.5–3.9 years of age (*N* = 29) | |
	TD (*n* = 38)	LT (*n* = 21)	TD (*n* = 20)	LT (*n* = 9)
Age (years)	2.2 (0.12)	2.2 (0.12)	3.7 (0.12)	3.8 (0.15)
Gender (male:female)	16:22	14:7	7:13	7:2
Receptive vocabulary[a]	384.00 (38.00)	258.00 (93.40)	119.45 (5.35)	111.11 (7.54)
Expressive vocabulary[a]	331.00 (73.20)	60.00 (49.50)	122.75 (8.43)	108.11 (12.90)
Social ability (SRS-2)[b]	27.90 (12.40)	32.10(10.80)		
Nonverbal IQ (Leiter-3)[c]			98.55 (6.68)	92.00 (12.80)

Note. T1, Timepoint 1; T2, Timepoint 2; TD, typically developing; LT, late talking; SRS-2, Preschool Social Responsiveness Scale-2.
[a] T1: Oxford Communicative Development Inventories (raw scores); T2: Receptive/Expressive One-Word Picture Vocabulary Tests (standardized scores).
[b] Higher scores indicate lower responsiveness/ability. Raw scores were used.
[c] Standardized scores were used.

tinct Label (*p* < .001), Unfamiliar–Matched Label (*p* = .007), and Unfamiliar–Distinct Label (*p* = .004) (Fig. 2). The difference between Familiar–Matched Label trials (*M* = 0.42) and Familiar–Distinct Label trials (*M* = 0.72) demonstrated that TD children were able to use verbal labels to scaffold their understanding of pictures and objects. In line with Callaghan (2000), children responded accurately when objects were familiar and had different basic labels, but responded inaccurately when familiar objects shared the same basic label. Performance on the Unfamiliar trial types indicated that when objects were unfamiliar, children were also able to use perceptual similarities between pictures and objects to select the correct object. Thus, not knowing the basic or subordinate label in these conditions was advantageous because it enabled children to use perceptual similarity.

LT children did not perform significantly above chance overall (*M* = 0.53, *t*(20) = 1.44, *p* = .083). They performed below or at chance on Familiar–Matched Label, Unfamiliar–Matched Label, and Unfamiliar–Distinct Label trials (Fig. 2). They performed above chance on Familiar–Distinct Label trials (*M* = 0.59, *p* = .021) and at a level similar to TD children on Familiar–Matched Label trials (*M* = 0.42). This suggests that LT children were able to use linguistic labels to support their picture–object matching when they were available. Scoring below chance when objects were unfamiliar suggested that LT children struggled to match pictures to unfamiliar objects based on perceptual similarities alone.

At the second timepoint (T2), when participants were 3.5 to 3.9 years of age, both TD and LT children performed above chance overall [TD: *M* = 0.80, *t*(19) = 10.93, *p* < .001; LT: *M* = 0.73, *t*(8) = 4.80, *p* < .001] (Fig. 2). Both TD and LT children performed at chance on Familiar–Matched Label trials but significantly above chance in all other trial types. These results indicate that LT children were largely able to use both perceptual information and linguistic labels in the task at T2.

GLME model analyses

All GLME analyses were undertaken with the same procedure. All models predicted child task accuracy as the dependent variable and were built in R (Version 1.1.463) using the *glmer* function in the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015). Models were built up sequentially, adding in one fixed effect at a time and comparing each model with the previous best-fitting model using log likelihood tests. Each model was built up from a null model containing random effects of participant and target. Random slopes of participant per target failed to converge. Where longitudinal data were analyzed, we also attempted to fit a random slope of timepoint per participant, but this failed to converge. To analyze fixed effects of trial type, we coded them as follows: *object familiarity* (Unfamiliar coded as 0 and Familiar coded as 1) and *language scaffolding* (Matched Label coded as 0 and Distinct Label coded as 1). Due to the number of analyses conducted, only results from best-fitting models that found significant effects of variables of interest are reported here. All best-fitting models were tested
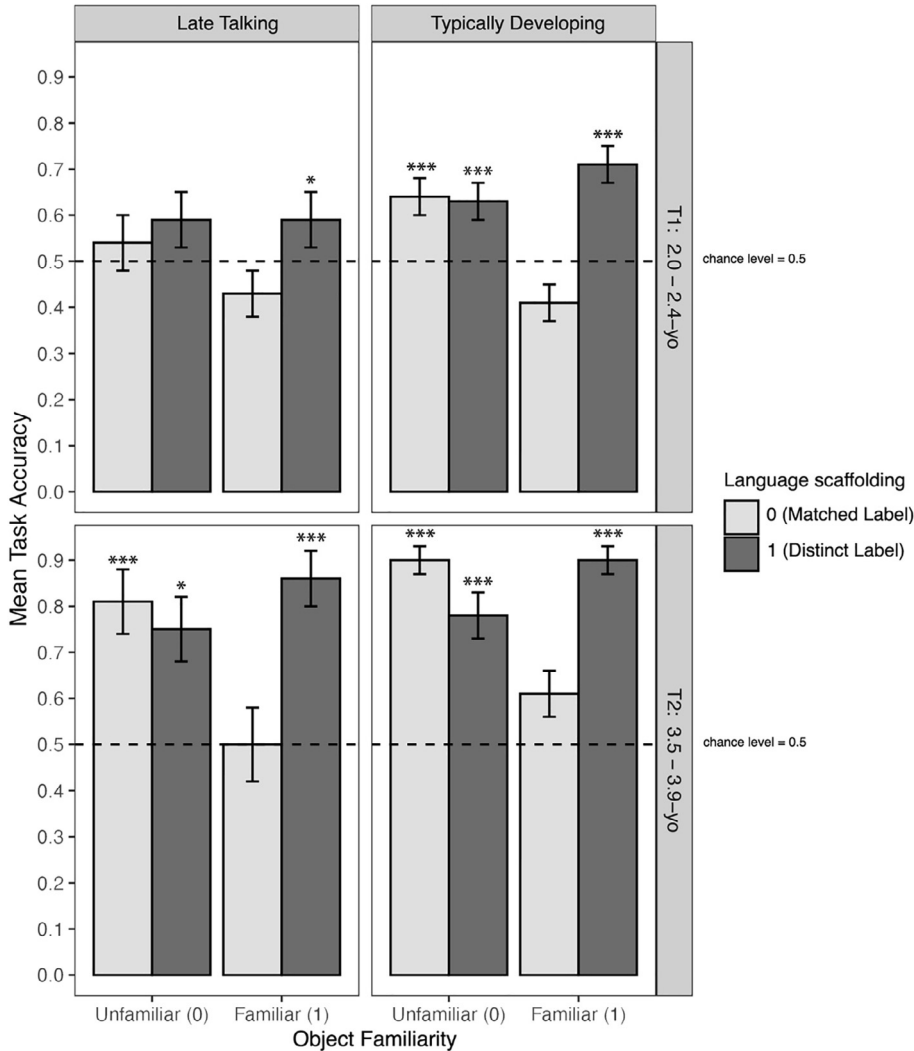
Fig. 2. Mean accuracy and standard errors at test across trial types per group over time. Trial types—Familiar: known objects to the child; Unfamiliar: unknown objects to the child; Matched Label: object pairs with the same global label and same basic label, inhibiting verbal scaffolding; Distinct Label: object pairs with the same global label and different basic labels, allowing verbal scaffolding. T1, Timepoint 1; T2, Timepoint 2; yo, years old. *$p < .05$; **$p < .01$; ***$p < .001$; p values to three decimal places; within-group one-sample Welch t tests against chance (50%).

for normality and overdispersion[1] and can be viewed on the OSF (https://osf.io/ywmx5/?view_only=14f51c730c4c47758893bc684d7cebf5).

All post hoc mediation analyses were undertaken using the *mediation* package in R (Version 1.1.463) (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). For each analysis, 1000 simulations were

---

[1] Due to the disparate scales used for each measure (i.e., accuracy as 0 or 1 and vocabulary as 0–416), in some cases convergence warnings were issued when fitting GLME analyses. Where this occurred, vocabulary measures were scaled by dividing the vocabulary score by 100 so that they were on a closer scale to accuracy. This is indicated in the tables reporting GLME result estimates. See R code on OSF for more details.

used to estimate model effects using the quasi-Bayesian Monte Carlo method (Imai, Keele, & Tingley, 2010).

Does LT status predict symbolic picture comprehension over time?

We conducted a GLME analysis with added fixed effects of population and timepoint to trial type. The best-fitting model to the data contained fixed effects of timepoint, population, language scaffolding, and object familiarity, with an interaction between language scaffolding and object familiarity and random effects of participant and target[model comparison: $\chi^2(3) = 13.51$, $p = .004$] (Table 3).

The pattern of accuracy for each trial type was consistent over both timepoints; relative to Unfamiliar–Matched Label trials where objects were unfamiliar and had the same unknown category label, children performed significantly less accurately on the Familiar–Matched Label trials where the objects were familiar and shared the same known category label ($p < .001$). The significant two-way interaction was caused by a significant difference between trial types involving familiar objects but not unfamiliar objects; participants performed significantly more accurately on Familiar–Distinct Label trials where verbal scaffolding could assist children's mapping of pictures to familiar objects ($p < .001$). Performance was highest on Familiar–Distinct Label trials and least accurate on Familiar–Matched Label trials (Fig. 2), consistent with Callaghan (2000). Children performed similarly to Unfamiliar–Matched Label trials on Unfamiliar–Distinct Label trials where objects were unfamiliar but had different basic category labels ($p = .603$).

The added effect of timepoint indicated that participants performed significantly more accurately at 3.5–3.9 years of age as compared with 2.0–2.4 years of age ($p < .001$), and the effect of population indicated that TD children performed significantly more accurately than LT children when data from both timepoints were combined ($p = .022$).

Thus, the longitudinal analysis indicated that there was a predictive effect of LT status on performance across time, with LT children attaining lower accuracy scores overall when total performance was assessed across both timepoints (Table 2). However, because there were no interactions between trial type and population, the results also suggested that the facilitative effect of linguistic scaffolding was stable across both populations.

How do concurrent receptive and expressive vocabulary contribute to picture comprehension at different ages?

Receptive vocabulary

We conducted three separate GLME analyses to identify the effects of receptive vocabulary on cross-sectional task performance at T1 and T2, collapsing across LT and TD data. For all analyses, fixed effects of trial type were used; only fixed effects of receptive vocabulary differed. When predicting T1 task performance, T1 receptive vocabulary (CDI) was used. When predicting T2 task performance, one model tested the effect of prior T1 receptive vocabulary (CDI) and the other model tested the effect of T2 receptive vocabulary (ROWPVT-4).

At T1, there was an added effect of concurrent receptive vocabulary to that of trial type predicting task performance [model comparison: $\chi^2(2) = 9.14$, $p = .010$] (Table 3). This indicated that children with higher concurrent receptive vocabularies performed significantly more accurately at 2.0 to 2.4 years of age ($p = .038$).

At T2, there was no added predictive effect of concurrent receptive vocabulary (ROWPVT-4) to that of trial type, and no interactions were found. However, prior receptive vocabulary at T1 did predict accuracy in addition to the effect of trial type [model comparison: $\chi^2(3) = 10.11$, $p = .018$] (Table 3), showing that children with higher receptive vocabularies at 2.0–2.4 years of age performed more accurately on the picture comprehension task when they were 3.5–3.9 years of age ($p = .018$).

Expressive vocabulary

We conducted three separate GLME analyses to identify the effects of expressive vocabulary on cross-sectional task performance at T1 and T2, collapsing across LT and TD data. For all analyses, fixed effects of trial type were used; only fixed effects of expressive vocabulary differed. When predicting T1

Table 2

Longitudinal analysis of task accuracy over time: Summary table of best-fitting model from general linear mixed-effects model analyses predicting accuracy over time using fixed effects of trial type (object familiarity or language scaffolding), population (TD or LT), and timepoint.

Fixed effect	Estimate	SE	z value	p value
(Intercept)[a]	0.34	0.20	1.67	.094
Familiar	−0.97	0.24	−3.96	<.001
Distinct Label	−0.13	0.25	−0.52	.603
Familiar * Distinct Label	1.34	0.35	3.82	<.001
Timepoint (T2: 3.5–3.9 years of age)	0.99	0.14	7.04	<.001
Population (TD)	0.34	0.15	2.28	.022

Note. TD, typically developing; LT, late talking; T2, Timepoint 2.

[a] Intercept corresponds to Matched Label (no language scaffolding; 0) and Unfamiliar (object unknown to child; 0), population LT, and Timepoint 1 (2.0–2.4 years of age).

Table 3

Cross-sectional analyses of predictive effect of receptive vocabulary on task accuracy: Summary of best-fitting model from general linear mixed-effects model analyses predicting T1 and T2 accuracy with fixed effects of trial type (object familiarity or language scaffolding) and T1 receptive vocabulary.

T1: 2.0–2.4 years of age				
Fixed effect	Estimate	SE	z value	p value
(Intercept)[a]	−0.13	0.31	−0.42	.674
Familiar	−0.75	0.21	−3.51	<.001
Distinct Label	0.06	0.21	0.28	.776
Familiar * Distinct Label	0.97	0.30	3.24	.001
T1 receptive vocabulary (CDI)[b]	0.16	0.08	2.07	.038
T2: 3.5–3.9 years of age				
Fixed effect	Estimate	SE	z value	p value
(Intercept)[a]	0.12	0.96	0.12	.904
Familiar	−1.78	0.62	−2.87	.004
Distinct Label	−0.76	0.63	−1.20	.229
Familiar * Distinct Label	2.84	0.90	3.14	.002
T1 receptive vocabulary (CDI)[b]	0.57	0.24	2.37	.018

Note. T1, Timepoint 1; T2, Timepoint 2; CDI, Oxford Communicative Development Inventories; ROWPVT-4, Receptive One-Word Picture Vocabulary Test.

[a] Intercept corresponds to Matched Labels (no language scaffolding; 0) and Unfamiliar (object unknown to child; 0).

[b] Rescaled using $x/100$ to allow model fit. Standardized scores were used.

Table 4

Cross-sectional analyses of predictive effect of expressive vocabulary on task accuracy: Summary of best-fitting model from general linear mixed-effects model analyses predicting T2 accuracy with fixed effects of trial type (object familiarity or language scaffolding) and T2 expressive vocabulary.

T2: 3.5–3.9 years of age				
Fixed effect	Estimate	SE	z value	p value
(Intercept)[a]	−2.93	1.56	−1.88	.060
Familiar	−1.79	0.62	−2.87	.004
Distinct Label	−0.76	0.63	−1.20	.229
Familiar * Distinct Label	2.85	0.91	3.14	.002
T2 expressive vocabulary (EOWPVT-4)[b]	0.04	0.01	3.32	<.001

Note. T2, Timepoint 2; EOWPVT-4, Expressive One-Word Picture Vocabulary Test.

[a] Intercept corresponds to Matched Label (no language scaffolding; 0) and Unfamiliar (object unknown to child; 0).

[b] Standardized scores were used.

task performance, T1 population (TD vs. LT) was used. When predicting T2 task performance, one model tested the effect of T1 population and the other model tested the effect of T2 expressive vocabulary (EOWPVT-4).

At T1, the GLME analysis did not find a predictive effect of population above that of trial type, and no interactions were found. The lack of a population effect suggested that at 2.0-2.4 years of age, expressive vocabulary was not predictive of pictorial understanding performance.

At T2, population at T1 did not predict accuracy. However, T2 expressive vocabulary (EOWPVT-4) did predict accuracy in addition to the effect of trial type [model comparison: $\chi^2(3) = 10.12$, $p = .018$] (Table 4). The best fitting model to the data demonstrated that as children's concurrent expressive vocabulary at 3.5–3.9 years of age increased, so did their picture comprehension accuracy ($p < .001$).

Relationship between receptive and expressive vocabulary in predicting task accuracy

The cross-sectional analyses indicated that early receptive vocabulary at ~2.0 years of age predicted both concurrent and later task accuracy at ~3.5 years of age and that later expressive vocabulary at ~3.5 years of age predicted concurrent task accuracy at ~3.5 years of age.

To tease apart the relative contribution of T1 receptive vocabulary and T2 expressive vocabulary to T2 task accuracy, we conducted a further post hoc mediation analysis (Fig. 3). The effect of T1 receptive vocabulary on T2 picture task accuracy was significantly mediated through T2 expressive vocabulary (*average casual mediation effects*: 0.07, *95% confidence interval (CI)*: [0.01, 0.12], *p = .016*). The results indicated that of the estimated increase in probability of task accuracy at ~3.5 years of age (*total effect*: 0.10) due to earlier receptive vocabulary at 2.0–2.4 years of age, 0.07 was estimated to be mediated through later expressive vocabulary at 3.5–3.9 years and 0.03 was estimated to be from earlier receptive vocabulary at 2.0–2.4 years.

Is the differential effect of expressive and receptive language in picture comprehension tasks mediated by social ability?

To test whether there was any effect of T1 social ability on task accuracy, we fitted an additional GLME model with SRS-2 as an additional fixed effect and compared it with the original best-fitting model for each timepoint.

For T1, adding SRS-2 to the best-fitting model with fixed effects of trial type and T1 receptive vocabulary was beneficial. Adding SRS-2 yielded a better fit to the data than a model without SRS-2 [model comparison: $\chi^2(1) = 5.40$, $p = .020$] (Table 5), suggesting that children with less social responsiveness were less accurate at matching pictures to symbolized objects ($p = .023$) regardless of language ability.

For T2, a GLME model with SRS-2 score as an additional fixed effect was not a better fit to the data when compared with the original models.

We conducted a post hoc mediation analysis to assess whether the effect of T1 receptive vocabulary on T1 task accuracy was mediated through concurrent T1 social ability (Fig. 4). This demonstrated a significant mediating effect of social ability (*average casual mediation effects*: 0.02, *95% CI*: [0.002, 0.03], *p = .020*). The results indicated that of the estimated increase in probability of task accuracy at 2.0–2.4 years of age (*total effect*: 0.04) due to concurrent receptive vocabulary, 0.02 was estimated to be mediated through concurrent social responsiveness and 0.02 was estimated to be from concurrent receptive vocabulary.

Discussion

Developmental theories propose that language scaffolds children's acquisition and understanding of the pictorial symbol system (Callaghan, 2000; Tomasello, 2003, 2010). Our results indicate not only that language ability affects the developmental trajectory of picture comprehension but also that receptive and expressive skills may differ in their contribution at different ages, subject to mediating effects of social ability.
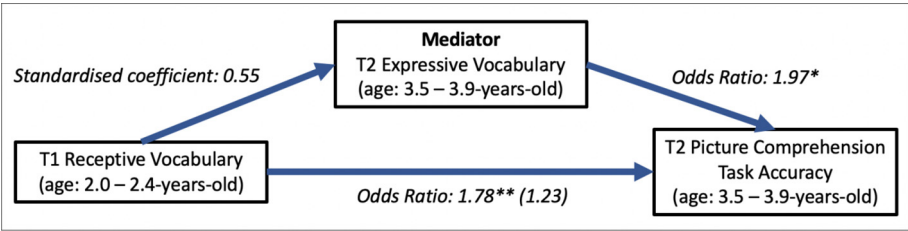
Fig. 3. Results of mediation analysis assessing indirect effect of Timepoint 1 (T1) receptive vocabulary on Timepoint 2 (T2) task accuracy through T2 expressive vocabulary. The value in parentheses indicates the direct effect of receptive vocabulary when the mediator is included. $^{*}p < .05$; $^{**}p < .01$.

Table 5

Cross-sectional analyses of added effect of social ability to predicting task accuracy: Summary of best-fitting model from general linear mixed-effects model analyses predicting T1 accuracy with fixed effects of trial type (object familiarity or language scaffolding), T1 receptive vocabulary, and T1 social ability.

T1: 2.0–2.4 years of age				
Fixed effect	Estimate	SE	z value	p value
(Intercept)[a]	0.54	0.42	1.28	.200
Familiar	−0.75	0.21	−3.54	<.001
Distinct Label	0.06	0.21	0.32	.790
Familiar * Distinct Label	0.98	0.30	3.14	.001
T1 receptive vocabulary (CDI)[b]	0.09	0.09	1.11	.265
T1 social ability (SRS-2)[c]	−1.45	0.63	−2.27	.020

Note. T1, Timepoint 1; CDI, Oxford Communicative Development Inventories; SRS-2, Preschool Social Responsiveness Scale-2.
  [a] Intercept corresponds to Matched Label (no language scaffolding; 0) and Unfamiliar (objects unknown to child; 0).
  [b] Standardized scores were used.
  [c] Rescaled using *x*/100 to allow model fit. Higher scores indicate less social responsiveness. Raw scores were used.
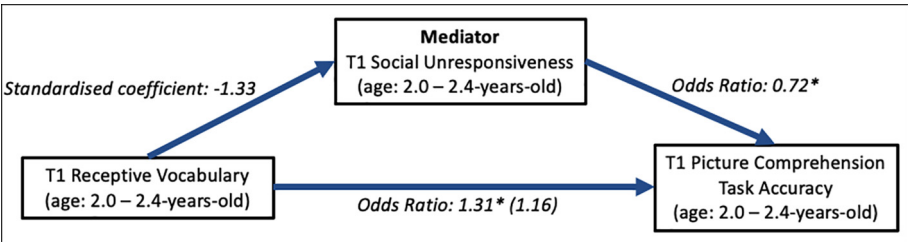

Fig. 4. Results of mediation analysis assessing indirect effect of Timepoint 1 (T1) receptive vocabulary on Timepoint 1 task accuracy through T1 social ability. Note that the Preschool Social Responsiveness Scale-2 is scored such that higher scores indicate lower ability and that the value in parentheses indicates the direct effect of receptive vocabulary when the mediator is included. $^{*}p < .05$.

The use of linguistic scaffolding in the picture comprehension task requires children to generate labels (albeit subvocally). When viewing the picture, children can either generate a label for the depicted object internally or store its visual features if the label is unknown and then use that information to match the picture to the referent object. There are two opportunities to generate a label: when the target is cued (e.g., a picture of a cat) and when the target object is selected (e.g., a plastic cat and a plastic rabbit on the tray). At an earlier age, receptive vocabulary skills might enable children to understand the task and, to some extent, use linguistic information to activate associated concepts that can be used to help scaffold picture comprehension. However, being more proficient in expressive vocabulary may facilitate children's ability to explicitly generate the label internally and activate asso-

ciated concepts both when the target is cued and when the object is selected and thus directly use that linguistic information to select the correct object.

More generally, our results suggest that at an earlier age children may rely more on understanding linguistic information and concurrent social ability, but at a later age maturing expressive vocabulary skills may become increasingly important in scaffolding picture comprehension. However, as our mediation analyses highlighted, receptive and expressive vocabulary skills and their influence on developing symbolic skills are tightly interwoven. We now outline the implications of these results for LT children, children's development more generally, and future considerations.

Implications for LT children

At both timepoints, LT children scored lower than TD children on the picture comprehension task. This was reflected in the longitudinal analysis that showed a general effect of population on task accuracy across both timepoints. One possibility is that the smaller expressive vocabularies of LT children might have meant they were less able to retrieve the words (e.g., "cat") and subsequent representations of the real object (e.g., the concept of a cat) when seeing the picture (e.g., a picture of a cat), resulting in more errors when identifying the depicted object. Similarly, Rescorla and Goossens (1992) suggested that reduced symbolic play in LT toddlers might be secondary to less fluent and less spontaneous retrieval and encoding of lexical entries for semantic representations across both referents and play scripts. However, because no significant effects of population were found cross-sectionally, any differences between the populations in our study were subtle. Furthermore, because there was no significant interaction between population or vocabulary measures and trial type in either the longitudinal or cross-sectional analyses, this suggests that the developmental trajectory for picture comprehension in LT children is not atypical—but rather, is delayed. The results also indicated that the effect of language in scaffolding pictures is stable even in early expressive language delay.

These findings are in line with outcome studies in LT children showing that the majority of children reach the same range as TD children in language skills by school age but fall on the lower end of this range (Rescorla, 2002, 2005). The predictive effect of receptive vocabulary at 2.0–2.4 years of age on picture comprehension in our study was also consistent with early receptive vocabulary being a better predictor of later outcomes than early expressive vocabulary in LT children (Fisher, 2017). Overall, although expressive language mediates linguistic scaffolding of picture comprehension at an older age, categorizing our participants using a dichotomous variable at an earlier age did not accurately represent the fine-grained detail contained in our sample as they grew older. This is also consistent with prominent theories suggesting that language ability, in LT children and DLD, falls on a spectrum (Bishop, 2017; Leonard, 2014).

We did not find any differences in SRS-2 scores in LT and TD children, indicating that early expressive language delay in our sample did not appear to coincide with reduced social proficiency. However, we did find that social ability at 2.0–2.4 years of age predicted task performance at the same age across the whole sample. The implications of this in typical development are discussed below, but of note is that social ability may actually help to mitigate delays that occur alongside, or as a result of, expressive language deficits. This adds to the evidence base for interventions for LT children that make use of social scaffolding to improve language outcomes (e.g., Alt, Meyers, Oglivie, Nicholas, & Arizmendi, 2014; Cable & Domsch, 2010; Robertson & Weismer, 1999). More prosocial toddlers may benefit from social scaffolding during interactions involving pictures at an early age even if their expressive vocabulary is less well developed. Children with higher social skills may also receive more exposure to pictures, and thus more exposure to adults labeling pictures, accelerating their acquisition of a linguistic strategy in pictorial understanding. It is also possible that this cycle may occur in the reverse; higher levels of parental input around pictures may encourage prosocial behavior in toddlers, leading to more opportunities for picture exposure and thus more opportunities for toddlers to use social scaffolding. A third possibility is that both children's social ability and parental input influence each other in a positive feedback loop. In any case, social scaffolding may be a beneficial strategy for future interventions targeting picture comprehension.

General implications for linguistic scaffolding and picture comprehension

Our findings also have several implications for how children comprehend pictures in both LT and TD populations. Across both timepoints, children struggled most with the Familiar–Matched Label trials where access to linguistic scaffolding was blocked. In theory, linguistic scaffolding was only available in Familiar–Distinct Label trials, and selecting the correct referent object in all other trial types required children to attend to the pictures' perceptual features. Although children were able to apply this strategy successfully in trials involving unfamiliar objects, the consistently lower performance in Familiar–Matched Label trials in both TD and LT children suggests that generating labels for pictures might not always be a beneficial strategy; the familiar linguistic label in these trials (e.g., "dog" when there are two types of dog to choose from) seemingly impeded comprehension of the picture based on perceptual resemblance. Callaghan (2000) suggests that when TD children are faced with unfamiliar objects with unknown labels (e.g., jetski vs. quadbike), they may use an attribute word (e.g., "wheels") to help scaffold performance alongside perceptual similarity. If the use of attribute words such as "wheels" for the distinction between "jetski" and "quadbike" is a strategy applied by TD children, then LT children may also have greater difficulty in comprehending unfamiliar pictures due to their comparatively limited vocabulary.

In both samples, children's accuracy on Unfamiliar trial types improved over time, indicating their developing ability to quickly encode mental representations of perceptual features when determining picture–object relationships. This improved accuracy over time may reflect more flexible cognitive strategies toward picture comprehension as children age; whereas linguistic scaffolding may be the default strategy, children showed the ability to adapt over time by using perceptual features as well. It is also possible that the lower performance of LT children overall in picture comprehension across all trials reflects a lack of flexibility or competency in applying different strategies. This would be consistent with word learning literature suggesting that LT children show less flexibility in adopting new strategies for learning novel words than TD children (e.g., Rice, Oetting, Marquis, Bode, & Pae, 1994; Stokes, Kern, & Dos Santos, 2012).

The function that language plays in aiding pictorial understanding may be in creating "cognitive distance" (Homer & Nelson, 2009, p. 132). By enabling children to treat pictures as distinct from real objects through labels, the salience of the picture itself as an object is reduced and its status as a symbolic representation is increased. This abstraction afforded by language is also found in category learning (Waxman & Markow, 1995). Children's language ability predicted performance across all trial types in our study, including those that relied on perceptual discrimination, indicating a robust relationship between pictorial understanding and language domains.

The ability to use language in this manner may depend on where in the trajectory of symbolic understanding children are located. At an earlier age, performance in the picture comprehension task was not dependent on being able to talk about pictures; rather, it was dependent on language comprehension ability and social ability. Social ability both predicted task performance at 2.0–2.4 years of age and mediated the effect of receptive vocabulary on task performance. The lack of interaction between condition and receptive vocabulary also suggests not only that language scaffolds picture comprehension—as evidenced by the highest accuracy scores being in Familiar–Distinct Label trials—but also that receptive vocabulary alongside social ability may mediate pictorial understanding more generally.

These results are consistent with a sociocognitive framework of symbolic understanding, where children at an earlier age rely more heavily on social scaffolding to interact with the world than children at later ages (Callaghan, Rochat, MacGillivray, & MacLellan, 2004). Striano, Tomasello, and Rochat (2001) found that when given uninteresting or ambiguous objects (e.g., a stapler), 2- and 3-year-olds did not perform symbolic actions spontaneously and largely declined to play at all without an experimenter modeling symbolic actions or actively engaging the children. However, 4-year-olds were better able to play with the items independently. In a longitudinal study, Callaghan and Rankin (2002) also found that symbolic scaffolding, consisting of explicitly highlighting the relationship between objects and pictures, improved children's picture comprehension and production in 28-month-olds.

Our results also indicate that children may be more vulnerable to interference of pictorial understanding when faced with more social difficulties early on, although none of the children in our sample reached clinical levels of impairment using the SRS-2. Rather, the results reflected individual differences in social proficiency. Future studies that examine clinical levels of social impairment and dual representation tasks in populations that are otherwise typical, or that manipulate social cues directly within the task, will help to elucidate these mechanisms.

At an older age, expressive vocabulary, rather than receptive vocabulary, predicted children's picture comprehension. This may reflect the shifting role of expressive vocabulary in facilitating symbolic understanding more generally at an older age. At ~3.5 years, the ability to actively talk about symbols and partake in social discourse, such as by asking questions and explicitly inviting caregiver expansions about pictures, may help children to understand how symbols depict real items but are not the same as the real item in itself (i.e. aiding children to understand dual representation; Deloache, 2004).

Similarly, Tomasello and colleagues (Rakoczy et al., 2005; Tomasello et al., 2005) described language as a means through which children are able to develop other symbolic functions such as pretend play. With advancing linguistic ability from 3 to 4 years of age, children are able to engage in meta-representational discourse—and it is this use of expressive discourse that affords them an appropriate vehicle to interpret mental states and broader symbols as referring to real-world concepts and objects. Nelson (2007) also describes an approach where children's external representations of meaning advance from nonintentional imitation of meaning as infants (e.g., copying gestures or early words) to intentional representation and sharing of meaning as school-aged children (e.g., using conventional symbolic systems like discourse). This process is facilitated by externalization of meaning within a social system such as by using words and gestures with caregivers.

Overall, our results indicate not only that pictorial understanding and language ability are developmentally inter-related but also that the importance of receptive and expressive vocabulary ability may be weighted differently as children develop symbolic understanding within a social context.

Limitations and future directions

There are a number of considerations that limit our findings. Our study was restricted by smaller sample sizes at T2 as a result of the COVID-19 pandemic. Because our task was designed to test children's understanding of symbolic relations between pictures and three-dimensional objects, data collection could not be completed online given that perceiving all stimuli via a two-dimensional screen would fundamentally change the nature of the task (Troseth & DeLoache, 1998). Thus, when face-to-face testing resumes, future directions include testing a larger sample.

We also did not have IQ data for the whole sample due to the interruption of testing; the Leiter-3 was to be collected at the oldest timepoint due to the increasing stability of IQ constructs with age (Gottfried, Gottfried, & Guerin, 2009; Schneider, Niklas, & Schmiedeler, 2014). However, the data we do have indicated no significant differences between populations. Furthermore, a mismatch between verbal and non-verbal ability is not sufficient evidence for diagnosis of DLD (Bishop, 2017), and nonverbal IQ may not predict symbolic or pictorial understanding (Kirkham et al., 2013). However, it is possible that we did not fully account for individual differences in attention and executive functioning.

Our sample also consisted of families with similarly high parental education levels and high interest in participating in developmental research. Consequently, although we can be confident that any differences between children in our samples were unlikely to be due to socioeconomic or environmental causes, we cannot extend these findings to populations with contrasting demographic characteristics without further testing. Furthermore, the use of pictures and symbols is subject to cultural differences; for example, Western cultures adopt a different pedagogical approach that entails more social scaffolding around pictorial understanding than non-Western cultures (Callaghan et al., 2011). Thus, our findings are applicable to a specific population where pictures and language have a privileged position in dual representation and broader symbolic understanding.

We also used a parent-report measure for vocabulary at T1 rather than an experimenter-administered measure. This may limit the comparison of timepoints because different measures were used to test vocabulary. However, because we used two distinct cutoffs for the two groups, it is unlikely that parent-report measures were so inaccurate as to incorrectly characterize group status at T1. Furthermore, the CDI can capture a broader assessment of how children use language in their everyday lives during the earlier stages of language development.

Of further note is that the ROWPVT-4 and EOWPVT-4 require picture recognition, whereas the CDI does not. Although both the picture comprehension task and the picture vocabulary task involve the use of pictures, it is important to note that picture vocabulary tests use picture *recognition* (i.e., identification of what a picture looks like without understanding the picture–referent representational relationship), a skill acquired very early in development that relies on perceptual ability (DeLoache, 2004), whereas our task tested the rather more advanced ability of picture *comprehension* (i.e., understanding the representational function of a picture and relating it to a specific real-world referent) as a type of symbolic understanding. Therefore, it is unlikely that picture vocabulary tests are truly tests of symbolic understanding that result in significant overlap with our picture comprehension task.

A further limitation is that the ROWPVT-4 and EOWPVT-4 were standardized on a U.S. population, whereas the Oxford CDI was normed on a U.K. population. This may mean that there are some cultural discrepancies between the two measures (Hamilton et al., 2000).

Conclusions

Our study has implications for both TD and LT children. Through a longitudinal study, we demonstrate, first, that LT children show evidence of less accurate picture comprehension skills over time when compared with a TD sample and, second, that these differences are subtle and subject to effects of participant heterogeneity. Thus, LT (in line with DLD) and its effects on pictorial understanding may be best considered on a dimensional scale rather than on a categorical one. Crucially, because the trajectory of development for LT children resembled that of earlier typical development, albeit developmentally delayed, this suggests that a significant early deficit in expressive language does not appear to cause any qualitative differences between domains; language still appears to be an important mediating factor across groups and ages. Thus, language appears to scaffold pictorial understanding not only in typical development but also in early expressive language delay.

We also demonstrate that the relationship between language and picture comprehension may be partly explained by differences in how receptive and expressive language ability help to scaffold picture comprehension over time, with receptive vocabulary predicting picture comprehension at 2.0 years of age and expressive vocabulary predicting picture comprehension at 3.5 years of age. This differential weighting may be secondary to the interplay of symbolic understanding and language with social ability and social scaffolding. At an earlier age children may rely on social scaffolding and language comprehension skills to understand pictures, but at an older age this may be superseded by the ability to talk about pictures to others. Overall, these findings advance understanding of both atypical and typical development and demonstrate how language ability, social ability, and pictorial understanding may inter-relate over time.

Acknowledgments

Appendix A. Stimuli used in picture comprehension task

Category	Control Familiar	Control Unfamiliar	Standard Familiar	Standard Unfamiliar
Animal	German shepherd (sitting) + Bernese mountain (standing)	Gray octopus (tentacles spread out, bulbous head) + white octopus (tentacles close together, long head)	Ginger cat (standing) + black and white rabbit (sitting)	Manatee (lying down) + narwhal (diving arc)
Vehicle	Red/orange boxed hatchback car + pink/purple flat sports car	Green short rounded submarine + yellow long box submarine	Yellow bicycle + red motorbike	Yellow quadbike + red jetski
Natural	White jagged limestone + black smooth pebble	Orange elkhorn coral + purple encrusting coral	Carrot + banana	Dragonfruit + arti choke
Household/ indoor artifacts	White porcelain Chinese spoon + silver metal Western spoon	Metal garlic press + wooden garlic press	Black hairbrush + wooden comb	Binoculars + safety science glasses

References

Alt, M., Meyers, C., Oglivie, T., Nicholas, K., & Arizmendi, G. (2014). Cross-situational statistically based word learning intervention for late-talking toddlers. *Journal of Communication Disorders, 52*, 207–220.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01.

Bishop, D. V. M. (2017). Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *International Journal of Language & Communication Disorders, 52*(6), 671–680.

Cable, A. L., & Domsch, C. (2010). Systematic review of the literature on the treatment of children with late language emergence. *International Journal of Language & Communication Disorders, 46*, 138–154.

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., & Tomasello, M. (2011). Early social cognition in three cultural contexts. *Monographs of the Society for Research in Child Development, 76*(2, Serial No. 299).

Callaghan, T. C. (2000). Factors affecting children's graphic symbol use in the third year: Language similarity and iconicity. *Cognitive Development, 15*(2), 185–214.

Callaghan, T. (2020). The origins and development of a symbolic mind: The case of pictorial symbols. *Interchange, 51*(1), 53–64.

Callaghan, T. C., & Corbit, J. (2015). The development of symbolic representation. In R. M. Lerner (Ed.). *Handbook of child psychology and developmental science* (Vol. 2, pp. 1–46). Hoboken, NJ: John Wiley.

Callaghan, T. C., & Rankin, M. P. (2002). Emergence of graphic symbol functioning and the question of domain specificity: A longitudinal training study. *Child Development, 73*(2), 359–376.

Callaghan, T. C., Rochat, P., MacGillivray, T., & MacLellan, C. (2004). Modeling referential actions in 6- to 18-month-old infants: A precursor to symbolic understanding. *Child Development, 75*(6), 1733–1744.

Collisson, B. A., Graham, S. A., Preston, J. L., Rose, M. S., McDonald, S., & Tough, S. (2016). Risk and protective factors for late talking: An epidemiologic investigation. *Journal of Pediatrics, 172*, 168–174.e1.

Constantino, J. N., & Gruber, C. P. (2012). *Social Responsiveness Scale 2 (SRS-2)*. Los Angeles: Western Psychological Services.

Conti-Ramsden, G. (1990). Maternal recasts and other contingent replies to language-impaired children. *Journal of Speech and Hearing Disorders, 55*(2), 262–274.

DeLoache, J. S. (1995). Early understanding and use of symbols: The model model. *Current Directions in Psychological Science, 4*(4), 109–113.

DeLoache, J. S. (2004). Becoming symbol-minded. *Trends in Cognitive Sciences, 8*(2), 66–70.

DeLoache, J. S., Pierroutsakos, S. L., Uttal, D. H., Rosengren, K. S., & Gottlieb, A. (1998). Grasping the nature of pictures. *Psychological Science, 9*(3), 205–210.

D'Odorico, L., & Jacob, V. (2006). Prosodic and lexical aspects of maternal linguistic input to late-talking toddlers. *International Journal of Language & Communication Disorders, 41*(3), 293–311.

Domsch, C., Richels, C., Saldana, M., Coleman, C., Wimberly, C., & Maxwell, L. (2012). Narrative skill and syntactic complexity in school-age children with and without late language emergence. *International Journal of Language & Communication Disorders, 47*(2), 197–207.

Eigsti, I.-M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders, 5*(2), 681–691.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5, Serial No. 242).

Fisher, E. L. (2017). A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *Journal of Speech, Language, and Hearing Research, 60*(10), 2935–2948.

Ganea, P. A., Allen, M. L., Butler, L., Carey, S., & DeLoache, J. S. (2009). Toddlers' referential understanding of pictures. *Journal of Experimental Child Psychology, 104*(3), 283–295.

Gottfried, A. W., Gottfried, A. E., & Guerin, D. W. (2009). Issues in early prediction and identification of intellectual giftedness. In F. D. Horowitz, R. F. Subotnik, & D. J. Matthews (Eds.), *The development of giftedness and talent across the life span* (pp. 43–56). Washington, DC: American Psychological Association.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language, 27*(3), 689–705.

Hartley, C., & Allen, M. L. (2014). Intentions vs. resemblance: Understanding pictures in typical development and autism. *Cognition, 131*(1), 44–59.

Hartley, C., & Allen, M. L. (2015). Iconicity influences how effectively minimally verbal children with autism and ability-matched typically developing children use pictures as symbols in a search task. *Autism, 19*(5), 570–579.

Hartley, C., & Allen, M. L. (2015). Symbolic understanding of pictures in low-functioning children with autism: The effects of iconicity and naming. *Journal of Autism & Developmental Disorders, 45*(1), 15–30.

Hartley, C., Trainer, A., & Allen, M. L. (2019). Investigating the relationship between language and picture understanding in children with autism spectrum disorder. *Autism, 23*(1), 187–198.

Henrichs, J., Rescorla, L., Schenk, J. J., Schmidt, H. G., Jaddoe, V. W. V., Hofman, A., ... Tiemeier, H. (2011). Examining continuity of early expressive vocabulary development: The Generation R study. *Journal of Speech, Language, and Hearing Research, 54*(3), 854–869.

Homer, B. D., & Nelson, K. (2009). Naming facilitates young children's understanding of scale models: Language and the development of symbolic understanding. *Journal of Cognition and Development, 10*(1-2), 115–134.

Horwitz, S. M., Irwin, J. R., Briggs-Gowan, M. J., Bosson Heenan, J. M., Mendoza, J., & Carter, A. S. (2003). Language delay in a community cohort of young children. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*(8), 932–940.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15*(4), 309–334.

Kirkham, J., Stewart, A., & Kidd, E. (2013). Concurrent and longitudinal relationships between development in graphic, language and symbolic play domains from the fourth to the fifth year. *Infant and Child Development, 22*(3), 297–319.

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

Longobardi, E., Spataro, P., Frigerio, A., & Rescorla, L. (2016). Language and social competence in typically developing children and late talkers between 18 and 35 months of age. *Early Child Development and Care, 186*(3), 436–452.

Lyytinen, P., Poikkeus, A.-M., Laakso, M.-L., Eklund, K., & Lyytinen, H. (2001). Language development and symbolic play in children with and without familial risk for dyslexia. *Journal of Speech, Language, and Hearing Research, 44*(4), 873–885.

Martin, N., & Brownell, R. (2011). *Expressive and Receptive One-Word Picture Vocabulary Tests-4 (EOWPVT-4/ROWPVT-4)* (4th ed.). Novato, CA: Academic Therapy Publications.

Nelson, K. (2007). *Young minds in social worlds: Experience, meaning, and memory*. Cambridge, MA: Harvard University Press.

Paul, R., & Elwood, T. J. (1991). Maternal linguistic input to toddlers with slow expressive language development. *Journal of Speech and Hearing Research, 34*(5), 982–988.

Pierroutsakos, S. L., & DeLoache, J. S. (2003). Infants' manual exploration of pictorial objects varying in realism. *Infancy, 4*, 141–156.

Preissler, M. A. (2008). Associative learning of pictures and words by low-functioning children with autism. *Autism, 12*(3), 231–248.

Preissler, M. A., & Bloom, P. (2007). Two-year-olds appreciate the dual nature of pictures. *Psychological Science, 18*(1), 1–2.

Preissler, M. A., & Carey, S. (2004). Do both pictures and words function as symbols for 18- and 24-month-old children? *Journal of Cognition and Development, 5*(2), 185–212.

Quinn, S., Donnelly, S., & Kidd, E. (2018). The relationship between symbolic play and language acquisition: A meta-analytic review. *Developmental Review, 49*, 121–135.

Rakoczy, H., Tomasello, M., & Striano, T. (2005). How children turn objects into symbols: A cultural learning account. In L. Namy (Ed.), *Symbolic use and symbolic representation: Developmental and comparative perspectives* (pp. 69–97). New York: Psychology Press.

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., ... Bretherton, L. (2010). Predicting language outcomes at 4 years of age: Findings from Early Language in Victoria Study. *Pediatrics, 126*(6), e1530–e1537.

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 45*(2), 360–371.

Rescorla, L. (2005). Age 13 language and reading outcomes in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 48*(2), 459–472.

Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews, 17*, 141–150.

Rescorla, L., Dahlsgaard, K., & Roberts, J. (2000). Late-talking toddlers: MLU and IPSyn outcomes at 3;0 and 4;0. *Journal of Child Language, 27*(3), 643–664.

Rescorla, L., & Goossens, M. (1992). Symbolic play development in toddlers with expressive specific language impairment (SLI-E). *Journal of Speech and Hearing Research, 35*(6), 1290–1302.

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 37*(1), 106–122.

Rice, M. L., Taylor, C. L., & Zubrick, S. R. (2008). Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *Journal of Speech, Language, and Hearing Research, 51*(2), 394–407.

Robertson, S. B., & Weismer, S. E. (1999). Effects of treatment on linguistic and social skills in toddlers with delayed language development. *Journal of Speech, Language, and Hearing Research, 42*(5), 1234–1248.

Rochat, P., & Callaghan, T. (2005). What drives symbolic development? The case of pictorial comprehension and production. In L. Namy (Ed.), *Symbolic use and symbolic representation: Developmental and comparative perspectives* (pp. 25–46). New York: Psychology Press.

Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (2013). *Leiter-3: Leiter international performance scale* (3rd ed.). Wood Dale, IL: Stoelting.

Roulstone, S., Loader, S., Northstone, K., & Beveridge, M. (2002). The speech and language of children aged 25 months: Descriptive data from the Avon Longitudinal Study of Parents and Children. *Early Child Development and Care, 172*(3), 259–268.

Schneider, W., Niklas, F., & Schmiedeler, S. (2014). Intellectual development from early childhood to early adulthood: The impact of early IQ differences on stability and change over time. *Learning and Individual Differences, 32*, 156–162.

Stokes, S. F., Kern, S., & Dos Santos, C. (2012). Extended statistical learning as an account for slow vocabulary growth. *Journal of Child Language, 39*(1), 105–129.

Striano, T., Tomasello, M., & Rochat, P. (2001). Social and object support for early symbolic play. *Developmental Science, 4*(4), 442–455.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software, 59*(5). https://doi.org/10.18637/jss.v059.i05.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2010). Language development. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (pp. 239–257). New York: John Wiley.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*(5), 675–691.

Troseth, G. L., & DeLoache, J. S. (1998). The medium can obscure the message: Young children's understanding of video. *Child Development, 69*, 950–965.

Vigil, D. C., Hodges, J., & Klee, T. (2005). Quantity and quality of parental language input to late-talking toddlers during play. *Child Language Teaching and Therapy, 21*(2), 107–122.

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29*(3), 257–302.

Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. W. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research, 50*(6), 1562–1592.