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

2021

Authors:

Rachael W. Cheung[1]

Calum Hartley[1]

Padraic Monaghan[1]

[1] Department of Psychology, Lancaster University, Bailrigg, Lancaster, LA1 4YF, United Kingdom

Keywords: late talkers, symbolic understanding, picture comprehension, language development

Declarations of interest: none**.**

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. While all children improved in accuracy over time as expected, TD children exhibited an advantage over LT children, despite both groups utilising 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-years-old, and expressive vocabulary predicted performance at ~3.5-years-old. Task performance at 3.5-years-old was predicted by earlier receptive vocabulary, but this effect was largely mediated by concurrent expressive vocabulary. Social ability across the whole sample at ~2-years-old also predicted and mediated the effect of receptive vocabulary on concurrent task performance. These findings suggest that LT 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 et al., 2005). A symbol is something that someone intends to represent something else, and can take many forms, e.g. gestures, graphics, text, words, maps, and so on (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 et al., 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 et al., 2018). This means that early language impairments have the potential to also affect children's understanding of non-linguistic 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

In order 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-old, infants manually investigate pictures as if they were real objects, grasping and plucking at depicted items (DeLoache et al., 1998).

By 18-months-old, 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, as verbal labels provide clues about how 2-D visual symbols relate to referents in the world (Callaghan, 2000; Ganea et al, 2008). When testing 2-year-olds, Preissler and Bloom (2007) demonstrated that labelling a picture of an unfamiliar object ('this is a *wug*. Can you show me another one?') directed children to identify the symbolised object 90% of the time, whereas children only identified symbolised referents 30% of the time when pictures were not labelled ('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 labelling cues children to view pictures as symbolic representations rather than objects. Children aged 15-, 18- and 24-month-olds will spontaneously extend a novel label (e.g. 'whisk') taught using a picture to its corresponding 3-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 that caregivers heavily invest in, going to considerable lengths to teach their children words. Children may thus 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 asked to choose their referents from pairs of objects. In Control trials, linguistic scaffolding was unavailable, as the two objects had the same category label (e.g.

two types of dog). In Standard trials, linguistic scaffolding could be used, as the two objects had distinct category labels (e.g. dog and cat). The study demonstrated that 2.5-year-olds only performed above chance when pictures could be unambiguously matched to objects using verbal labels, while 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-old ($r = .41$), whereas concurrent expressive measures better predicted symbolic play in studies of children over 3-years-old ($r = .36$). However, this interaction was driven by a difference in effect sizes for receptive, rather than expressive vocabulary, as the expressive effect size remained stable across ages, making any differential effects at different ages hard to clearly identify. As 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 graphic symbol comprehension and production at 28, 36, and 42-month-old and found that graphic comprehension scores positively correlated with receptive language and graphic production scores positively correlated with expressive language. Kirkham et al. (2013) also assessed the relationship between language, graphic symbols and symbolic play. They found that Mean Length of Utterance of Five Words at 4 years predicted symbolic play and graphic symbolism at 5 years, and that receptive and expressive language score combined at age 4 predicted symbolic play at age 5. Receptive vocabulary has also been found to correlate with performance on scale model search tasks (finding a real hidden object in a room, based on the location of a miniature object positioned in a scale model of the same room; Homer & Nelson, 2009).

In summary, cross-sectional and longitudinal evidence show that linguistic and non-linguistic symbolic domains are developmentally inter-related. Verbal labelling 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 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 defined as 18–30-month-old children at or below the 10th percentile of expressive vocabulary compared to other children their age, without neurodevelopmental or sensory deficits (Fisher, 2017). Although the majority of LT children

recover by approximately age 5 (Rescorla, 2011), a minority – between ~12% – 25% (Collisson et al., 2016; Henrichs et al., 2011; Reilly et al., 2010; Roulstone et al., 2002; Zubrick et al., 2007) – develop Developmental Language Disorder (DLD). Although many LT children reach 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 et al., 2000; Rice et al., 2008).

LT children are characterised by expressive vocabulary deficits, yet can have varying receptive vocabulary skills (Fisher, 2017), whereas in typically developing (TD) children, expressive and receptive vocabulary are tightly intertwined. Evidence that expressive and receptive vocabulary might exert differential effects on pictorial understanding can be found in autistic spectrum disorder (ASD) studies, as children with ASD typically exhibit a range of language difficulties (Eigsti et al., 2011). Studies that test the extension of words from pictures to symbolised 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-old; Hartley & Allen, 2015; 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-old), Hartley et al. (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, but not receptive, deficits might struggle utilising 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 utilised. Many accounts of symbolic understanding rely on a foundation of socio-cognitive 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) argue that pictures are inherently communicative, and understanding them is driven by a 'basic affiliative need' to communicate and identify with other humans. They describe pictorial understanding development in stages that are built on social factors, beginning from infants (12-months-old) who imitate the actions of adults when given pictorial symbols, to toddlers (2 – 4-years-old) who use social scaffolding through language and imitation to understand pictures, and finally to school-aged children (4 – 5-years-old) who begin to understand not only symbol-referent relations, but also intentions of the symbol-creator.

Differences in socio-cognitive ability may contribute towards some of the differences in pictorial understanding found in ASD 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), less lexical and prosodic information (D'Odorico & Jacob, 2006) and produce fewer expansions, less self-directed speech and less general responses (Vigil et al., 2005). Others have found no difference in maternal input, but rather found that as LT children simply say less, caregivers have less to expand upon (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 et al., 2016).

Overall, despite socio-cognitive skills forming the basis of theoretical accounts of symbolic development, we do not know how individual differences in social ability in non-clinical populations 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 behaviour than those who are not.

The current study

In sum, there are three distinct areas in which further research is necessary. Firstly, although symbols form a key part of communication throughout life, and TD children use language before 3 – 4-years-old 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 et al., 2001; Rescorla & Goossens, 1992). As symbolic play, pictorial understanding, and language are developmentally inter-related (Callaghan & Rankin, 2002; Kirkham et al., 2013), LTs may also exhibit deficits in pictorial understanding.

Secondly, 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 we do not know how emerging language skills interact with pictorial understanding at different ages.

Thirdly, regardless of theoretical literature maintaining that social scaffolding and language are crucial to pictorial understanding, the relationships between individual social

ability, language delay, and pictorial understanding have not been directly investigated in non-clinical populations.

We address 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 – 2.4-years-old (timepoint 1; T1) and 3.5 – 3.9-years-old (timepoint 2; T2). We also considered the effect of social ability measured at 2 – 2.4-years-old.

We hypothesised 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 hypothesised 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.[1] As an exploratory analysis, we also hypothesised that children with less sophisticated social ability would score lower on picture comprehension accuracy.

Material and Methods

Participants

Participants were part of a longitudinal project intended to capture differences between LT and TD children for 18 months, between the ages of 2 – 2.4-years-old to 3.5 – 3.9-years-old. The picture comprehension task was administered at the first and last time points.

---

[1] We originally hypothesised that LT children who had not recovered would perform less accurately with linguistic scaffolding, and that recovered-LT children would perform on par with TD children (see preregistrations). However, as we could only test half of the original sample due to COVID-19, resulting in small subgroups, we utilised concurrent expressive vocabulary across all participants at T2.

Participants were recruited using flyers from Lancaster 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-CDI (Hamilton et al., 2000) and included if they met one of the following criteria: TD with productive vocabulary score $\geq 25^{th}$ percentile, or LT with productive vocabulary score $\leq 10^{th}$ percentile. These criteria were chosen to ensure two distinct groups, with the LT criterion 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 time point aged 2.0 – 2.4-years-old (T1); however, 2 TDs did not complete the pictorial understanding task due to fussiness and so were excluded from the final sample of 59 children (38 TD and 21 LT). At 3.5 – 3.9-years-old (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 was used as a measure of individual social proficiency (raw scores). 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 et al., 2013) at the T2 test visit.

Picture comprehension task

Objects: We used the same criteria as Callaghan (2000) for selecting relevant stimuli. There were 32 different objects in total, split into 16 pairs. For each condition, there were four trials, one from each of four groups: animals, natural, household/indoor artifacts, and vehicles (16 trials in total). For the Control-Familiar 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 Control-Unfamiliar 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 Standard-Familiar condition, pairs of familiar objects had the same global label (e.g. animal) but different basic labels (e.g. cat and rabbit). For the Standard-Unfamiliar condition, pairs of unfamiliar objects had the same global label (e.g. vehicle) but different subordinate 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 colours and poses were chosen (sitting vs. standing dogs); for artifacts, different colours, 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-old, range = 10–16-months-old). Example stimuli can be seen in Figure 1 (see Appendix A for all stimuli used).

Pictures: Sixteen 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 cut out at the back that allowed the experimenter to rearrange objects out of sight of the child. 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 (Control trials; two objects with the same basic label, e.g. two types of dog) and conditions where it could be (Standard trials; two objects with different basic labels, e.g. rabbit and cat). We also manipulated the familiarity of objects depending on the child's knowledge of the labels and objects (Familiar and Unfamiliar) within the Control and Standard trials. The order of trial types was randomised per participant, with no more than two trial types of the same time presented consecutively.

The task was administered at T1 and T2. Participants were tested with the same mobile set-up for the task, either at the participant's home or in a designated room at the Babylab depending on the family's preference. Where visits took place at 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-metre wide, low fold-out table. The experimenter held up the relevant trial picture card (e.g. cat) and said "Look!". The picture was presented for 4 seconds 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, 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).
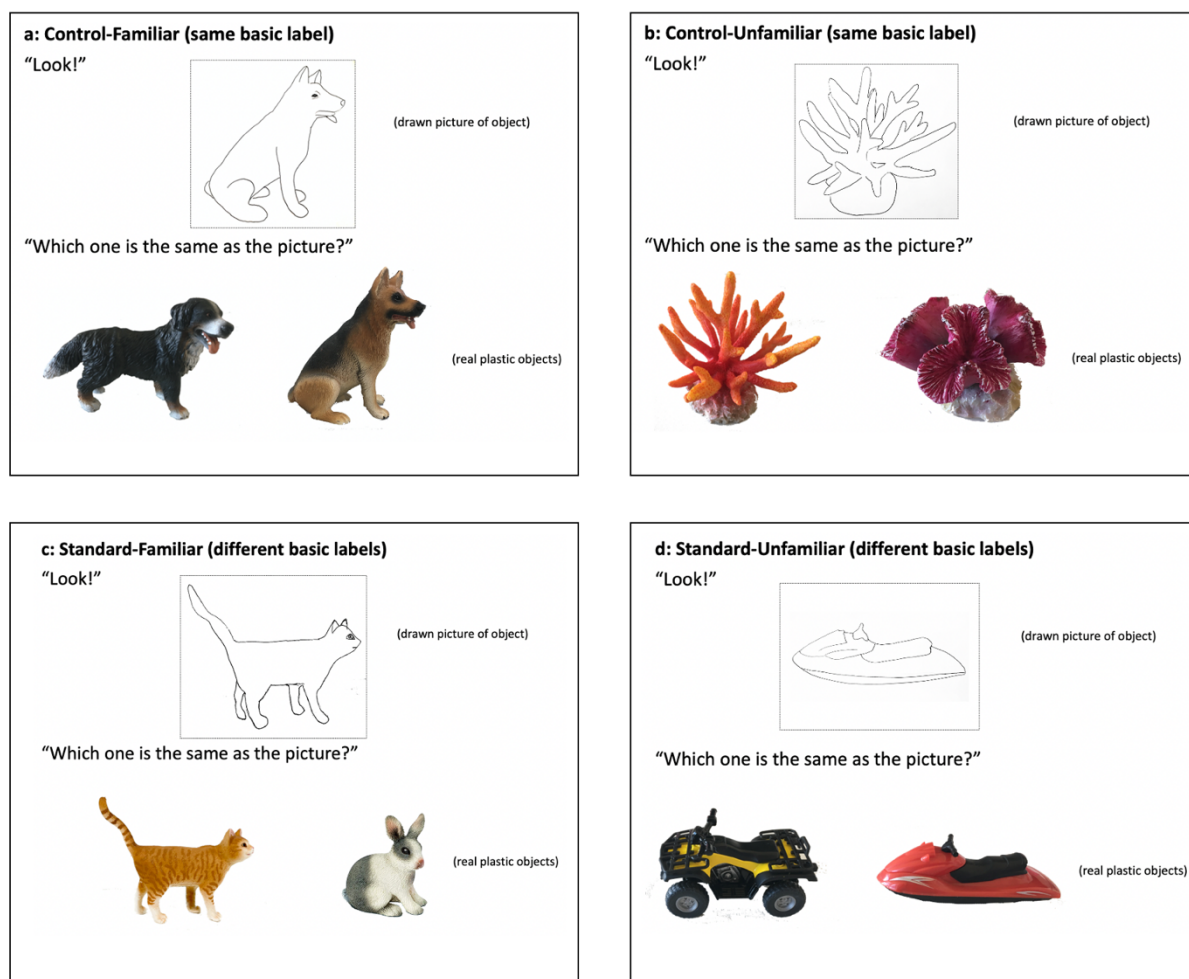
Figure 1. Example of stimuli and trial types used: a) Control-Familiar; b) Control-Unfamiliar; c) Standard-Familiar; d) Standard-Unfamiliar.

Results

All data and code can be found at:

https://osf.io/ywmx5/?view_only=14f51c730c4c47758893bc684d7cebf5, alongside pre-registrations with a document that explains deviations due to the COVID-19 pandemic.

Sample

Table 1 contains T1 and T2 final sample demographics, questionnaire, and vocabulary scores. 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 ($t(46.39) = 1.35$, $p = .183$) or Leiter-3 scores ($t(10.02) = -1.45$, $p = .178$).

Table 1. Mean and standard deviation for demographic, questionnaire, and vocabulary scores for samples at first timepoint (T1) and second timepoint (T2).

Timepoint	T1: 2.0 – 2.4-years-old (N = 59)		T2: 3.5 – 3.9 years-old (N = 29)	
	TD (n = 38)	*LT (n = 21)*	*TD (n = 20)*	*LT (n = 9)*
Age (years)	2.19 (0.12)	2.19 (0.12)	3.73 (0.12)	3.75 (0.15)
Gender (ratio, m : f)	16 : 22	14 : 7	7 : 13	7 : 2
Receptive vocab[a]	384 (38.00)	258 (93.40)	119.45 (5.35)	111.11 (7.54)
Expressive vocab[a]	331 (73.20)	60 (49.50)	122.75 (8.43)	108.11 (12.90)
Social ability (SRS-2)[b]	27.9 (12.40)	32.1(10.80)		
Non-verbal IQ (Leiter-3)			98.55 (6.68)	92 (12.80)

LT = late talker; SRS-2 = Social Responsiveness Scale-2; TD = typically developing

[a] T1: Communicative-Development Inventories; T2: Receptive/Expressive One-Word Picture Vocabulary Tests

[b] Higher scores indicate lower responsiveness/ability.

Descriptive task results

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%). At the first timepoint (T1), when participants were aged 2.0 – 2.4-years-old, TD children performed significantly above chance overall ($M = 0.60$; $t(37) = 4.64$, $p < .001$). TD children performed below chance on the Control-Familiar trials, but above chance in all other trial types: Standard-Familiar ($p <.001$), Control-Unfamiliar ($p = .007$), and Standard-Unfamiliar ($p = .004$; Table 2, Figure 2). The difference between Control-Familiar ($M = 0.42$) and Standard-

Familiar ($M = 0.72$) trials 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 utilise perceptual similarities between pictures and objects to select the correct object. Not knowing the basic or subordinate label in these conditions was thus advantageous, as it enabled them to utilise perceptual similarity only.

LT children did not perform significantly above chance overall ($M = 0.53$; $t(20) = 1.44$, $p = .083$). They performed below or at chance in Control-Familiar, Control-Unfamiliar and Standard-Unfamiliar trials (Table 2, Figure 2). They performed above chance on Standard-Familiar trials ($M = 0.59$; $p = .021$), and at a level similar to TD children in Control-Familiar trials ($M = 0.42$). This suggests that LT children were sometimes able to use verbal labels when they were available, but did not make use of them to the same degree as TD children. Scoring below chance when objects were unfamiliar suggested that LT children also struggled to match pictures to unfamiliar objects based upon perceptual similarities alone.

At the second timepoint (T2), when participants were aged 3.5 – 3.9-years-old, both TD and LT children performed above chance overall (Table 2, Figure 2; TD: $M = 0.80$; $t(19) = 10.93$, $p < .001$; LTs: $M = 0.73$; $t(8) = 4.80$, $p < .001$). Both TD and LT children performed at chance in Control-Familiar trials, but significantly above chance in all other trial types. These results indicate that LT children were largely able to utilise both perceptual information and linguistic labels in the task at T2.

Table 2. Mean accuracy and standard error at test at each timepoint per group. Trial types: *Control* = object pairs with the same global label and same basic label, inhibiting

verbal scaffolding; *Standard* = object pairs with the same global label and different basic labels, allowing verbal scaffolding; *Familiar* = known objects to the child; *Unfamiliar* = unknown objects to the child.

	Trial Type	**Typically Developing**	**Late Talker**
		Mean (SE)	*Mean (SE)*
	Control Familiar	0.42 (0.04)	0.42 (0.05)
Time 1	Control Unfamiliar	0.65 (0.04)***	0.54 (0.05)
(2-2.4-years-old)	Standard Familiar	0.72 (0.04)***	0.59 (0.04)*
	Standard Unfamiliar	0.64 (0.04)***	0.57 (0.06)
	Control Familiar	0.61 (0.07)	0.50 (0.09)
Time 2	Control Unfamiliar	0.90 (0.03)***	0.81 (0.06)***
(3.5 – 3.9-years-old)	Standard Familiar	0.90 (0.03)***	0.86 (0.06)***
	Standard Unfamiliar	0.78 (0.05)***	0.75 (0.09)*

*p <.05, ** p <.01, *** p <.001; p-values to 3 decimal places; within-group One-Sample Welch T-Tests against chance (50%)
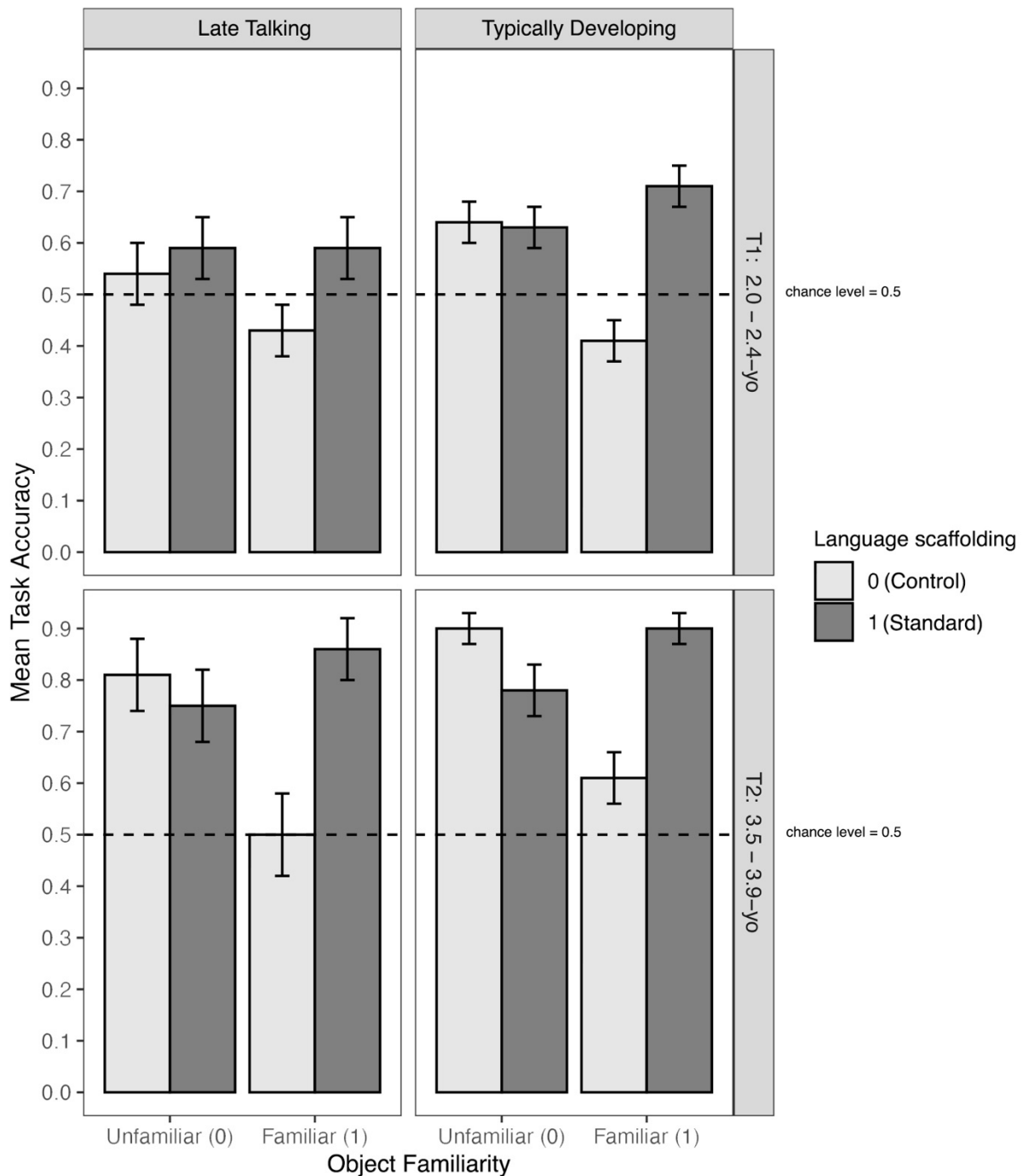
Figure 2. Mean accuracy and standard error at test across trial types per group over time. Trial types: *Control* = object pairs with the same global label and same basic label, inhibiting verbal scaffolding; *Standard* = object pairs with the same global label and different basic labels, allowing verbal scaffolding; *Familiar* = known objects to the child; *Unfamiliar* = unknown objects to the child.

Task analyses: overview

We conducted three analyses to assess our research hypotheses. The first tested the longitudinal predictive effect of LT status over time using generalised linear mixed effects modelling (GLME). 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 to the data with and without social ability, and by using a post-hoc mediation analysis.

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 package *lme4* (Bates et al., 2015). Models were built up sequentially, adding in one fixed effect at a time and comparing each model to 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 was analysed, we also attempted to fit a random slope of timepoint per participant, but this failed to converge. To analyse fixed effects of trial type, we coded them as follows: *object familiarity*: Unfamiliar coded as 0, and Familiar coded as 1, and *language scaffolding*: Control coded as 0, and Standard 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.[2] All models run can be viewed on the Open Science Framework (https://osf.io/ywmx5/?view_only=14f51c730c4c47758893bc684d7cebf5).

---

[2] Due to the disparate scales utilised 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 they were on a closer scale to accuracy. This is indicated in the Tables reporting GLME result estimates. Please see R code on OSF for more details.

All post-hoc mediation analyses were undertaken using the *mediation* package in R [version 1.1.463] (Tingley et al., 2014). For each analysis, 1000 simulations were used to estimate model effects using the quasi-Bayesian Monte Carlo method (Imai et al., 2010).

Does late talking 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 (Table 3; $\chi^2(3)$ = 13.51, p = .004).

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

The added effect of timepoint indicated that participants performed significantly more accurately at age 3.5 – 3.9-years-old as compared to 2.0 – 2.4-years-old ($p < .001$), and the effect of population indicated 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 late-talking status on performance across time, with LT children attaining lower accuracy scores overall when total performance was assessed across both timepoints. However, as 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.

Table 3. Longitudinal analysis of task accuracy over time. Summary table of best-fitting model results from general linear mixed effect model analyses on accuracy over time, using fixed effects of trial type, population 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
Standard	-0.13	0.25	-0.52	.603
Familiar * Standard	1.34	0.35	3.82	<.001
Timepoint (T2: 3.5 – 3.9-years-old)	0.99	0.14	7.04	<.001
Population (TD)	0.34	0.15	2.28	.022

LT = late talker; TD = typically developing

[a]*Intercept corresponds to no language scaffolding (0) and object unfamiliar (0), population LT, and timepoint T1 (2.0 – 2.4-years-old).*

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

<u>Receptive vocabulary:</u> 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 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 (Table 4; model comparison: $\chi^2(2) = 9.14$, $p = .010$). This indicated that children with higher concurrent receptive vocabularies performed significantly more accurately at 2 – 2.4-years-old ($p = .038$).

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 (Table 4; model comparison: $\chi^2(3) = 10.11$, $p = .018$), showing that children with higher receptive vocabularies at 2.0 -2 .4-years-old, performed more accurately on the picture comprehension task when they were 3.5 – 3.9-years-old ($p = 0.18$).

Table 4. Cross-sectional analyses of predictive effect of receptive vocabulary on task accuracy. Summary table of best-fitting model from general linear mixed effect model analyses on T1 and T2 accuracy as dependent variable (cross-sectional) with fixed effects of trial type (familiarity of objects and availability of labels) and T1 receptive vocabulary.

T1: age 2.0 – 2.4-years-old

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
Standard	0.06	0.21	0.28	.776

Familiar * Standard	0.97	0.30	3.24	.001
T1 receptive vocabulary (CDI)[b]	0.16	0.08	2.07	.038

T2: age 3.5 – 3.9-years-old

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
Standard	-0.76	0.63	-1.20	.229
Familiar * Standard	2.84	0.90	3.14	.002
T1 receptive vocabulary (CDI)[b]	0.57	0.24	2.37	.018

CDI = Oxford Communicative Development Inventories; ROWPVT-4 = Receptive One Word Picture Vocabulary Test

[a]*Intercept corresponds to Control (no language scaffolding; 0) and Unfamiliar (object familiarity; 0)*

[b]*Rescaled using x/100 to allow model fit*

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 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 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-old, 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 (Table 5; model

comparison: $\chi^2(3) = 10.12$, $p = .018$). The best fitting model to the data demonstrated that as children's concurrent expressive vocabulary at 3.5 – 3.9-years-old increased, so did their picture comprehension accuracy ($p <.001$).

Table 5. Cross-sectional analyses of predictive effect of expressive vocabulary on task accuracy: Summary table of best-fitting model from general linear mixed effect model analyses on T2 accuracy as dependent variable (cross-sectional) with fixed effects of trial type (familiarity of objects and availability of labels) and T2 expressive vocabulary.

T2: age 3.5 – 3.9-years-old				
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
Standard	-0.76	0.63	-1.20	.229
Familiar * Standard	2.85	0.91	3.14	.002
T2 expressive vocabulary (EOWPVT-4)	0.04	0.01	3.32	<.001

EOWPVT-4 = Expressive One Word Picture Vocabulary Test

[a]*Intercept corresponds to no language scaffolding (0) and object unfamiliar (0)*

Relationship between receptive and expressive vocabulary in predicting task accuracy:
The cross-sectional analyses indicated that early receptive vocabulary at ~2-years-old predicted both concurrent and later task accuracy at ~3.5-years-old, and later expressive vocabulary at ~3.5-years-old predicted concurrent task accuracy at ~3.5-years-old.

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 (Figure

4). 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% CI: [0.01, 0.12]; *p* = .016). The results indicated that of the estimated increase in probability of task accuracy at ~ 3.5-years-old (total effect: 0.10) due to earlier receptive vocabulary at 2 – 2.4-years-old, 0.07 was estimated to be mediated through later expressive vocabulary at 3.5 – 3.9-years-old, and 0.03 was estimated to be from earlier receptive vocabulary at 2 – 2.4-years-old.
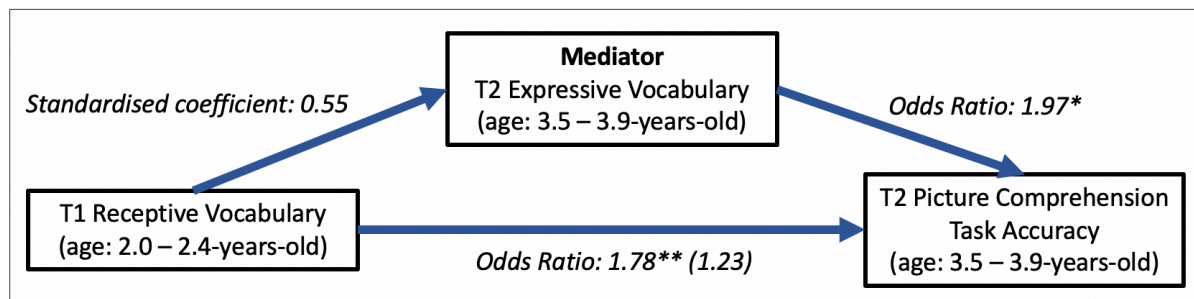

Figure 4. Results of mediation analysis assessing indirect effect of T1 receptive vocabulary on T2 task accuracy through T2 expressive vocabulary. The value in parentheses indicates the direct effect of receptive vocabulary when the mediator is included. **p < .05; **p = .01*

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 to the original best-fitting model for each time point.

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 was a better fit to the data than a model

without SRS-2 (Table 6; model comparison: $\chi^2(1) = 5.40$, $p = .020$), suggesting that children with less social responsiveness were less accurate at matching pictures to symbolised objects ($p = .023$) regardless of language ability.

For T2, a GLME model with SRS-score as an additional fixed effect was not a better fit to the data when compared to 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 (Figure 5). 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 – 2.4-years-old (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.

Table 6. Cross-sectional analyses of added effect of social ability to predicting task accuracy: Summary table of best-fitting model from general linear mixed effect model analyses on T1 accuracy as dependent variable (cross-sectional) with fixed effects of trial type (familiarity of objects and availability of labels), T1 receptive vocabulary, and T1 social ability.

T1: age 2.0 – 2.4-years-old				
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
Standard	0.06	0.21	0.32	.790
Familiar*Standard	0.98	0.30	3.14	.001

T1 receptive vocabulary (CDI)	0.09	0.09	1.11	.265
T1 social ability (SRS-2)[b]	-1.45	0.63	-2.27	.020

CDI = Communicative Development Inventories; SRS-2 = Social-Responsiveness Scale-2

[a]*Intercept corresponds to no language scaffolding (0) and object unfamiliar (0)*

[b]*Rescaled using x/100 to allow model fit. Higher scores indicate less social responsiveness.*
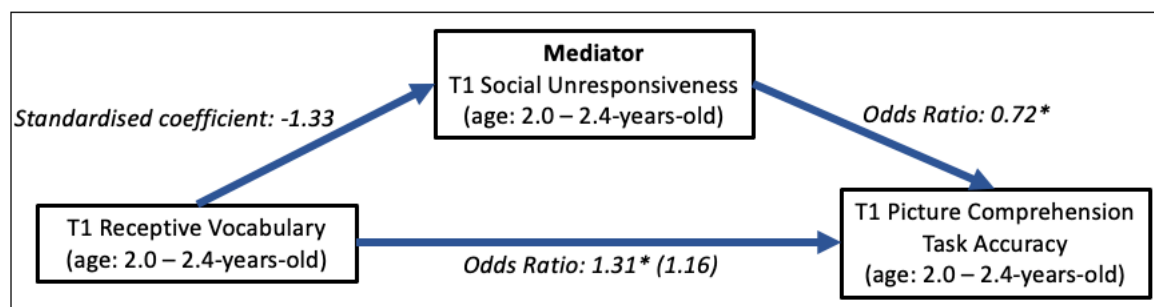

Figure 5. Results of mediation analysis assessing indirect effect of T1 receptive vocabulary on T3 task accuracy through T1 social ability. Note that the SRS-2 is scored as such that higher scores indicate lower ability, and that the value in parentheses indicates the direct effect of receptive vocabulary when the mediator is included. *$p <.05$

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.

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 (i.e. a picture of a cat) and when the target object is selected (i.e. a plastic cat and a plastic rabbit on the tray). In Standard-Familiar trials, if the participant generated a label when the target was cued, they could achieve a correct response by identifying the target object based on its matching label rather than responding based on perceptual similarity (however, this strategy is unavailable when both referent objects share the same label as the picture, or labels are unknown). 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 associated concepts both when the target is cued and when the object is selected, and thus directly utilise that linguistic information to select the correct object.

More generally, our results suggest that at an earlier age, children rely on understanding linguistic information and concurrent social ability, but at a later age, they shift to using their expressive vocabulary skills to scaffold picture comprehension. We now outline the implications of these results for LT children, typical development, and future considerations.

Implications for late-talking children

At both time points, 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 (i.e. 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, as no significant effects of population were found cross-sectionally, any differences between the populations in our study were subtle. Furthermore, as there was no significant interaction between population or vocabulary measures with 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, just 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 age 2 – 2.4 years 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, categorising 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 which suggest that language ability, in LT children and DLD, falls upon 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 age 2 – 2.4 years predicted task performance at the same age across the whole sample. The implications for this in typical

development are discussed below, but of note is that social ability may actually help 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 et al., 2014; Cable & Domsch, 2010; Robertson & Weismer, 1999). More pro-social 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 labelling pictures, accelerating their acquisition of a linguistic strategy in pictorial understanding.

Implications for typical development

Across both timepoints, children struggled most with the Control-Familiar trials where access to verbal scaffolding was blocked. In theory, linguistic scaffolding was only available in Standard-Familiar tasks, and selecting the correct referent object in all other trial types required children to attend to the pictures' perceptual features. While TD children applied this strategy successfully in trials involving unfamiliar objects, the consistently lower performance in Control-Familiar trials suggests that generating labels for pictures may 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. Moreover, 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.

The function that language plays in aiding pictorial understanding may be in creating 'cognitive distance' (p.132, Homer & Nelson, 2009). By enabling children to treat pictures as distinct to 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, but rather on language comprehension ability and social ability. Social ability both predicted task performance at age 2.0 – 2.4 and mediated the effect of receptive vocabulary on task performance. The lack of interaction between condition and receptive vocabulary also suggests that language not only scaffolds picture comprehension – as evidenced by the highest accuracy scores being in Standard-Familiar trials – but also that receptive vocabulary alongside social ability may mediate pictorial understanding more generally.

These results are consistent with a socio-cognitive 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 et al., 2004). Striano et al. (2001) found that when given uninteresting or ambiguous objects (e.g., a stapler), 2 – 3-year-olds did not perform symbolic actions spontaneously and largely declined to play at all without an experimenter modelling symbolic actions or actively engaging the child. However, 4-year-olds were better able to play with the items independently. In a longitudinal study, Callaghan and Rankin (2002) also found that cultural scaffolding, consisting of explicitly highlighting the relationship between objects and pictures, improved children's graphic symbol 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 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 manipulate social cues directly within the task, will help elucidate these mechanisms.

At an older age, expressive, 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-old, language again forms a central component of how representations of the world are understood, but the ability to actively talk about symbols and partake in social discourse about them may actually afford a stronger abstraction of pictures from referring objects than simply understanding what others say. Tomasello and colleagues (Rakoczy et al., 2005; Tomasello et al., 2005) describe language as a means through which children are able to develop other symbolic functions, such as pretend play. With advancing linguistic ability from 3-4 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 non-intentional imitation of meaning as infants (such as copying gestures or early words), to intentional representation and sharing of meaning as school-aged children (such as using conventional symbolic systems like discourse). This process is facilitated by externalisation 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. As our task was designed to test children's understanding of symbolic relations between pictures and 3-D objects, data collection could not be completed online, as perceiving all stimuli via a 2-D screen would fundamentally change the nature of the task (Troseth & DeLoache, 1998). When face-to-face testing resumes, future directions include testing a larger sample to assess the links between social ability and picture comprehension.

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 et al., 2009; Schneider et al., 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 non-verbal IQ may not predict symbolic or graphic understanding (Kirkham, 2013). However, it is possible that individual differences in attention and executive functioning were not fully accounted for.

Our sample also consisted of families from mid to high income backgrounds. Consequently, although we can be confident that any differences between the children in our sample were less likely to be due to socioeconomic or environmental causes, we cannot extend these findings without further testing. Furthermore, the use of pictures and symbols are 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 utilised a parent-report measure for vocabulary at T1 rather than an experimenter-administered measure. However, as we used two distinct cut-offs for the two groups, it is unlikely that parent-report measures were so inaccurate as to incorrectly characterise group status at T1. Furthermore, the ROWPVT-4 and EOWPVT-4 are, to some extent, measures of pictorial understanding in themselves that children might struggle with before the age of 3. Parent-reported CDIs, on the other hand, can capture a broader assessment of how children utilise language in their everyday lives during the earlier stages of language development.

Conclusions

Our study has implications for both TD and LT children. Through a longitudinal study, we demonstrate firstly that LT children show evidence of less accurate picture comprehension skills over time when compared to a TD sample and, secondly, that these differences are subtle and subject to effects of participant heterogeneity. These findings suggest that late talking (in line with DLD) and its effects on pictorial understanding may be best considered on a dimensional scale, rather than a categorical one. Crucially, as the trajectory of development for LT children resembled that of earlier typical development, albeit developmentally delayed, this also 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 scaffold picture comprehension over time, with receptive vocabulary predicting picture comprehension at 2-years-old, and expressive vocabulary predicting picture comprehension at 3.5-years-old. 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 as well as 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.

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. https://doi.org/10.1016/j.jcomdis.2014.07.002

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. https://doi.org/10.1111/1460-6984.12335

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, 100824014249025.

https://doi.org/10.3109/13682822.2010.487883

Callaghan, T. C. (2000). Factors affecting children's graphic symbol use in the third year: Language similarity and iconicity. *Cognitive Development*, *15*(2), 186–214. https://doi.org/10.1016/S0885-2014(00)00026-5

Callaghan, T. C. (2020). The origins and development of a symbolic mind: The case of pictorial symbols. *Interchange*, *51*(1), 53–64. https://doi.org/10.1007/s10780-020-09396-z

Callaghan, T. C., & Corbit, J. (2015). The development of symbolic representation. In *Handbook of Child Psychology and Developmental Science* (pp. 1–46). American Cancer Society. https://doi.org/10.1002/9781118963418.childpsy207

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.

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), ii–128.

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. *The Journal of Pediatrics*, *172*, 168-174.e1. https://doi.org/10.1016/j.jpeds.2016.02.020

Constantino, J. N., & Gruber, C. P. (2012). *Social Responsiveness Scale 2 (SRS-2)*. 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.

   https://doi.org/10.1044/jshd.5502.262

DeLoache, J. S. (1995). Early understanding and use of symbols: The Model Model. *Current

   Directions in Psychological Science*, *4*(4), 109–113. https://doi.org/10.1111/1467-

   8721.ep10772408

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

   70. https://doi.org/10.1016/j.tics.2003.12.004

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.

   https://doi.org/10.1111/1467-9280.00039

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. https://doi.org/10.1080/13682820500342976

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. https://doi.org/10.1111/j.1460-6984.2011.00095.x

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. https://doi.org/10.1016/j.rasd.2010.09.001

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), 1–173; discussion 174-185.

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. https://doi.org/10.1044/2017_JSLHR-L-16-0310

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. https://doi.org/10.1016/j.jecp.2009.05.008

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). American Psychological Association. https://doi.org/10.1037/11867-003

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. https://doi.org/10.1017/S0305000900004414

Hartley, C., & Allen, M. (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. https://doi.org/10.1007/s10803-013-2007-4

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. https://doi.org/10.1177/1362361317729613

Henrichs, J., Rescorla, L., Schenk, J. J., Schmidt, H. G., Jaddoe, V. W. V., Hofman, A., Raat, H., Verhulst, F. C., & 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. https://doi.org/10.1044/1092-4388(2010/09-0255)

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. https://doi.org/10.1080/15248370903041298

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. https://doi.org/10.1097/01.CHI.0000046889.27264.5E

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309–334. https://doi.org/10.1037/a0020761

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. https://doi.org/10.1002/icd.1786

Leonard, L. B. (2014). Part II: Desrcibing the data: Lingustic and nonlingustic findings: Chapter 5: Exploring the boundaries of SLI. In *Children with Specific Language Impairment* (pp. 151–180). 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. https://doi.org/10.1080/03004430.2015.1039529

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.

https://doi.org/10.1044/1092-4388(2001/070)

Martin, N., & Brownell, R. (2011). *Expressive and Receptive One-Word Picture Vocabulary*

Tests–4 (EOWPVT-4/ROWPVT-4) (4th ed.). Academic Therapy Publications.

Nelson, K. (2007). *Young Minds in Social Worlds: Experience, Meaning, and Memory :*

Experience, Meaning, and Memory. Harvard University Press.

http://ebookcentral.proquest.com/lib/lancaster/detail.action?docID=3300052

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.

https://doi.org/10.1044/jshr.3405.982

Pierroutsakos, S. L., & DeLoache, J. S. (2003). Infants' manual exploration of pictorial

objects varying in realism. *Infancy*, *4*(1), 141–156.

https://doi.org/10.1207/S15327078IN0401_7

Preissler, M. A. (2008). Associative learning of pictures and words by low-functioning

children with autism. *Autism*, *12*(3), 231–248.

https://doi.org/10.1177/1362361307088753

Preissler, M. A., & Bloom, P. (2007). Two-Year-Olds Appreciate the Dual Nature of

Pictures. *Psychological Science*, *18*(1), 1–2. https://doi.org/10.1111/j.1467-

9280.2007.01837.x

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.

https://doi.org/10.1207/s15327647jcd0502_2

Quinn, S., Donnelly, S., & Kidd, E. (2018). The relationship between symbolic play and

language acquisition: A meta-analytic review. *Developmental Review*, *49*, 121–135.

https://doi.org/10.1016/j.dr.2018.05.005

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).

Psychology Press

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., Conway, L., Eadie,

P., & Bretherton, L. (2010). Predicting language outcomes at 4 years of age: Findings

from Early Language in Victoria Study. *Pediatrics*, 236(6), e1530-7.

https://doi.org/10.1542/peds.2010-0254

Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist? *Developmental*

Disabilities Research Reviews, *17*, 141–150. https://doi.org/10.1002/ddrr.1108

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.

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.

https://doi.org/10.1044/1092-4388(2002/028)

Rescorla, L. (2005). Age 13 language and reading outcomes in late-talking toddlers. *Journal*

of Speech, Language, and Hearing Research, *48*(2), 459–472.

https://doi.org/10.1044/1092-4388(2005/031)

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.

https://doi.org/10.1017/S0305000900004232

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.

https://doi.org/10.1044/1092-4388(2008/029)

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. https://doi.org/10.1044/jslhr.4205.1234

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). Psychology Press

Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (2013). *Leiter-3: Leiter International Performance Scale* (3rd ed.). Stoelting Company.

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. https://doi.org/10.1080/03004430212126

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. https://doi.org/10.1016/j.lindif.2014.02.001

Striano, T., Tomasello, M., & Rochat, P. (2001). Social and object support for early symbolic play. *Developmental Science*, *4*(4), 442–455. https://doi.org/10.1111/1467-7687.00186

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*. Harvard University Press.

Tomasello, M. (2010). Language Development. In U. Goswami (Ed), *The Wiley-Blackwell Handbook of Childhood Cognitive Development* (pp. 239–257). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444325485.ch9

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. https://doi.org/10.1017/S0140525X05000129

Troseth, G. L., & DeLoache, J. S. (1998). The medium can obscure the message: Young children's understanding of video. *Child Development*, *69*(4), 950–965. https://doi.org/10.1111/j.1467-8624.1998.tb06153.x

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. https://doi.org/10.1191/0265659005ct284oa

Vygotsky, L. S. (1980). *Mind in Society: The Development of Higher Psychological Processes*. 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. https://doi.org/10.1006/cogp.1995.1016

Zubrick, S. R., Taylor, C. L., & Rice, M. L. (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. https://doi.org/10.1044/1092-4388(2007/106)

Appendix A: Stimuli used in picture comprehension task

Category	Control Familiar	Control Unfamiliar	Standard Familiar	Standard Unfamiliar
Animal	German shepherd (sitting) + Burmese mountain (standing)	Grey 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 + artichoke
Household/ indoor artifacts	White porcelain Chinese spoon v silver metal Western spoon	Metal garlic press + wooden garlic press	Black hairbrush + wooden comb	Binoculars + safety science glasses