Characterising language learning in typical and atypical populations

Rachael W. Cheung

This thesis is submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Psychology

Lancaster University

June 2021

Table of contents

[submission in alternative format form; removed as per Lancaster University guidelines]

Declaration

I declare that this thesis is entirely my own work completed under the supervision of Padraic Monaghan and Calum Hartley (author contributions are listed at the start of chapters). None of the work in this thesis has been submitted elsewhere in support of application for another degree at this or any other institution.

The parts of this thesis that I have submitted or published in academic journals during the course of this doctoral degree have been indicated at the beginning of the relevant chapter.

- Rachael W. Cheung (2021)

List of tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

dependent variable (cross-sectional) with fixed effects of trial type (familiarity of objects and availability of labels) and T1 receptive vocabulary.

Table 5. Cross-sectional analyses of predictive effect of expressive vocabulary on task accuracy. General linear mixed effects model results predicting T2 accuracy as dependent variable (cross-sectional) with fixed effects of trial type (familiarity of objects and availability of labels) and T2 expressive vocabulary.

Table 6. Cross-sectional analyses of added effect of social ability to predicting task accuracy: General linear mixed effect model results predicting 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.

List of figures

Chapter 2

Chapter 3

Chapter 5

Chapter 6

List of abbreviations

ALSPAC	Avon Longitudinal Study of Parents and Children
ASD	Autism Spectrum Disorder
ASQ	Ages and Stages Questionnaire
CDI	Communicative Development Inventories
CELF	Clinical Evaluation of Language fundamentals
CSWL	Cross-situational word learning
ECM	Emergent Coalition Model
ELVS	Early Language in Victoria Study
EOWPVT	Expressive One Word Picture Vocabulary Test
GCA	Growth curve analysis
GEE	Generalised estimated equations
GLME	General linear mixed effects
HSP	Human Simulation Paradigm
LDS	Language Development Survey
LME	Linear mixed effects
LT	Late talking
ME	Mutual exclusivity
MIM	Multi-modal Integration Model
MLU	Mean Length of Utterance
ND	Neighbourhood density
OSF	Open Science Framework
PP	Phonotactic probability
RDLS	Reynells Developmental Language Scales
ROWPVT	Receptive One Word Picture Vocabulary Test
SES	Socioeconomic status
SRS	Social Responsiveness Scale
TD	Typically developing

Acknowledgements

I cannot stress how lucky I've been to have the benefit of not one, but two, highly supportive supervisors. Padraic Monaghan and Calum Hartley have provided excellent supervision and sincere comradery across years, countries, and even during a global pandemic. Padraic, you have introduced entirely new ways of thinking and opportunities abound. Your ability to distill elegance out of chaos is deeply inspiring. I hope the future of academia will continue to be led by those like you, and by a similar willingness to advocate for early career researchers. Calum, thank you for always keeping your door open. I remain amazed by your unwavering enthusiasm, endless novel objects, and uncanny insights into research. Your dedication to your work and to your students is something I aspire to. Both of you have made my PhD not only fruitful, but also thoroughly enjoyable, and for this I am very grateful. Thank you.

Thank you to my family and to my friends for literally everything; as you know, you are the constant source of all good things. To Rens B. van Hoef; Здравствуй, темнота, мой старый друг – thank you for making me laugh every day, and for everything else.

I am grateful also to Agata Dymarska, Shal Joshi, and David Elliott, for being confidantes and for making the transition to research fun; and to James Brand, Jacky Chan, Shirly Ma, and Leigh Keating, who have provided both friendship and research insights in equal measure.

I also thank Professor Louise Connell at Lancaster University and Dr Julius Rentergent at TTD, for their sage advice and willingness to share their wisdom; and Professor David Reiss M.D. at Yale University for the invaluable lessons that he imparted many years ago, and for taking an interest in a young, naïve doctor. David, it is not hyperbole to state that your words have shaped my world view permanently for the better.

This thesis was made possible by the Leverhulme Trust, and by my wonderful fellow Babylab students and the wider Babylab team at Lancaster University. A particular thanks to dynamic duo Barrie Usherwood and Peter Tovee, who have helped many PhD students achieve their goals, and always with a smile.

Lastly, I must express my gratitude to all the parents and children who participated - particularly those in the longitudinal study, many of whom stayed engaged even during the pandemic. Working with these preschoolers, and in particular seeing them so well cared for, has been brilliant. Their experiences are a continual reminder of what developmental research ought to strive for: individuals who begin life with as many opportunities as possible. That might sound rather idealistic, but that's not necessarily a bad thing.

Thesis abstract

Young children learn words rapidly and amongst substantial environmental variation. How they manage to do so, and with relatively consistent results, is the topic of much debate in the developmental literature. Recent research has turned to the environmental variation surrounding children, and the vital information it may hold to help children to learn words more efficiently by way of statistical learning. The responsiveness of caregivers to this variation however – and the subsequent effects of the cues that they provide in real time – remains under-investigated.

The first part of this thesis investigates two key questions: 1) do caregivers alter the cues they provide in response to variation in the environment during word learning, and 2) does both environmental variation and caregiver response affect how their children learn words? The results of Study 1 demonstrate not only that caregiver cue use is dependent upon the amount of variation present, but also that children can learn more effectively in the face of variation. Study 2 then explores how these cues affect learning in real time, addressing the following questions: 1) how do adult learners make use of visual cues in relation to auditory labels as the learning process unfolds temporally, and 2) does interfering with this process affect word learning? This second study shows that it matters most *when* such cues occur in relation to the given label as the word learning process unfolds in time. These studies use a multi-disciplinary approach (computational modelling, child and adult experimental studies, and eyetracking) to address the multi-factorial process of language acquisition, and show how investigating the interaction of cues with environmental variation and within-trial learning processes can help us understand how children manage to learn words so consistently.

The multi-factorial process of word learning is then further explored through the lens of atypical language development, and offers a longitudinal perspective of word learning. Whereas part 1 of the thesis addresses receptive ability in cross-sectional studies, part 2

addresses the additional effects that expressive ability have on word learning processes over time. Late talkers are children who are developing typically with the exception of significant expressive language delay, producing fewer words than approximately 90% of their peers. Their unique deficit offers the chance to elucidate the differences between receptive and expressive language, and to study how language scaffolds development in other domains, such as symbolic understanding of pictures. However, late talking is also problematic: it is a risk factor for Developmental Language Delay, yet late talking children are notoriously heterogenous as a group, making predicting outcomes difficult. Crucially, determining whether or not late talking children utilise word learning mechanisms differently to typically developing children can provide an evidence-base for predicting outcomes from a clinical perspective.

The second part of this thesis reports a longitudinal study over 2 years in a cohort of late talking and typically developing children. Two research questions are examined: 1) do late talkers show deficits in word learning mechanisms as compared to typically developing children? 2) do late talking children show an impaired understanding of pictorial symbols as a result of their language delay, and how does expressive language affect symbolic understanding more generally?

This longitudinal study is unique in that it takes into account individual variation within the sample, and it also provides further evidence that a multiple hit hypothesis may best reflect the data, where a deficit in one area of ability does not necessarily lead to poor outcomes unless further deficits in other areas are present (i.e. there are multiple hits to language development ability). Study 3 shows that late talking children are impaired in some, but not all, word learning mechanisms; even when late talking children reach typical expressive vocabulary levels, their phonological abilities still lag behind those of their peers and they may struggle to retain statistical information, although certain key receptive abilities remain intact. Study 4 reports that although late talking children show deficits in symbolic understanding of pictures, their development in this domain follows a delayed trajectory, rather than one that is functionally different to typically developing children. The results also indicate that expressive and receptive language skills differentially support symbolic understanding of pictures, mediated by individual variation in social ability.

By examining language acquisition through typical and atypical development, this thesis aims to not only advance understanding of word learning as a process that inevitably involves, and makes use of, variation that exists in a child's environment, but also examines how expressive language ability – arguably the most clearly observable outcome of word learning for caregivers and early years professionals – interacts with how children come to understand the world around them.

“Helping a tiny baby to learn your language is like building a bonfire with words for twigs. Nothing happens for ages. You keep putting the bloody twigs on and trudging back and forth in a cold, damp field. You may have a faulty pelvic floor and much rather be watching something on the telly with a towel under your bum, but bonfires don’t build themselves, do they?

But there’s a problem. No matter how many words you pile on, nothing catches. At first, you try to build it properly, sentence by sentence, with full stops and proper pauses, but by the end, you’re just flinging random words on top of each other, sweating and slightly mad. You stand back. It’s taken more than a year or longer. You now have a huge pile of impressive but slightly useless wood. You try singing nursery rhymes to it, but it stares blankly back before doing a poo and crying.

You give up and are about to put the kettle on. Then you hear a roar and a crackle behind you. The fire has caught. Everything you piled on that bonfire, even the words you thought didn’t go in, is playing its part, burning brightly with the sheer exuberance of language. You stand back to bask in the heat and the magic and the wildness of the flames, rubbing your hands and telling all your neighbours: “Yep, I built that. Oh, it was nothing. Just love and patience, really.”

From that moment, the fire burns forever.”

- Skinner, N. 2016 Jan 9. When kids mangle language we all benefit. *The Guardian*.

1. Chapter 1: Word learning in typical development

1.1 How do children learn the meaning of words?

Between 12 – 24-months-old, children show rapid acceleration in their ability to comprehend and produce words (Fenson et al., 1994). A typically developing child may progress from understanding around 50 words at 12-months-old and saying one or two words, to understanding around 370 words and saying approximately 270 by 24-months-old (Hamilton et al., 2000).

By the time children have begun to learn words, they have already acquired a certain degree of knowledge around the speech sounds that make up their language and how to then extract words from speech streams (Werker & Yeung, 2005). However, in order to learn a novel word, the child must attach some meaning to a given label. This is not a trivial task; when a child hears a label for the first time, they do not necessarily know which referred-to item (*referent*) in the immediate vicinity is the correct item. This problem is perhaps best illustrated by Quine (1960), who described a non-native speaker and a native speaker witnessing a rabbit running past, and the native speaker utters ‘gavagai’. The non-native speaker is then faced with a problem of *referential ambiguity* (Markman, 1989): they do not know whether ‘gavagai’ refers to the rabbit itself (a small, herbivorous mammal), to what the rabbit represents (food, an omen, etc.), to the action the rabbit is performing (bouncing, running, etc.), or so on. Children appear to face the same problem when learning a novel word – it could refer to innumerable potential referents in their environment – but without a native language to build upon.

Despite this challenge, children show the remarkable ability to accurately learn a new word-referent pair in lab-based experiments following minimal exposure, through a process known as *fast mapping*. Carey and Bartlett (1978) demonstrated that 3-year-old children were able to quickly map a novel label (‘chromium’) to the correct colour (olive green) simply by being asked to retrieve ‘the chromium tray, not the blue one, the chromium one’. Similar studies have demonstrated that when children are shown novel and familiar objects, and request children to select both familiar and novel objects (*referent selection*), toddlers aged between 17 – 30-months-old have shown remarkable success in doing so (e.g. Golinkoff et al., 1992; Halberda, 2003; Mervis & Bertrand, 1994).

To explain how children manage to select the correct referent for a novel word so quickly, a number of theories have been proposed. These relate to the use of *lexical* principles, which are specific constraints that children appear to apply to a situation when learning new words (e.g. Golinkoff et al., 1994; Markman, 1989), and *socio-pragmatic* principles, which refer to more general mechanisms that may predispose children to learn words to begin with, such as joint attention and communicative abilities (e.g. Baldwin & Tomasello, 1998). More recently, *statistical learning* has also been proposed as a mechanism through which children resolve referential ambiguity. This broadly refers to the principle that learning can be driven by accruing data from the environment, for example about regularities in the speech steam, and deriving informative patterns (Romberg & Saffran, 2010).

1.2 Lexical principles

Lexical principles refer to learner-based assumptions that the child applies to a word learning situation that leads to them selecting the correct referent for a novel label. One of the most well-documented is *mutual exclusivity* (ME; Markman & Wachtel, 1988). ME proposes that every object has one label, and that when there are two objects and the label for one is known, children assume that the novel label must refer to the unknown object. Some research indicates that children apply this strategy from the early stages of word learning (Markman et al., 2003). A similar constraint is the novel name-less category constraint (N3C, Golinkoff et al., 1994; Mervis & Bertrand, 1994), which states that children will allocate a novel label to a novel object, rather than to an object which already has a known label. Others include the *whole object assumption*, whereby children assume a novel label refers to a whole object rather than just a part of an object, and the *taxonomic assumption*, the assumption that words belong to the same category (Markman & Wachtel, 1988).

Some of these constraints may share a common source. For example, the use of ME and similar mechanisms could fall under Barrett's (1978) *contrastive hypothesis* (Merriman et al., 1989), which proposes that children contrast negative examples of objects with positive ones, e.g. a ‘blue’ car can be identified by contrasting with cars that are *not* blue, such as ‘red’, ‘yellow’, and so on. It could also be argued that these mechanisms are derivatives of Clark's (1983; 1987) *principle of contrast*, which proposes that children assume each word holds a different meaning to other words, e.g. ‘blue’ is different to ‘red’.

However, a problem with lexical constraints is that it is not clear how they may interact with one another, nor is it clear how they scale up to the myriad of complexity that exists in a lexicon, such as taxonomic hierarchies, the different features of objects shared by the same label, synonyms, and so on. Furthermore, they assume that children have sufficient prior knowledge in order to apply them, but do not indicate where this prior knowledge originates from. By applying specific principles to complex behaviour, the isolated use of lexical constraints ignores the natural variability involved in building language, and the components that comprise the developmental abilities to apply such constraints (Deák, 2000).

Instead, these constraints may be part of a more general-purpose learning mechanism, rather than specific principles that only apply to word learning (McMurray et al., 2012). For example, Halberda (2006) argued that three main mechanisms – ME, contrast, and pragmatics – could be explained by a more globally applicable process-of-elimination strategy that goes beyond lexical processing (*disjunctive syllogism,* otherwise described as ‘A or B; not A, therefore B’). Similar to Barrett (1978), Halberda also proposed that negative evidence (e.g. knowing the name of a distractor) should be weighted more than positive (e.g.

N3C, or more generally, novelty) during novel word learning. Using eyetracking, Halberda showed that when faced with a familiar and novel object and asked to find their respect referents, both adults and children aged 3-4-years-old systematically disregard the known distractor before selecting the correct referent for a novel label. These results suggest that lexical constraints may actually be explained by more general cognitive mechanisms.

Furthermore, the use of lexical constraints by way of specific mechanisms or a general cognitive mechanism only explains how children might solve the referential ambiguity problem when one of the objects is already known. It does not explain how children manage to build a vocabulary over time, nor how children identify a correct word-referent pair when there is no familiar object to compare with the unfamiliar object.

1.2.1 If children use lexical constraints, when do they use them?

Lexical constraints were described as potential solutions for the referential ambiguity problem during fast mapping. However, as Carey and Bartlett (1978) stated, accurate initial selection of word-referent pairs does not necessarily reflect long term learning. Children in their study retained some, but not complete, knowledge of ‘chromium’ one week later, suggesting their knowledge of the word was fragile as well as subject to individual variation: 63% correctly identified the word-referent mapping, whereas some knew the referent had its own colour, but could not recall its unique label. Thus, fast mapping may only show the ability of children to correctly *identify* word-referent pairs, rather than the ability to *learn* them. When testing 24-month-olds, Horst and Samuelson (2008) found that children were able to correctly fast map novel words to referents , but did not retain the correct mappings just 5 minutes later. Similarly, Bion et al. (2013) found that 24-month-olds were able to correctly select referents during fast mapping tasks but could retain them, and retention of novel words was still fragile at 30 months. Vlach and Sandhofer (2012, 2014) have further demonstrated that both children (36 – 48-month-olds) and adults show poor retention of novel words when tested after delays of 1 week and 1 month. Thus, fast mapping does not necessarily equate to longer term learning, and lexical constraints alone cannot explain how children build vocabularies over time.

Rather, lexical constraints may be just one part of a much larger model of word learning. For example, some conceptualise learner-based constraints as prior probabilities that inform and interact with the present situation, as well as stored knowledge, to highlight the most likely solution (Xu & Tenenbaum, 2007). Alternatively, McMurray et al. (2012) in the *dynamic associative model* proposed that fast mapping reflects initial problem solving by the learner only, where selecting the correct word-referent pairs under referential ambiguity occurs as a ‘fast’, in-moment process. By contrast, longer term learning and retention of those words occurs as a ‘slow’ gradual process, where multiple instances are used to strength or weaken word-referent associations over time.

McMurray et al.’s (2012) model proposes that, during referent selection, lexical principles help to identify the relevant information. This process does not involve learning, but rather involves activating potential word-referent pair candidates that then compete for relevancy. This process can be modulated by paying attention to some candidates over others, e.g. through highlighting one particular pair over another. Over time and repeated exposures to word-referent pairs, the associations between words and lexical concepts are weighted using Hebbian learning (if a word and lexical concept are activated closely together in time, the connection between the two is strengthened; this can also be summed up by the principle: ‘neurons that fire together, wire together’; Hebb, 1949). Over time, spurious connections are pruned, whereas others are strengthened, and it is this process that builds eventual knowledge. According to McMurray et al.’s model, it is these associative processes that lead to longer term learning of correct word-referent pairings.

In sum, this research demonstrates that constraints may aid referent selection during fast mapping of novel words, but also that fast mapping a word is not equivalent to learning it. Rather, repeated exposures to word-referent mappings lead to retention and subsequent acquisition of those words. This process is inherent in cross-situational word learning, which instead of considering internal constraints, considers the environmental factors that may guide word learning.

1.3 Cross-situational word learning

Statistical learning has been applied to language acquisition over the last two decades, with research demonstrating that infants can apply these principles from a young age (Romberg & Saffran, 2010). Cross-situational word learning makes use of this concept. During cross-situational word learning, the referent for a word may be ambiguous on a single trial, but over several trials, this ambiguity can be narrowed by tracking which words and referents co-occur (Yu & Smith, 2007).

Yu and Smith (2007) were one of the first to apply cross-situational statistics to word learning in a lab-based experimental format. Adult learners were given a series of referentially ambiguous trials presenting novel objects with novel words, and were instructed to learn which words paired with which objects. The number of words and objects presented on each trial varied between two, three, or four. Their results showed that adults were able to make use of statistical co-occurrences across trials to correctly identify word-object pairs at test, even when four pairs were presented per trial, although their accuracy did drop at this higher level of referential ambiguity.

L. B. Smith and Yu (2008) further demonstrated that infants aged 12- and 14-months-old could also make use of co-occurrences between words and referents to facilitate word learning. Across 30 trials, infants saw two objects and heard two words, with each correct word-object pair occurring 10 times overall (totaling 6 word-object pairs to be learnt). At test, infants looked significantly longer at targets over distractors, indicating that they were able to use cross-situational statistics to identify correct word-referent pairs. Following this seminal study, both adult and child learners have demonstrated the ability to learn new words from cross-situational statistics (Bunce & Scott, 2017; Fitneva & Christiansen, 2011; Monaghan & Mattock, 2012; K. Smith et al., 2009; Yurovsky et al., 2013; Yurovsky & Frank, 2015).

One perspective is that competition plays a vital role in trial-by-trial learning. Under typical circumstances, the presence of several potential referents for a given label results in competition between potential word-referent pairs that helps learners to identify correct pairings. When each novel object has only one novel word associated with it, competition limits the mapping process – i.e. if object A and object B are presented with labels X and Y, and prior exposures to object A co-occurring with label X have yielded a robust association, label X will be designated to object B. This process therefore involves of the application of mutual exclusivity or a similar contrast within trials, but crucially, depends upon multiple occurrences across different trials. The use of competition in cross-situational word learning is thus both *local*, where possible word-referent pairs are weighted within a trial against each other, and *global,* where word-referent pairs are weighted across trials based on prior knowledge, allowing the stabilisation of word-referent pairings over time (Yurovsky et al., 2013). In short, whilst learner-based constraints such as the principle of contrast focus on a single naming event, cross-situational word learning relies on the aggregation of knowledge across multiple naming events.

Precisely how word learners utilise cross-situational statistics is subject to debate. One the one hand, the *associative learning theory* that where across multiple trials, learners assign certain weights to the associations between words and referents to converge upon correct word-referent pairs (MacWhinney, 2005; McMurray et al., 2012; Xu & Tenenbaum, 2007). For example, when testing cross-situational word learning in adults, Yu and Smith (2007) tested accuracy at the end of all trials, finding that learners were able to score well above chance despite referential ambiguity. They proposed that this must be because learners were able to keep track of multiple co-occurrences during learning.

Alternatively, the *hypothesis testing* theory argues that learning is the result of confirming or rejecting hypotheses about each word-referent pair on a trial-by-trial basis (Halberda, 2006; *'propose-but-verify'*, Medina et al., 2011; Trueswell et al., 2013). Here, accuracy of word-referent mappings is not the end result after accumulating all information over all instances as in the associative account, but rather, the result of candidate word-referent mappings that are examined dynamically at each instance. In a cross-situational word learning task, adult learners did not show evidence that they were tracking multiple hypotheses, but rather, identifying one word-referent mapping on one trial, then on the next, either confirming it discarding it (Medina et al., 2011). When using eyetracking, target looks only exceeded looks to distractors when they participant had been correct on the previous trial; where they were incorrect, the proportion of looking to target and competitors was similar (Trueswell et al., 2013).

Across a series of simulations, Yu and Smith (2012) proposed a model of word learning that showed how both hypothesis testing and associative learning could lead to correct word-referent pairs. Similar to McMurray et al. (2012), they theorised that lexical principles such as ME and the novelty bias could be integrated by constraining initial referent selection, but that hypothesis testing and associative learning made use of these constraints slightly differently, although with relatively similar results. For example, in Yu and Smith's (2012) hypothesis testing model, ME is added as a constraint to maintain the same level of certainty across hypothesised word-referent pairs – i.e. each referent has just one word – and pairs are deemed either correct or incorrect. With more and more trials, the model converges to confirm correct pairs, and disregard incorrect ones. However, their associative model accumulates evidence across many conflicting word-referent associations across trials, and therefore allows for effects of referential ambiguity. These potential associations are stored, and then at test, the word-referent pair with the strongest association is selected (also utilising Hebbian learning).

Both Yu and Smith’s (2007) hypothesis testing and associative models were able to complete a cross-situational word learning task to a comparable degree of accuracy when adding in familiarity and novelty biases. The difference between them was that associative learning favoured the *amount* of information, whereas hypothesis testing favoured the *kind* of information – preferring familiarity over novelty. Thus, where the two differ most appeared to be on decision-making based on retrieval of information; whereas hypothesis testing kept correct pairs on one list and incorrect on another, associative learning retained more information about statistics in the environment that could then be utilised to achieve more flexible decision making. Despite a preference for the associative model, they also urged a ‘*deliberate move away from the main theoretical question (…) being to show that one grand idea or principle beats another’* (p.34), and instead advocated for considering how information is selected on a trial-by-trial basis and how learning alters as a result of the task.

Overall, cross-situational word learning can offer a flexible model of word learning that may account for both short term and longer-term learning. In terms of mechanisms that underpin cross-situational word learning, both hypothesis testing and associative learning accounts have contributed to theoretical understanding, and increasingly, a dual approach is advocated for. Continuing evidence indicates that learners switch between the two strategies depending on how difficult the task is (Khoe et al., 2019; K. Smith et al., 2011; Yurovsky & Frank, 2015) or even depending upon the types of cues available (MacDonald et al., 2017). For example, Yurovsky and Frank (2015) identified that when there were many potential word-referent mappings, learners relied more on a single hypothesis and were not able to represent multiple candidates; however, when there were less candidates, learners relied more on multiple co-occurrences and were able to represent more potential word-referent pairs. They proposed that reliance on either mechanism was secondary to demands of attention and memory within the task itself.

Rather than splitting contributions to the field into two opposing camps, L. B. Smith et al. (2014) recommended examining theoretical commonalities including that word-scene co-occurrences have extractable structure, that statistical learning requires multiple co-occurrences, that word-referent pairs compete with one another, and that statistical models inherently encompass the process of learning more efficiently over time. Similarly, Roembke and McMurray (2016) advocate for an integrative approach to cross-situational statistical learning. Across a series of cross-situational word learning experiments with adults, they found evidence that learners both track and use associative information such as previous target location and gradual accumulation of evidence, but also make use of prior accuracy that falls under a hypothesis testing approach.

In conclusion, cross-situational models of word learning may be able to explain how children learn words across varying degrees of referential ambiguity. However, the aforementioned mechanisms tests word learning in largely adult populations in stable laboratory environments, and thus far have neglected the role of the caregiver and the social abilities of children that may also guide word learning. A potential limitation for cross-situational word learning is thus that it may ignore the context in which real infant word learning takes place.

1.4 Socio-pragmatic principles

Socio-pragmatic principles guiding word learning follow a more domain-general approach than lexical principles. Under socio-pragmatic accounts, children's success at word learning is predicated upon their socio-cognitive development and their interaction with the world (Baldwin & Tomasello, 1998; Tomasello, 2003). Tomasello (2003) describes two dynamic factors at play: the child's own developing socio-cognitive skills, and the impact of those that provide the socio-cultural context of word learning, such as caregivers.

A crucial socio-cognitive ksill that can support children's early word learning is *joint attention*, where caregivers and infants both concurrently attend and engage with the same object from approximately 9-months-old (Tomasello et al., 2005). Akhtar et al. (1991) found that joint attention in 13-month-olds correlated with productive vocabulary 9 months later. Tomasello and colleagues (Carpenter et al., 1998; Tomasello & Farrar, 1986) have also demonstrated that joint attention in infants aged 12-months-old correlated with their expressive and receptive vocabulary both concurrently and also three months later.

Joint attention provides the foundation for infants being able to infer the communicative intentions of adults (intention reading). In finding games where adults voice their intentions (e.g. 'I'm going to find the toma') before searching for a novel object, children aged 18- and 24- are able to learn the novel word-object mapping at test by following the adult's communicative intentions (Tomasello et al., 1996; Tomasello & Barton, 1994). Likewise, when the presence or absence of adults is manipulated relative to a target event, infants appear to take this into account. In Akhtar et al. (1996), 24-month-olds played with four objects; the first three were played with whilst the adults were present, and the last object – the target – was played with whilst the adults were absent. When all four objects were placed together, and the adult said 'oh look! A modi', children correctly selected the target object, suggesting that they were able to understand that the adult was identifying what was novel to them, rather than to the child (indicating that they could understand the adult's intentions; Baldwin & Tomasello, 1998).

Within the socio-pragmatic account, adults help form the context that allows a child to construct meaning. Under this account, referential ambiguity during fast mapping is not really a problem at all. Adults are pointing their children towards the correct referent to begin with, and providing the relevant word for the child, who is already focused on the referent (Nelson, 2007). Words are thus acquired through understanding the intentions of others, requiring only that caregivers communicate with their infants, and that infants are sensitive to this communication (Tomasello, 2003). Relative to retention, this theory does not explicitly distinguish between referent selection and retention of words in the way that associative learning models do (e.g. McMurray et al., 2012), but rather contends that common ground by way of shared intentions constrains the learning situation, and where common ground does not exist, children do not learn new words. Tomasello (2003) argues more generally that it cannot be simply that children learn the words they hear the most often, as they hear ‘the’ and ‘a’ very frequently but do not learn them early on. Rather, learning words happens as a byproduct of social interaction with adults, and develops on a timescale that is dependent on children’s developing socio-cognitive skills – children begin to learn words towards the end of the first year of life because they develop the skills required to understand intentions at this time.

However, others assert that there may be again more general cognitive mechanisms that explain these results without reference to socio-pragmatic theory. These concern basic attention and memory processes. Samuelson and Smith (1998) argued that the children in Akhtar et al. (1996) chose the correct novel object because the context was novel, not because they were able to infer that the adult was seeing the object for the first time. To demonstrate this, Samuelson and Smith made the target object contextually novel; the first three objects were dropped down a chute, whereas the fourth target object was played with in a different location. All four objects were then shown to the child, and the experimenter announced ‘there’s a gazzer in there’. Children selected the target object under contextual novelty just as frequently as they did in Akhtar et al., suggesting that the key feature for children was the saliency of the fourth object being contextually different, which could be explained by memory and visual attentional processes, rather than by inferring the speaker’s referential intent.

Furthermore, some contend that it is not joint attention, but rather *sustained attention* – which does not necessarily rely on socio-pragmatic factors – that is responsible for longer term learning of words. Using head-mounted eye-tracking for both infants and adults, Yu et al. (2019) measured eye gaze to determine joint and sustained attention during toy play with

9-month-olds and their caregivers, measuring vocabulary at 12- and 15-months old. Their results showed that children's sustained attention to an object, rather than joint attention with caregivers, predicted larger vocabularies. They suggested that previous work has conflated the two, as they are correlated with one another, but that joint attention supported sustained attention to objects, rather than the other way around. This would indicate that intention reading on the part of the *infant* is not the key factor in labelling objects correctly, but rather, the infant maintaining attention on objects, supported by *caregivers* being able to read infants intentions to provide the right label at the right time, is the most important factor.

Finally, both infants who have not yet developed joint attention or intention reading abilities, and children with autism who typically have trouble with socio-pragmatic skills, are still able to learn words and to communicate. This indicates that there cannot be just one route to language. For example, infants as young as 6-months-old are able to communicate using gestures to represent items before they are able to utilise joint attention mechanisms, which typically start at around 9-months-old (Johnston et al., 2005), and even show evidence of fast-mapping (Friedrich & Friederici, 2011). Children with autism spectrum disorder, despite struggling with language, are also still able to learn words and in some cases have not shown functional differences in how they use word learning mechanisms such as referent selection and the formation of cross-situational associations (e.g. Hartley et al., 2020; Luyster & Lord, 2009). Thus, socio-pragmatic accounts alone cannot provide a comprehensive explanation for children's language acquisition. Furthermore, they do not provide a detailed account of how children retain words they have learnt over time.

1.5 Multiple cue models of word learning

Ultimately, it would seem logical that if single accounts cannot comprehensively explain word learning, then combining multiple accounts might represent a solution. However, fitting these accounts together is difficult. For example, if we accept McMurray et al.'s (2012) conception of word learning as being distinctly split into referent selection (that occurs through online competition of candidate word-referent pairs) and retention (aided by cross-situational word learning and association over time), then how might socio-cognitive principles fit in? Even if socio-pragmatic theories cannot explain every instance of word learning, Tomasello and colleagues (2003) have demonstrated that children are sensitive to social cues from an early age and do use them to learn. How might these social cues interact with statistical learning? Additionally, when there is very high referential ambiguity – such as multiple referents for a given word, or even multiple words with multiple referents – how can a learner identify the correct word-referent pair based on general purpose learning strategies or purely associative learning alone?

Hollich et al. (2000) were amongst the first to describe language acquisition as the result of multiple sources of information that included attention, social, and linguistic factors in the *Emergent Coalition Model* (ECM). They argued that humans as learners are adaptive and resourceful, and likely to make use of multiple cues in multiple ways. In a series of experiments, 12–24-month-olds were exposed to novel words and objects and tested on word-referent mapping, manipulating perceptual salience (testing preferences for an exciting, brightly-coloured object, or a dull object) and social cues (testing preferences for one of the two objects depending on which the experimenter was looking at). Crucially, infants of all ages showed an awareness of attentional, social, and linguistic cues even when they did not use them to directly map a label to an object, indicating that even as early as 12-months, children are sensitive to multiple cues in their environment. Their preferential use of the cues, however, was dependent on their age. Infants aged 12 month relied more on perceptual salience, looking longer at the bright object at test even if the dull one was labelled, whereas those aged 19- and 24-months-old relied more on social cues (e.g. looking longer at the dull object when it was labelled). Thus, Hollich et al. (2000) demonstrated not only that different cues could be used, but that the way in which they were used depended on the age of the children in question.

However, the ECM does not account for statistical information that may be in the environment, as Hollich et al. (2000) only accounted for how social cues could be integrated with lexical principles, and largely focussed on children aged 12-months-old and above. As infants show the ability to extract and learn from statistical regularities in their native language from as early as 7 – 8 months (Saffran et al., 1996), this represents a potential limitation in how widely the ECM can be applied to word learning.

In a computational model, Yu and Ballard (2007) combined specific social cues with cross-situational word learning. They focused upon joint visual attention and prosodic cues (quantified by voice pitch) in child-mother interactions and used them to assign weights to words and referents within the framework of a statistical learning model. This combined model and a model of statistical learning only were tested on precision (percentage of words spotted by the model which were correct) and recall (percentage of correct words that the model actually learnt out of all words expected to be learnt). A model combining attention and prosodic cues with cross-situational statistics outperformed a model of statistical learning alone (83% precision and 77% recall, versus 75% precision and 58% recall). These results indicate that both attentional and social cues can contribute to facilitate accurate cross-situational word learning.

It is largely accepted that children learn words under variable input, and yet many models of word learning consider the environment to be relatively stable. In recognition of this, Monaghan (2017) proposed the *Multi-modal Integration Model* (MIM) as a computational approach to integrating multiple cues with cross-situational word learning. Unlike previous models, rather than characterising the instability of the child's environment as a barrier to learning, the MIM suggested this variability might actually aid more robust learning, with the interplay between different cues being crucial to how we end up with relatively consistent results in language. Beginning with the recognition that language otuput is broadly similar despite variation in the environment, Monaghan proposed a mechanism borrowed from genetics – canalisation – which posits that greater interaction between multiple sources of information (similar to genes) yields narrower and more stable outcomes (phenotypes).

Testing this theory, the MIM investigated the use of gesture, distributional cues (grammatical consistencies across the language, such as articles preceding nouns), and prosody during cross-situational word learning in a computation model. The MIM demonstrated that combination of these cues boosted the model's learning. The model also showed that variability of the cue was important; when cues were always present, the model was brittle and prone to error, but when cues were sometimes present and other times absent, the model's learning was far more robust. Similarly, cues do not always occur with perfect reliability in natural language learning – sometimes adjectives might precede a noun rather than an article, other times the stress of a word may not indicate the target, and so on. If a learner only learnt a word-referent mapping when a pointing gesture was there, any subsequent situations without the gesture cue would result in errors at test. The model itself was tested in an adult word learning study (Monaghan et al., 2017), which found that participants scored most accurately at test when a pointing gesture appeared 75% of the time during training for novel words.

Overall, multiple cue models that combine cues that children attend to with statistical information gleaned from the environment may go some way to explain how children end up with broadly similar language abilities, despite the vast amount of variation in the input. However, the way in which these cues interact with cross-situational word learning during within-trial learning, and how such principles work in experimental child studies, requires further examination. In particular, the first part of this thesis focuses on the role of gestures as cues in word learning. The following section presents a brief overview of the role of gestures in language development, with a consideration of how and which type of gestures work as cues to meaning during word learning.

1.6 - The role of gestures in vocabulary development

Gestures are an integral aspect of how children interact with caregivers and the world at large during language acquisition (Iverson & Goldin-Meadow, 2005; Southgate et al., 2007), aiding effective communication when verbal ability has not yet been fully realized (O'Neill, 1996). In particular, gesture use appears to be facilitative of vocabulary development, with increased child gesture use predicting larger future vocabulary size (Brooks & Meltzoff, 2008; Fenson et al., 1994; Rowe et al., 2008). Caregiver gesture use also appears to be predictive of child gesture use (Rowe et al., 2008), and can help highlight referents during word learning (Cartmill et al., 2013; Iverson et al., 1999). The *quality* of gesture also contributes to word-referent mappings. In the Human Simulation Paradigm (Cartmill et al., 2013; Medina et al., 2011), adult participants are asked to guess words from muted videos of parent-child interactions, providing a measure of caregiver input quality. Over half of the 'high quality' vignettes (where participants guessed words to a high degree of accuracy) involved caregivers using gesture close to the onset of the mystery word (Cartmill et al., 2013). Children of parents who offered higher quality input at 14-18 months of age also had higher receptive vocabulary at 53 months of age.

Precisely how gesture contributes to vocabulary development in children remains uncertain. Children pair gestures with words before they begin to produce two-word combinations (Iverson & Goldin-Meadow, 2005). O'Neill (1996) found that in a toy retrieval task, 32-month-olds preferred to use gesture to indicate the location of the toy despite being able to name the locations. Fenson et al. (1994) postulated that gesture use in infants might serve as a bridge between between passive comprehension of words (receptive vocabulary) and active participation in producing them (expressive vocabulary) during the process of language acquisition. This idea was echoed by Goldin-Meadow (2000, 2007), who described gesture as a way to bridge between concepts that cannot be expressed in speech during both language acquisition and in learning more generally. This may occur through shifting verbal cognitive load to a visuospatial modality instead (Goldin-Meadow & Wagner, 2005). Similarly, in a study that manipulated homonyms (two words that sound the same, but have different meanings, e.g. ‘glasses’ meaning both spectacles and drinking receptacles), children aged 4-5-years-old used gestures to help distinguish between the two meanings under these ambiguous conditions (E. Kidd & Holler, 2009).

Thus, gesture has a vital role in language acquisition and learning from the child’s perspective. However, the nature of the relationship between gesture and vocabulary development may derive from the informative role of gestures in word learning during active communication between parent and child. Gesture use by caregivers may provide valuable information about intended referents during rapid vocabulary development. Some evidence indicates that once verbal input is accounted for, parent gesture does not correlate with child vocabulary scores, suggesting that the value of caregiver gesture use may be embedded in the information it provides simultaneously with speech (Iverson et al., 1999; Pan et al., 2005; Rowe et al., 2008). The value of gestures may be in presenting visual information that is not present in speech, such as the hand action of a bird’s wings flapping when saying the word ‘eagle’, or by reinforcing what is said during speech, such as pointing at an intended referent whilst naming it (Goldin-Meadow, 2000; Goldin-Meadow & Wagner, 2005).

Infant gesture appears to predict language development, and gesture use in children appears to be related to parental gesture use (although the majority of studies focus on infant gesture during language acquisition, rather than parent gesture). A longitudinal intervention study with infants aged 11 months to 36 months found significantly higher receptive and expressive child vocabulary in a gesture-trained parent group at endpoint compared to a control group (Goodwyn et al., 2000), although these results have not been replicated when methodological improvements were made (Kirk et al., 2013). LeBarton et al. (2015) found that training infants to use gesture in a shorter 8-week intervention study led to increased parent gesture use, and that infant gesture correlated with increased child speech during spontaneous interactions with caregivers. However, increased parental gesture did not relate to increased child speech. Taken together, these results suggest that although caregiver and infant gesture use are linked, enforced caregiver gesture use may not correlate with longitudinal vocabulary development. While these studies do not demonstrate whether parental gesture has an effect during immediate word learning itself, they do, however, suggest that caregiver gesture use can be manipulated.

Caregiver gesture as a cue for word learning may help delineate correct word-referent pairs. For example, in Monaghan (2017), pointing gestures were used to highlight correct referents for given words. Similarly, although Yu and Ballard (2007) did not use body movements such as hand gestures as part of their model, they did envisage these as part of their larger model for combining social cues with cross-situational information.

Gesture itself then appears to be a useful candidate for determining how multiple cues may interact with statistical information, being both correlated with – and perhaps even underpinning – language development, but also as a way to highlight specific referents during word learning itself.

Deictic gestures as cues for word learning

Gestures come in many different forms and, subsequently, have a multitude of classifications. Very broadly speaking, one convention is to divide gesture into those that are deictic (highlighting attention by showing, giving, or pointing) and those that are representational (sometimes known as symbolic or iconic; gestures that represent features of a referent, e.g. flapping arms to indicate a bird; Capone & McGregor, 2004; Rowe et al., 2008). Deictic gestures precede the development of representational gestures, with infants using them from around 9-months-old, and being able to reliably follow adults' pointing from approximately 12-months-old (Carpenter et al., 1998).

Deictic gestures such as pointing can serve as a useful cue to disambiguating meaning when used by caregivers. Iverson et al. (1999) found that mothers used pointing gestures in 15% of word-learning exchanges with their infants to delineate a target object. When coding a corpus of mother-infant interactions for social cues highlighting referents within discourse, Frank et al. (2013) found maternal pointing was highly precise in predicting object reference, indicating that pointing is a reliable cue to meaning. Goldin-Meadow (2007) also emphasise the role of pointing gestures in specifically highlighting referents, describing the use of pointing combined with representational gestures in a string of gestures similar to spoken sentences in users of home sign language. Within these gesture sentences, pointing gestures act as nouns and pronouns referring to specific objects.

One avenue for debate concerns whether or not the mechanisms that underlie the use of pointing gestures as word learning cues are socio-pragmatic or attentional, echoing the arguments around the different types of constraints in language acquisition. Although it is not the aim of this thesis to advance this debate, a brief overview is presented here.

Proponents of a socio-pragmatic approach focus on pointing gesture use by infants, and argue that children point to influence the mental state of the caregiver (Tomasello et al., 2007). Under this interpretation, pointing at an object provides a frame of reference for joint attention, and is thus linked to understanding the recipient's intentions. For example, infant gesture may be interrogative in nature, acting as a 'signal' to caregivers to instigate verbal input so the infant may gain critical information about a specific object (Iverson & Goldin-Meadow, 2005; Southgate et al., 2007). Moreover, caregiver pointing is useful because children interpret it as a signal for sharing experience. For example, Liebal et al. (2009) tested 18-month-old infants on their ability to respond to an experimenter's pointing gesture. Children played a puzzle game with a first experimenter who, at the end of the game, highlighted one puzzle piece was missing (the target) before leaving the room. A second experimenter then entered to play a separate 'clean-up' game using an identical puzzle with the child, placing all pieces into the basket, except for the missing piece. The missing target piece was then placed on the floor out of sight of the child. At test, the first experimenter re-entered, and either the first or the second experimenter pointed at the target piece saying 'oh, look!'. They found that when the first experimenter pointed, children put the target piece with the puzzle and completed it, but when the second experimenter pointed, children put the target piece in the clean-up basket. This result indicates that children used their previous shared experience to interpret the caregivers' pointing gestures; the first experimenter wanted to complete the puzzle, whereas the second wanted to tidy up the puzzle.

However, others might argue that pointing gestures support word learning in much the same way that any non-social cue that increases the attentional saliency of an object. L. B. Smith (2000) describes word learning generally as a process that involves associating the most perceptually salient object with an auditory label and, using head-mounted cameras, has shown that infants tend to restrict their visual field to single objects at a time (Pereira et al., 2014). This might indicate that basic perceptual processes are responsible for how cues restrict learning. Under this account, pointing would simply highlight the saliency of an object without the need for an infant to infer any intentionality behind the point, similar to an arrow cue. For example, some literature suggests that arrows and eye-gaze similarly visually orient towards objects as a result of general associative or automatic mechanisms that respond to directional information, rather than because information is socially relevant (e.g. Brignani et al., 2009; G. Kuhn & Kingstone, 2009).

Others propose that attentional processes are key to highlighting correct word-referent pairs. For example, Horst and Samuelson (2008) and Axelsson et al. (2012) argue that the most important role of cues in word learning is the attentional highlighting of a target object during ostensive naming and the simultaneous dampening of competitors, allowing for effective encoding of word-referent associations. In Axelsson et al. (2012), children aged 24-months-old were tested on referent selection and retention. Children were asked to select referents in all conditions, but feedback on their selection was given afterwards by illuminating the target, covering distractors, or both, as compared to where children's attention was directed by the use of a pointing gesture. In retention trials, children in the pointing condition did not score above chance, whereas they did in other conditions. They concluded that attention-directing feedback led to more accurate word learning than pointing cues did.

In sum, there are two main possibilities for why caregiver pointing gestures may be useful during word learning: because infants (and caregivers) are socially motivated, or because infants are perceptually cued towards things that attract their attention. It is even possible that the nature of pointing gesture cues does not have to be mutually exclusive, and that a pointing gesture can be both a social cue *and* an attentional cue. In some cases, pointing may tell us something deeply profound about how humans communicate (Tomasello et al., 2007); in others, the task at hand may be achieved by using a light (Axelsson et al., 2012). Here, the crucial point is that children *can* and *do* make use of pointing gesture cues to identify referents. What remains to be investigated is how learners make use of pointing gestures during within-trial learning as the process unfolds, and whether caregivers adapt their gestures to the environment to facilitate their child's word learning.

1.7 Summary and thesis outline

Word learning is a multi-facted process, and a wide variety of theories have attempted to explain how children manage to learn words with apparent ease within typical development. More recent models of word learning suggest an integrative approach may best characterise the complexity that stems from not only multiple sources of information, but multiple ways in which they may interact with each other, with the surrounding environment, and also with the age and general developmental abilities of the child in question.

However, to understand precisely how such models work in practice, closer attention must be paid to, firstly, the nature of interactions between cues and environmental variability and the subsequent effect such interactions have on children's accuracy in word learning and, secondly, how cues are integrated with the information present during in-moment learning itself.

The first part of this thesis makes use of cross-sectional studies across multiple methods, combining eye tracking methods that allow for assessing the temporal dynamics of learning, with computational methods that allow the precise control of parameters, and also experimental methods that highlight the need for any theory to be tested in practice. The first two papers focus on gesture as one of multiple cues that can contribute to word learning. The first paper reports a computational model of gesture and word learning across referential ambiguity (Chapter 2), and tests the model's predictions in a behavioural study of word learning 18-24-month-olds and their caregivers. The second paper (Chapter 3) reports three experiments in adult learners, where referential ambiguity and both the timing and presence of a pointing gesture cue is manipulated, testing both effects during training using an eyetracker and effects at test.

However, to gain a comprehensive understanding of how children learn words, the effects of atypical development must be considered alongside what happens during typical development, as well as how word learning mechanisms interact with vocabulary over time. The second part of this thesis thus concerns a longitudinal study of late talking and typically developing children, beginning with a review of the late talking literature (Chapter 4). The third paper (Chapter 5) examines word learning mechanisms in the cohort and tests whether performance on different word learning tasks can predict both late talking status and later expressive vocabulary outcomes. The fourth paper (Chapter 6) demonstrates how expressive delay can affect other areas of development, by testing how language delay can affect symbolic understanding through the use of a picture comprehension task.

To gain a comprehensive understanding of how children learn words requires not only a consideration of typical and atypical development, but also requires a multi-disciplinary approach. Overall, this thesis heeds the words of Hollich et al. (2000, p.14), who state: ‘*without recognizing the enormity of the word learning problem a theory cannot support the weight of lexical acquisition (…) Just as a one-legged table is inherently unstable, scientific explanations of complex process that force either/or decisions are not as powerful as those that embrace different perspectives*.’

2 Chapter 2: Caregivers use gesture contingently to support word learning

2.1 Chapter introduction

The Multimodal Integration Model (MIM, Monaghan, 2017) highlighted how variability in cue availability might support word learning, and even lead to more robust learning that is less prone to error. In particular, gesture was found to be a useful cue when the MIM was tested in adults (Monaghan et al., 2017), and gestures in general support children's word learning and vocabulary development (Rowe et al., 2008). What we do not know is whether caregivers are sensitive to how gesture can support word learning, and whether this depends on the amount of variation in the environment, such as the number of possible referents for a novel word. We also do not know how children might respond to variability in both the gestures that caregivers provide during word learning, and the variability in the number of potential referents for a given word.

This paper tests caregiver response to referential ambiguity by first testing the MIM with a pointing gesture cue across conditions, where the number of potential referents for a word differs. The predictions of this model are then tested in children aged 18–24-months-old and their parents during a word learning task, tracking the gestures that caregivers make during training and children's subsequent word learning accuracy.

Author contribution for Chapter 2: *Rachael W Cheung:* design, data collection (behavioural), analysis (behavioural), writing, review. *Calum Hartley:* design, review. *Padraic Monaghan:* design, analysis (computational modelling), review

Published as: Cheung, R.W., Hartley, C., Monaghan, P. (in press) Caregivers use gesture contingently to support word learning. *Developmental Science*, e13098, https://doi.org/10.1111/desc.13098

2.2 Abstract

Children learn words in environments where there is considerable variability, both in terms of the number of possible referents for novel words, and the availability of cues to support word-referent mappings. How caregivers adapt their gestural cues to referential uncertainty has not yet been explored. We tested a computational model of cross-situational word learning that examined the value of a variable gesture cue during training across conditions of varying referential uncertainty. We found that gesture had a greater benefit for referential uncertainty, but unexpectedly also found that learning was best when there was variability in both the environment (number of referents) and gestural cue use. We demonstrated that these results are reflected behaviourally in an experimental word learning study involving children aged 18-24-month-olds and their caregivers. Under similar conditions to the computational model, caregivers not only used gesture more when there were more potential referents for novel words, but children also learned best when there was some referential ambiguity for words. Thus, caregivers are sensitive to referential uncertainty in the environment and adapt their gestures accordingly, and children are able to respond to environmental variability to learn more robustly. These results imply that training under variable circumstances may actually benefit learning, rather than hinder it.

2.3 Introduction

Word learning is a complex process, requiring children to individuate words from continuous speech and pair them with intended referents in the environment. However, there are multiple possible references within multiword utterances (Monaghan & Mattock, 2012; Yu & Ballard, 2007) and multiple potential referents in the environment for each word (Quine, 1960; Siskind, 1996; L. B. Smith & Yu, 2008). Although internal constraints may aid special cases of language acquisition (Carey, 1988; Golinkoff et al., 1992; Markman & Wachtel, 1988; Mervis, 1987), alternative accounts have explored how constraints present in the environment can be utilised by more general purpose learning mechanisms.

The environment contains multiple sources of information that can help to constrain word-object mappings. This includes cross-situational statistics, where possible links between words and referents may be resolved by tracking co-occurrences between them across multiple situations (Siskind, 1996; L. B. Smith & Yu, 2008). Other cues include prosody, such as the referring word having the highest amplitude (Fernald & Mazzie, 1991), and distributional information from syntax, such as nouns and verbs being preceded by frequently-occurring articles (Fries, 1952; Mintz, 2003; Monaghan et al., 2007). Gestural cues also contribute vital information, forming an integral part of communication from early infancy (Iverson & Goldin-Meadow, 2005; Southgate et al., 2007), and helping caregivers delineate referents during word learning (Cartmill et al., 2013; Iverson et al., 1999).

Despite huge environmental variation across learning situations, word learning studies generally assume a relatively stable environment for children (McMurray et al., 2012; Yu et al., 2012). Importantly, this variability may actually be useful. In a computational model of word learning, Monaghan (2017) developed the multimodal integration model (MIM; A.C. Smith et al. 2017) to explore the role of multiple cues – distributional, prosodic, and gestural – in supporting language acquisition. The model was trained to learn word-object pairings when words and objects were presented among multiple possibilities and when cues were present or absent. Although learning benefited from all cues, learning was more efficient and more accurate when cues occurred 75% of the time, rather than when they were present 100% of the time (Monaghan, 2017). This was confirmed in behavioural studies with adults (Monaghan et al., 2017). The MIM showed that multiple cues support learning over single cues, and that the model learnt most robustly when the cues were individually variable. This prevented the model from relying too heavily on single cues in the environment, akin to dropout training, in which input units are stochastically dropped to improve model generalisation and avoid overfitting (Srivastava et al., 2014). In the MIM, the existence of variability within the environment itself circumvents the requirement for this to be incorporated into the learner, providing the necessary degree of dropout to maintain the learner's sensitivity to multiple cues in the environment. These results indicate that although word learning occurs in noisy contexts with multiple, variable cues, learners are able to make use of this variability to benefit learning.

However, the MIM did not test the extent to which variability in cues may be contingent on the informational content of situations. For instance, when there is only one possible referent in the environment, gesture may be redundant. Alternatively, when there are many possible referents, gesture may be crucial. Thus, during learning situations, if the speaker is sensitive to this environmental ambiguity, we may see cues deployed differently according to the situation.

Speakers adjust their prosody, syntax, word selection, and phonology according to context and the listener's perspective (Brown-Schmidt & Duff, 2016; Gorman et al., 2013), and children also adapt speech and gesture according to the perspective of adults (Bahtiyar & Küntay, 2009; Bannard et al., 2017; Nadig & Sedivy, 2002; Nilsen & Graham, 2009; O'Neill, 1996). In contrast, how caregivers adapt to the environment is less established. Caregivers demonstrate patterns of behaviour when labelling objects that align with children's internal constraints, such as naming whole objects rather than parts (Masur, 1997), or using one label per object, encouraging mutually exclusive labelling (Callanan & Sabbagh, 2004). Caregivers also adjust how they use labels according to their child's knowledge (Luce & Callanan, 2010; Masur, 1997); for example, by placing unfamiliar nouns and verbs saliently in an utterance and physically presenting unfamiliar objects more clearly (Cleave & Bird, 2006). However, these adaptations depend on perceived levels of familiarity in the child, rather than perceived uncertainty in the environment when the level of familiarity is consistent (such as when all objects are novel). These studies show that caregivers are sensitive to the informational content of cues relative to their child, but whether this sensitivity exists when environmental variability itself is manipulated has not yet been tested.

Gesture offers a prime candidate for further exploration of how caregivers might adapt contingently during word learning. Not only is gesture facilitative of vocabulary development, with increased early child gesture use predicting larger future vocabulary size (Brooks & Meltzoff, 2008; Fenson et al., 1994; L. J. Kuhn et al., 2014), but caregiver gesture use can predict early child gesture use (Rowe et al., 2008) and offer highly valuable information for word-referent mapping (Cartmill et al. 2013). Caregivers also alter gestures according to whether an object is familiar to their child as well as present or absent (Vigliocco et al., 2019), and in response to increased task complexity when communicating with children with delayed language development (Wray & Norbury, 2018).

The types of gestures produced by caregivers and children are rich and varied (Capone & McGregor, 2004; Özçalışkan & Dimitrova, 2013). They may occur in isolation or combined with speech, providing information that may overlap, complement, or even mismatch speech content – all of which offer valuable communicative insight (Goldin-Meadow & Wagner, 2005). Yet, when faced with high referential ambiguity during word learning, the most informative caregiver gestures may be those that clearly delineate the target of a novel label. Children follow deictic gestures such as pointing from approximately 12-months-old (Carpenter et al., 1998), and caregivers also use deictic gestures more than other gestures with children under 22-months-old (Özçalişkan & Goldin-Meadow, 2005). Whether caregivers alter these useful gestures based on the presence of environmental referential ambiguity remains unexplored.

In this paper, we examined how environmental variability might affect word learning by testing the contingency of caregiver gesture use to support word learning under referential uncertainty. We first adapted an established computational model of word learning (MIM; Monaghan, 2017) to test the benefit of contingent gestural cues for word learning when the number of possible referents for speech varies. We then conducted a behavioural study to determine whether caregivers varied in their gesture use when teaching novel words under different degrees of referential uncertainty, and whether the predictions of the computational model for optimal behaviour are exhibited in naturalistic exchanges. We thus considered the presence and interaction of two distinct aspects of variability: *referential uncertainty*, conferred by differing numbers of potential referents for a given word, and the *availability of gestural cues*, with their role determined firstly by altering the occurrence of such cues systematically in a computational model, and then by examination of naturally-occurring differences in caregiver cue use during a behavioural study.

2.4 Computational model

We adapted Monaghan's (2017) implementation of the MIM by varying the number of possible referents in the visual field during training to test the effect of environmental indeterminacy on cue influence. Monaghan's (2017) implementation is an adaptation of A.C. Smith et al. (2017), and simulates word learning via acquiring the correspondence between one of several words heard in an utterance and one of several objects in the environment. The model is a neural network that learns through backpropagation, operating on principles of acquiring associations between representations. The MIM is similar in principle to other associative models of word learning (e.g. McMurray et al., 2012; Yu & Smith, 2012), but extends these to test multiple cues in the child's immediate environment that provide information about the intended reference of speech. Our aim in this paper is to examine how such a simple associative learning system might respond to variation in environmental cues in terms of how associations between words and objects cohere.

We trained and tested the MIM (Monaghan, 2017) under three conditions that allowed us to investigate the effects of a gestural cue on learning during: 1) a condition with no referential ambiguity, where the object presented must be the target (one object); 2) a condition with some referential uncertainty, where one object was the target and one was the foil (two objects); and 3) a condition with a higher degree of referential uncertainty, where one object was the target and there were five foils (six objects). Enumeration tasks suggest that observers are able to rapidly report the numbers of objects in a visual display between one to four objects with ease; however, above four, they switch to slow counting of individual objects (Cowan, 2001; Xu & Chun, 2009). Thus, our aim was to crowd the visual display in the six-object condition.

An increase in potential referents for a given novel word has led to less reliable learning in behavioural studies (K. Smith et al., 2011; Trueswell et al., 2013). We therefore predicted that the model would learn more quickly from the one-object than the two-object condition, which in turn would be learned more quickly than the six-object condition. We also predicted that the effect of the gestural cue would be largest when there were two objects compared to one, and six objects compared to two: as indeterminacy of the intended referent increases, gesture may become more important to support and constrain word-referent mappings.

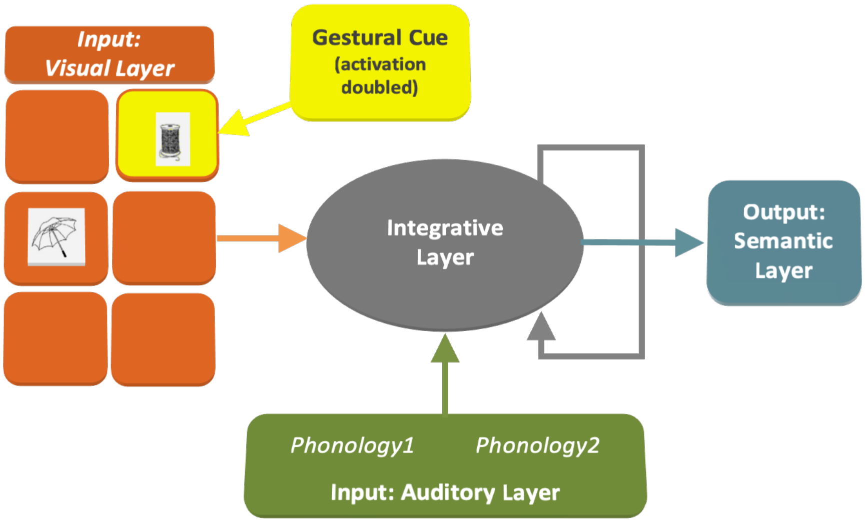
2.4.1 Method

Architecture

The model's architecture is shown in Figure 1. The model had an auditory input, comprising 80 units, where sets of spoken words were presented, and an 80-unit visual input, where sets of objects were presented. Each unit in the auditory and visual inputs was capable of representing one piece of information (i.e. a phoneme feature within a word, or a visual feature of an object). Input from these auditory and visual inputs projected to a central integrative layer of 100 units, each of which combined and processed input from the set of auditory and visual inputs. This integrative layer was self-connected, and was also connected to a semantic output layer comprising 100 units, where the model had to generate the meaning representation of the target word-object pairing.

For the current simulation, we expanded the number of objects that could appear in the visual input from two (as in the original simulation; Monaghan, 2017) to six. For the one-object condition, the object could appear in any of the six possible object locations. For the two-object condition, any two of the six locations presented the objects. For the six-object condition, one object appeared in each of the six locations. The model was otherwise identical to the original simulations.

Figure 1. Architecture of the multimodal integration model (MIM) for word-object mapping (example of two-object training condition with gesture cue present).



Representations

The auditory, visual, and semantic representations for each word-object mapping were identical to Monaghan (2017).

When the gestural cue was present, the activation of the target object's location was doubled, enhancing the influence that the visual features of the object in that position had on the model's learning. The role of gesture was thus implemented as increasing the salience of one position in the visual display of the model, and the effect of gesture is akin to increasing attention to a region of visual space, as implemented of visual processing in dynamic systems models (Samuelson et al., 2017). Across simulation runs, we varied the availability of the gestural cue by altering its presence across individual trials, where the cue was present 0%, 33%, 67% or 100% of the time. For example, in the 33% gesture cue availability condition, there was a 1/3 chance for each trial that the cue was present.

For each simulation, there were 100 word-object mappings to be learned, with the auditory and visual representation of each word-object mapping randomly generated for each simulation run.

Training

The model was trained to learn correspondences between 100 spoken words and 100 visual objects through cross-situational statistics.

For each training trial, the model was presented with two auditory words – one corresponded to a visual object appearing in the visual input, and the other was randomly selected from the other 99 words. The model was required to produce the semantic representation corresponding to the overlap between the target word and target object at the output.

For the one-object condition, only the target object corresponding to one of the spoken words was presented. For the two-object condition, two objects were presented – one corresponding to one of the spoken words and the other randomly selected from the other 99 objects (but not corresponding to the other, foil word). For the six-object condition, five foil objects were selected. In all conditions the positions of objects were randomised. For the one-object condition, the target object appeared in one of the six locations, and the other five locations were empty. For the two-object condition, the target and a foil object appeared in random locations across the six possible positions. For the six-object condition, the target object appeared randomly in one location, and five other foil objects filled the five remaining locations. The gestural cue was present for either 0%, 33%, 67%, or 100% of the individual trials in each condition.

Activation in the model passed between layers for five time steps. At time 1, the auditory and visual input was presented to the model. At time 2, the activation from these input layers reached the integrative layer. At time steps 3 to 5, the model was required to produce the semantic representation for the word-object pairing, with recurrent activation cycling through the integrative layer's self-connections and from the integrative layer to the semantic output layer. At the end of each training trial, the model's error was calculated across the semantic output layer as the cross-entropy error of the difference between the model's actual activation of units and the target activations. Connections were adjusted between units in the model according to the backpropagation through time learning algorithm (Pearlmutter, 1989). The model's connections were initially randomised in the range [-0.1, 0.1], and the learning rate was set at 0.01.

After 1000 learning trials had been presented to the model, its performance on each of the 100 word-object mappings was tested. The model was judged to be accurate if it produced a semantic representation closer to the target than to any of the other 99 semantic representations. The point in training at which the model was able to identify 95% of the word-object mappings correctly in four consecutive tests was identified as reflecting the ease of the model's ability to learn the words. If the model failed to learn by the end of training, then the end of training was taken to be the length of training time. Training finished after 100,000 learning trials had been presented to the model, and then the model was tested.

We formulated 10 different versions of the training patterns. For each training pattern, we ran 12 different versions of the model, with different randomised starting weights, different gesture cue availability, and a different number of objects during training. In total, there were 120 simulation runs: 10 versions of pattern x 4 gesture cue availability (0%, 33%, 67%, and 100%) x 3 numbers of objects (1, 2, and 6). We treated each of the 10 different versions of the training patterns as a separate subject during analysis, and treated gesture cue availability and number of objects as within-subject variables.

Testing

The model's ability to accurately detect the word-object mapping for each of the 100 pairings was tested under different conditions than its training: the model was tested instead where the target object appeared along with two other foil objects (simulating a three-alternative forced choice test). To assess the robustness of learning, we also determined whether the model could identify the target pairing without any gestural cue being present. The model's accuracy was determined in the same way as during training: if it produced a semantic representation closer to the target than to any of the other 99 semantic representations.

Data, code, and models run are available on the Open Science Framework (OSF) (http://osf.io/6frcw/?view_only=72344789a6294aa19d63a8bd93a628f3).

2.4.2 Results and Discussion

Length of training

Figure 2A shows the time taken for the model to identify 95% or more of the word-object patterns in four consecutive tests. Additional simulations that were trained to a lower threshold of 90% correct criterion were also run, as some initial simulation runs failed to reach the 95% criterion by the end of training (Supporting Information, Appendix A, Figure S2).

We tested linear mixed effects (LME) models on length of training time (*lmer* and *lme4;* R [v3.6.3, 2020]), with number of objects during training (condition: 1, 2, or 6) as a categorical fixed effect (categorical so the difference between each of these contextual conditions on performance could be determined), gesture cue condition (0%, 33%, 67% and 100%) as a numeric fixed effect, and simulation run (1 to 10) as a random effect. We included number of objects during training and gesture condition as random slopes, but adding gesture cue condition, or the interaction between number of objects and gesture cue condition, resulted in the model not converging. The models were built including one fixed effect at a time, and using log-likelihood comparison to compare the contribution to model fit of each fixed effect (Barr et al., 2013).

Cues during training

Adding number of objects during training resulted in a significant improvement in fit, ($\chi^2(2) = 10.10$, $p = .006$). Quicker word learning was achieved with one object than two objects ($t(106.89) = 5.075$, $p < .001$), and two objects than six objects ($t(106.99) = 18.129$, $p < .001$). Gesture cue also significantly improved fit ($\chi^2(1) = 45.70$, $p < .001$), with greater cue availability resulting in quicker learning. The interaction also significantly improved fit ($\chi^2(2) = 14.23$, $p < .001$), with increasing availability of gesture cue having a stronger effect on learning speed in the two- and six-object conditions compared to the one-object condition ($t(114) = -3.572$, $p < .001$; $t(114) = -2.881$, $p = .005$, respectively). The effect of gesture cue on the two- and six-object conditions was not significantly different ($t(114) = 0.690$, $p = .491$). The resulting model is shown in Table 1 and the mean learning times for each object condition is shown in Figure 2A.

The model could learn word-referent mappings using cross-situational statistics and performed better with a cue: the addition of gesture (enhancing input activation from one location in the visual input layer) increased the associative learning signal from this region of the visual input. The model learned more quickly when there was no referential uncertainty about the target object – the one object condition learned faster than when two or six objects were present, but as we predicted, the gesture cue had a larger influence on learning under conditions of referential uncertainty. This of course makes perfect sense: when there is only one object, the model does not need support for disambiguating the referent. There was also a larger effect of gesture cue availability on the two-object than the six-object condition.

Accuracy at test

For testing performance, we constructed a series of generalised LME models in a similar way to the analyses of training length, with fixed effects of number of objects present during training and gesture cue condition, and random effects of simulation, but also an additional random effect of test item. Slopes for both fixed effects and their interaction were included for each random effect.

Number of objects present during training contributed significantly to fit ($\chi^2(2) = 18.43$, $p < .001$), with one object resulting in lower accuracy than two and six objects ($z = 18.77$, $z = 12.033$, both $p < .001$, respectively), and six objects resulting in lower accuracy than two objects ($z = -3.34$, $p < .001$). Adding gesture cue did not significantly improve fit ($\chi^2(1) = 0.936$, $p = .333$), but the interaction between gesture cue and number of objects during training was significant ($\chi^2(2) = 23.54$, $p < .001$). As with the training time analysis, the effect of gesture cue availability had a stronger facilitative effect on accuracy for the two- and six-object conditions compared to the one-object condition ($z = -8.64$, $z = -5.88$; both $p < .001$, respectively), and the effect of gesture cue availability on the two- and six-object conditions was not significantly different ($z = 1.30$, $p = .194$). The final model is shown in Table 1 and Figure 2B.

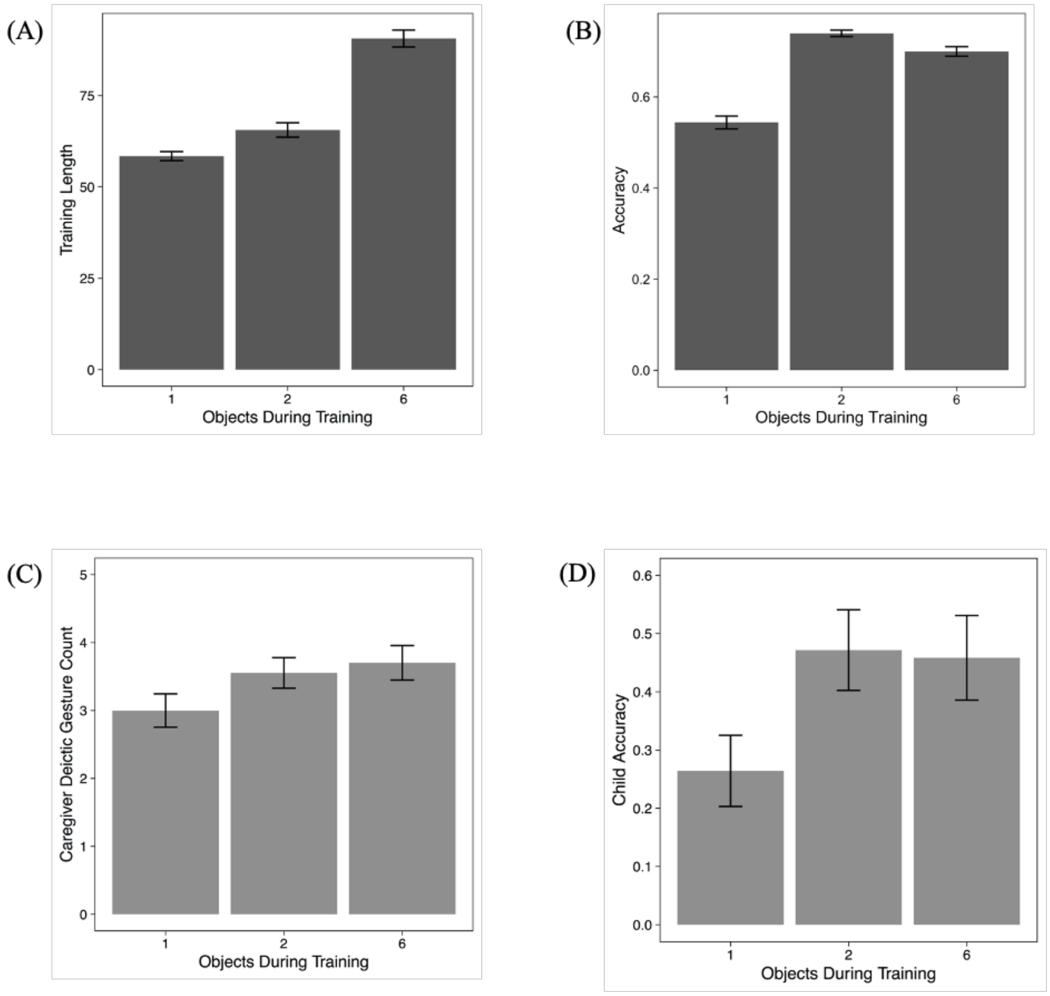
Table 1.

Computational model: linear mixed effects model results of the MIM computational model's performance, testing the effects of number of objects during training and gesture cue condition on length of training time and accuracy.

Dependent variable	*Independent variables*	*Estimate*	*SE*	*df*	*t*	*p-value*
Length of training time	(intercept – one object)	68.62	1.49	114	46.05	< .001
	One v. two objects	13.21	2.11	114	6.27	< .001
	One v. six objects	37.04	2.11	114	17.58	< .001
	Two vs. six objects	23.84	2.11	114	11.31	< .001
	Gesture cue	-20.39	2.39	114	-8.54	< .001
	One v. Two object x Gesture cue	-12.06	3.38	114	-3.57	< .001
	One v. Six object x Gesture cue	-9.73	3.38	114	-2.88	.005
	Two v. Six object x Gesture cue	2.33	3.38	114	0.69	.491
		Estimate	*SE*		*z*	*p-value*
Testing accuracy after training to criterion	(intercept – one object)	-0.53	0.16		-3.36	< .001
	One v. two objects	2.67	0.19		13.91	< .001
	One v. six objects	1.94	0.21		9.30	< .001
	Two vs. six objects	-0.66	0.20		-3.35	< .001
	Gesture cue	0.40	0.07		5.58	< .001
	One v. two objects x Gesture cue	-0.54	0.08		-6.80	< .001
	One v. six objects x Gesture cue	-0.45	0.09		-4.98	< .001
	Two v. six objects x Gesture cue	0.07	0.07		0.91	.365
		Estimate	*SE*		*z*	*p-value*
Testing accuracy after extended training	(intercept – one object)	-0.70	0.15		-4.51	< .001
	One v. two objects	2.95	0.18		16.20	< .001
	One v. six objects	2.12	0.21		9.93	< .001
	Two vs. six objects	-0.45	0.13		-3.50	< .001
	Gesture cue	0.48	0.07		6.68	< .001
	One v. two objects x Gesture cue	-0.66	0.08		-8.64	< .001
	One v. six objects x Gesture cue	-0.55	0.09		-5.88	< .001
	Two v. six objects x Gesture cue	0.26	0.20		1.30	.194

Figure 2.

Mean and standard error bars for results of the MIM and behavioural study. Note that for testing accuracy, there were three objects present and no gesture cue. (A) MIM: Training length time by number of objects present during training (calculated across gesture cue condition);† (B) MIM: Testing accuracy proportion correct by number of objects present during training (calculated across gesture cue condition);† (C) Behavioural study: Count of caregiver deictic gesture use by number of objects present during training; (D) Behavioural study: Child testing accuracy proportion correct by number of objects present during training.



†For MIM results by number of objects present during training and by individual gesture cue condition, please see Supporting Information, Figure S1.

As there was a confound between training length and availability of gestural cues, additional simulations were run where the model was trained to the same amount of exposure for each of the different levels of availability of gestural cues, with similar accuracy results (Supporting Information, Appendix A, Table S1, Figure S2).

Unexpectedly, the model demonstrated more robust retention of the word-object mappings during testing when it had been trained under referential uncertainty; the two- and six-object conditions achieved higher accuracy than the one-object condition. In Monaghan (2017), the MIM performed best when there was some variability in the cues (when present 33% or 67% of the time) rather than with no variability (present 0% of the time) or a large degree of variability (present 100% of the time). However, in the current simulations, the effect of altering the number of potential referents in the environment for a given word also affected learning – some, but not a great deal, of referential uncertainty resulted in better learning, with the model demonstrating the highest accuracy in the two-object condition.

Thus, the computational model confirms our expectations about gesture being more important in the presence of referential uncertainty. We predict that if caregivers are sensitive to the potential value of a cue, then they ought to use more gestures in word learning situations when two unfamiliar referents are present rather than one. We might also predict that gestural cue use increases when six potential referents are present, though the model learned under these conditions to a similar degree irrespective of gesture cue availability.

However, the model also generated additional predictions that were unexpected: that word learning could actually be more successful when learning takes place under conditions of referential uncertainty. These results imply that variability in the environment can support learning. These hypotheses generated by the MIM were then tested in a behavioural word learning study with children aged 18–24-months-old and their caregivers.

2.5 Behavioural study

This experiment examined gesture use when caregivers taught their children novel word-object mappings under different degrees of referential uncertainty, and also explored whether gesture use under referential uncertainty predicts word learning. During training, caregivers taught their child three novel word-object pairs across the same conditions of referential uncertainty as simulated in the computational model – one, two, or six novel objects with a single target object per condition. Children were then tested on the novel word-object pairs taught by their caregiver during training.

2.5.1 Method

Participants

Forty-seven caregiver and child dyads, recruited through Lancaster Babylab, completed training (*M* = 20.5 months, *SD* = 1.7, male = 27; Table 2). All caregivers gave informed consent for the dyad. All dyads were from monolingual English homes, with no history of developmental or sensory disorders. The data from an additional six dyads were excluded due to child fussiness (Supporting Information, Appendix A, Table S9). Twenty-seven of the dyads that completed training also completed testing (*M* = 20.8 months, *SD* = 1.6, male = 13), with the remaining dyads excluded due to incomplete trials (16) or child fussiness (4). Dyads received a storybook for participation and reimbursement for travel expenses.

Stimuli

Three novel words were used: *darg*, *noop*, and *terb* (NOUN database; Horst & Hout, 2016). Nine similarly sized novel objects with different colours and shapes were used as stimuli (e.g. Figure 3). Three of these objects were randomly paired with the three novel words per participant. The remaining six objects then served as foils.

Training

Caregivers were familiarised with the three novel word-object pairs prior to the experiment without the child present. During training, the novel word and a three-word description of the target object were visible to the caregiver as a memory aid. Caregivers were told to imagine they were in an everyday setting, such as a shop with items on a shelf out of reach, and instructed to teach the novel words to their children as if they were real words for objects that the child had not seen before. Children then sat on their caregiver's lap and were presented with stimuli on a tray 70 cm away for 30 seconds, during which caregivers taught their child the novel word-object mapping (three training trials; 30 seconds each; one per novel word-object mapping). During training, dyads could not reach or handle the objects.

Dyads began with a warm-up trial where a red ball was presented on the tray and caregivers practised teaching their child the word 'ball'. All dyads were then administered all three conditions where target objects would appear alone (one-object condition), with another foil (two-object condition), or with five foils (six-object condition), reflecting the computational model's learning conditions (Figure 3A). A Latin Square was used to counterbalance the order in which training conditions were administered, and the position of targets per condition was also randomised in the same way as the computational model's training.

Testing

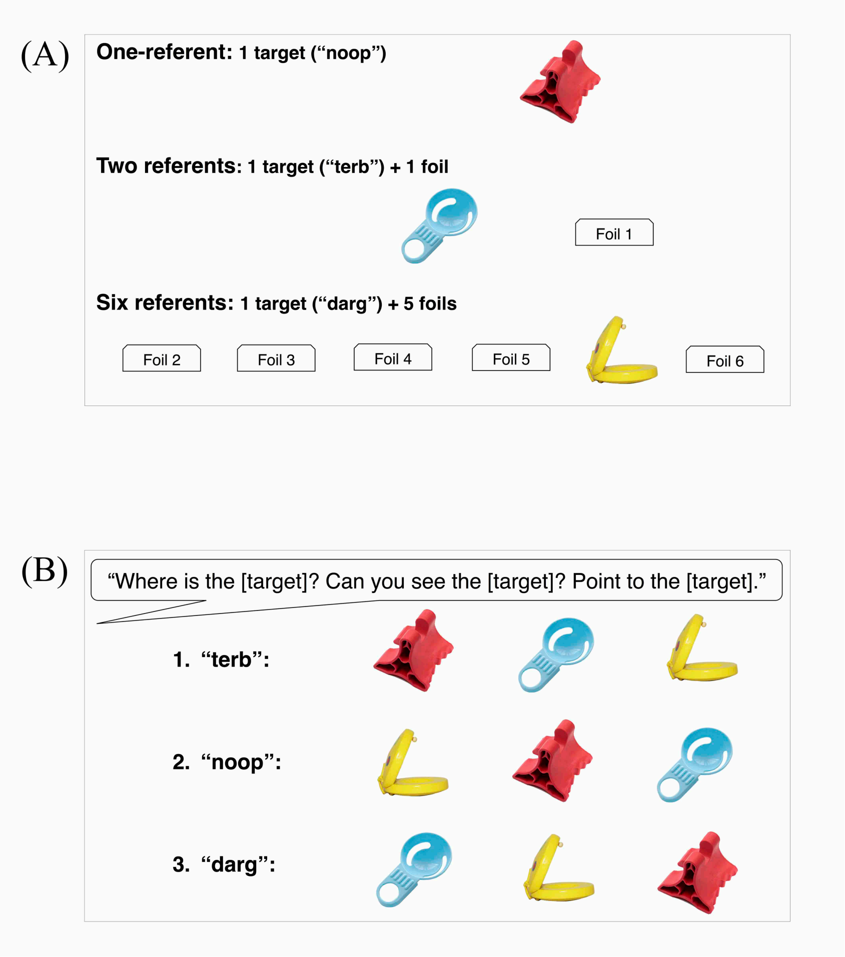
After training, children were tested by the experimenter on the three novel word-object mappings they had just learnt in a three alternative forced choice test, mirroring the computational model, with each word tested on separate trials (each word tested twice, six test trials in total; Figure 3B).

For each trial, the tray was arranged out of sight and then made visible. The then experimenter asked the child "Where is the [novel word]? Can you see the [novel word]?

Point to the [novel word].” The tray was moved forward within the child’s reach, and the child pointing towards, reaching for, or touching an object was recorded as a response. If the child did not respond, this was repeated; if the child still did not respond, the experimenter advanced to the next test trial. A Latin Square was used to counterbalance the order of conditions during testing across participants.

Figure 3.

Behavioural study: (A) Example of training trials; (B) Example of testing trials.


(A)
One-referent: 1 target (“noop”)
Two referents: 1 target (“terb”) + 1 foil
Foil 1
Six referents: 1 target (“darg”) + 5 foils
Foil 2
Foil 3
Foil 4
Foil 5
Foil 6
(B)
“Where is the [target]? Can you see the [target]? Point to the [target].”
1. “terb”:
2. “noop”:
3. “darg”:

Coding

Training trials were video-recorded and coded per utterance for total gestures and speech co-occurring with gesture by a trained coder (see Supporting Information, Appendix A, for details). An independent second rater coded 20% of the videos (randomly selected), with an inter-rater reliability of Cohen's $\kappa = 0.78$ for categorisation of gesture into subtypes (*deictic, representational, other*; $N = 284$; 85.21% agreement) and Cohen's $\kappa = 0.86$ for categorisation of speech with gesture into subtypes (*complementary* or *supplementary*; $N = 160$; 92.5% agreement).

An utterance was defined as a string of words or gestures preceded and followed by a pause or changes in conversation turn or intonation (Rowe et al., 2008). For gesture subtypes, we adapted Rowe et al.'s (2008) coding system: *deictic* gestures were intentional, clear movements that singled out the target, including pointing towards the target (e.g. finger points with the arm in extension) and reaches towards the target (e.g. extension of the arm with the palmar aspect of the hand exposed, or extension of the arm with the fingers in extension). *Representational* gestures included upper limb or body movements depicting object attributes such as shape or size (e.g. indicating a ball is round with two hands cupped and fingers flexed) and actions with the object (e.g. cupping the palmar aspect of one hand with fingers flexed, followed by arm movement forward from the shoulder joint, to indicate a ball rolling). *Other* gestures included all gestures not directed towards the referent; these included both deictic and representational gestures towards foils, to the experimenter, or caregiving-related gestures such as a parent hugging a child.

We adapted Iverson and Goldin-Meadow's (2005) coding system for speech with gesture in order to account for the effect of combined gesture and speech on learning as either *complementary*, where speech contained the target label, or as *supplementary*, where speech contained related information about the target referent such as size, colour, or function. Deictic gestures and occurrences of complementary speech with gesture correspond to the gestural cue conditions of the computational model. We also recorded the total number of times the *referent label* was used.

Vocabulary measures

Caregivers completed a demographics questionnaire that included socioeconomic status (SES; determined by parent education level). A parent-report measure of child vocabulary, the UK Communicative Development Inventories (CDI; Alcock et al., 2017) was also administered. The UK CDI measures expressive, receptive, and gesture vocabulary (communicative and symbolic). Communicative gestures include declarative and imperative gestures. Symbolic gestures are representational gestures that include actions, games, and pretend play.

2.5.2 Results and Discussion

Data, code, and models run are available on OSF (http://osf.io/6frcw/?view_only=72344789a6294aa19d63a8bd93a628f3). All dyads were from similar, mid-high SES backgrounds. Dyads that only completed training and those that completed both training and testing, did not yield any significant differences in demographics or CDI scores (Table 2).

Table 2.

Behavioural study: demographics and child vocabulary scores. Measured by the UK-Communicative Development Inventories with Welch two sample t-tests comparing those that completed training only, and those that completed training and testing.

	Completed training (total sample; N = 47)	*Completed training + testing trials (n = 27)*	*Completed training only (n = 20)*	*Welch two sample t-tests (completed training + testing, v. completed training only)*		
Sex (m:f ratio)	27:20	14:13	13:7			
	mean (sd)	*mean (sd)*	*mean (sd)*	*t (df)*	*95% CI*	*p-value*
Age (months)	20.5 (1.7)	20.8 (1.6)	20 (1.8)	-2 (38)	[-1.85, 0.22]	.100
Receptive vocab.	276 (91.5)	294 (87.9)	251 (92.9)	-2 (40)	[-96.5, 11.7]	.100
Expressive vocab.	146 (114)	159 (119)	129 (108)	-0.9 (43)	[-97.2, 36.8]	.400
Comm. gesture	19.9 (3.79)	20.5 (3.9)	19.1 (3.6)	-1 (43)	[-3.60, 0.83]	.200
Symb. gesture	41.1 (6.9)	41.4 (7.4)	40.5 (6.4)	-0.4 (33)	[-5.40, 3.58]	.700

comp. = complementary; symb. = symbolic; vocab = vocabulary

To compare behavioural results to the computational model prediction that cue importance increased with referential ambiguity, we tested whether the number of objects during training affected caregiver behavioural cue use; in particular, deictic gesture use. LME models (*lmer* and *lme4*; R [v3.4.1, 2017]) were constructed to predict caregiver deictic gesture use, complementary speech with gesture, and referent label use separately. For each analysis, the number of objects during training (condition: 1, 2, or 6) was included as a categorical fixed effect, and child vocabulary was included as a numeric fixed effect. Due to high correlation between expressive and receptive vocabulary, separate linear mixed effects models were carried out – one with fixed effects of expressive, symbolic, and communicative gesture vocabulary, and one with receptive, symbolic, and communicative gesture vocabulary. Only the latter analysis is included here as the task required children to understand, rather than produce, novel words. Analyses with expressive vocabulary resulted in similar effects and are reported in the Supporting Information (Appendix A, Tables S3-S4). The models also contained random effects of participant, child age, target word, and target item. Slopes of condition per participant resulted in the model not converging. As for the computational model analysis, we included one fixed effect at a time, and used log-likelihood comparison to compare the contribution to model fit for each fixed effect (Barr et al., 2013). Separate LME models were also constructed in the same way to predict caregiver and child behaviour for each subtype described in our coding scheme to examine the range of caregiver communication with their children. We report here complementary speech with gesture and referent label use as these also highlight the referent in a similar manner to deictic gestures; all other subtypes can be found in Supporting Information (Appendix A, Tables S3-S4, Figure S4).

Cues during training

Caregiver data demonstrated a significant effect of condition on overall gesture use ($\chi^2(2) = 11.73$, $p = .003$). Consistent with the MIM results, this was largely due to deictic gesture cues ($\chi^2(2) = 9.48$, $p = .009$; Table 3, Figure 2C), with caregivers using more deictic gesture cues in the two-object ($t(90.24)= 2.32$, $p = .023$) and six-object ($t(91.79) = 3.08$, $p = .003$) conditions when compared to the one-object condition. Caregivers demonstrated no significant increase in deictic gesture use between two- and six-object conditions ($t(93.35) = 0.77$, $p = .445$). There were no significant fixed effects of child vocabulary or significant interactions found, and representative and other gestures did not yield any significant effects or interactions (Supporting Information, Appendix A, Figure S4A).

Table 3.

Behavioural study: linear mixed effect model (LME) results testing the effects of number of objects during training and child vocabulary scores on caregiver deictic gesture use during training trials, and generalised estimated equation (GEE) results on the effects of number of objects during training and child vocabulary scores on child accuracy at test.

Dependent variables	*Independent variables*	*Estimate*	*SE*	*df*	*t*	*p-value*
Caregiver deictic gestures during training (LME)	(intercept – one object)	2.99	0.31	12.58	9.76	<.001
	One v. two objects	0.57	0.25	90.24	2.32	.023
	One v. six objects	0.76	0.25	91.79	3.08	.003
	Two v. six objects	0.19	0.25	93.35	0.77	.445
		Estimate	*SE*		*Wald*	*p-value*
Child testing accuracy (GEE)	(intercept – one object)	-1.76	0.66		7.05	.008
	One v. two objects	0.90	0.43		4.36	.037
	One v. six objects	0.85	0.46		3.32	.068
	Two v. six objects	-0.05	0.50		0.01	.921
	Receptive vocabulary	0.002	0.002		1.24	.265
	Caregiver deictic gesture	0.03	0.10		0.10	.749

When examining caregiver complementary speech with gesture the addition of child symbolic gesture vocabulary improved model fit with a main effect of condition ($\chi^2(3) = 0.43$, $p < .001$; Table 4). Caregivers used more complementary speech with gesture in the two-object than the one-object condition (*t*(80) = 2.58, *p* = .012), but there was no significant difference between the two-object and six-object conditions (*t*(80) = -0.89, *p* = .375). A significant effect of condition on their overall use of the novel label was also found ($\chi^2(2)$ = 11.90, *p* = .003, Table 4). The novel label was uttered significantly more by caregivers in the two-object compared to the one-object condition (*t*(89.49) = 2.37, *p* = .020), but significantly less in the six-object compared to the two-object condition (*t*(89.66) = -3.52, *p* < .001). No other significant effects of child vocabulary or interactions were found.

Overall, these results were consistent with the MIM model showing the largest effect of gesture availability in the two- and six-object conditions.

Table 4.

Behavioural study: linear mixed effects model results testing the effects of number of objects during training and child vocabulary scores on caregiver gesture and speech with gesture subtypes during training trials.

Dependent variable	*Independent variables*	*Estimate*	*SE*	*df*	*t*	*p*-value
Referent label use	(intercept – one object)	6.49	0.57	8.11	11.44	<.001
	One v. two objects	0.77	0.33	89.49	2.37	.020
	One v. six objects	-0.39	0.33	91.12	-1.18	.242
	Two v. six objects	-1.16	0.33	89.66	-3.52	<.001
Comp. speech with gesture	(intercept – one object)	0.43	1.04	41.57	0.41	.069
	One v. two objects	0.65	0.25	80.00	2.58	.012
	One v. six objects	0.43	0.25	80.00	1.68	.096
	Two v. six objects	-0.23	0.25	80.00	-0.89	.375
	Symb. gesture vocab	0.03	0.02	36.27	1.33	.193

comp. = complementary; symb. = symbolic; vocab = vocabulary

Accuracy at test

We used Generalised Estimated Equations (GEE; *geeglm* and *geepack*; R[v3.4.1, 2017]) to examine the effect of condition, caregiver behaviour, and child behaviour during training on test trial accuracy.[1] Separate GEEs were constructed to examine child vocabulary variables, condition, and each training behaviour gesture subtype as independent variables; here we report the effect of caregiver deictic gesture use with child receptive vocabulary. For all other subtypes and child vocabulary variables, please see Supporting Information (Appendix A, Tables S5-S8).

In line with the computational model results, children performed most accurately in the two-object condition (Table 3, Figure 2D), although there was no significant difference in accuracy between the two-object and six-object condition (*Wald* = 0.01, *p* = .921). However, children responded significantly more accurately in the two-object than the one-object condition, even when child receptive vocabulary and caregiver deictic gesture use were accounted for (*Wald* = 4.36, *p* = .037).

Although the lack of referential ambiguity would suggest that word-object mapping should be easier in the one-object condition, a higher success of word learning in the two- and six-object conditions was consistent with the MIM computational results. Additionally, although children were offered the least amount of gesture information by caregivers in the one-object condition, adding caregiver behaviour subtypes during training to the analysis did not contribute any significant value to predicting accuracy during testing (Table 3; Supporting Information, Appendix A, Tables S5-S8).

2.6 General Discussion

Natural language learning environments are noisy and variable, and yet children still manage to accurately map words to objects. In this study, we predicted that a computational model of word learning (MIM) trained under conditions of varying referential uncertainty would learn faster with fewer potential referents. We also predicted that a gestural cue would be most helpful to word-referent mapping when there was an increase in potential referents.

[1] General linear mixed effects models (*glmer* package; *lme4* in R [v3.4.1, 2017]) were originally used but failed to converge, so GEEs were employed.

Contrary to our first prediction, but consistent with literature highlighting the value of variability during word learning (e.g. Apfelbaum & McMurray, 2011; Monaghan, 2017), the computational model predicted the most robust learning when there were several potential referents, rather than just one. Although the MIM learnt quickest in the one-object condition, there was higher accuracy at test when it had been trained under referential uncertainty during the two- and six- object conditions. The addition of a gestural cue during training significantly improved learning when there were more potential referents as predicted, but the model also benefited from the presence of variability via the availability of gestural cues, learning most robustly when cues were presented 33% and 67% of the time.

This generated two hypotheses for testing in behavioural settings. Firstly, if caregivers are sensitive to the role of gestural cues in supporting word learning, they ought to use more gestures when there is referential uncertainty, and secondly, children might actually learn best when trained under referentially uncertain conditions. The experimental study did identify that caregivers adapt their gestural cues to support learning in the face of referential uncertainty, but with significant increases only from the one-object to the two-object or six-object condition, and no significant increase from the two-object to the six-object condition. Finally, the experimental study also found that children learnt best under referential uncertainty, performing most accurately in the two- and six-object conditions, in line with the model's surprising predictions.

These results were somewhat counterintuitive; one might expect the highest test accuracy in the behavioural study for words learnt in the one-object condition. This would be consistent with the fast-mapping literature, where children are able to identify a new word after a single exposure (Carey & Bartlett, 1978), and with cross-situational word learning in adults that indicates increasing the number of potential referents results in less accurate and slower learning (K. Smith et al., 2011; Trueswell et al., 2013; Yu & L.B. Smith, 2007). Despite this, our task differed in several ways that could have affected performance at test.

Firstly, children were not tested on each word after the corresponding training trial as in referent selection trials during fast-mapping tasks (Horst & Samuelson, 2008) – they were tested after all training trials. Secondly, the co-occurring foils were novel, whereas fast-mapping tasks involve familiar objects alongside novel objects. Cross-situational word learning paradigms also usually offer the opportunity to learn from within- and across-trial competition as all objects are named (Yurovsky et al., 2013). In our study, there was no such opportunity, as different foils were used within-subject for each condition, and testing trials consisted of forced-choice between the three target objects.

Rather, it is possible that the presence of referential uncertainty in the two- and six-object conditions might have supported learning through enabling comparison. The role of two or more competing alternatives is well established in internal constraint accounts of language learning, including mutual exclusivity (Halberda, 2006; Markman & Wachtel, 1988) and the novel name-nameless category principle (Golinkoff et al., 1992). Similarly, children's learning of categories is aided by having an alternative, either by using comparison, where one object appears with others in the same category, or by contrast, where an object appears with a non-category object (Ankowski et al., 2013). Such a beneficial effect may also apply in our study where the referent is identified among a range of other unknown objects.

Few studies have examined cross-situational referential ambiguity in infants and children, with most limiting referential ambiguity to two potential referents per training trial (e.g. L.B. Smith & Yu, 2008; Yu & Smith, 2011). Those that have examined older children (5–7-years-old) suggest that they may struggle most when a specific foil, termed a high probability competitor, co-occurs with a target more often than other foils (Suanda et al., 2014). Bunce and Scott (2017) examined 2.5-year-old children with four potential referents per trial. Children could identify the correct target using cross-situational statistics with four potential referents without exhaustive labelling when all distractors were different (no across-trial competition), and even with a high probability competitor – but only if a different foil appeared by the last trial, allowing disambiguation at the end of training. This suggests that children are able to learn under certain circumstances with increased referential ambiguity, subject to limitations in cognitive and memory capacity.

Another potential explanation for performance in the one-object condition is that children were less interested compared to when there were several objects present. Future research could use an eye-tracker to measure attention more precisely and determine how foils are fixated on alongside targets. Testing immediately after training trials using both target and foil objects may also help illuminate whether children process all objects present.

The present computational model and experimental study also highlighted that some variability in both the environment and in the use of cues in communication may facilitate learning. We have demonstrated that the former influences the latter, establishing that caregiver gesture cue use when teaching their children novel words was contingent on the presence of referential uncertainty. This is consistent with the theory that gestures singling out target referents are particularly valuable during word-object mapping (Cartmill et al., 2013; Rader & Zukow-Goldring, 2012). However, although we expected gesture use during training to increase from the two- to the six-object condition, this was not the case. Hence, caregivers gestured and offered cues according to the presence, rather than the degree, of referential uncertainty, and did not offer significantly more cues when referential uncertainty was high.

Taken together, these results indicate that referential uncertainty is perhaps subject to some degree of cognitive management by both the caregiver and child, where high uncertainty can be reduced to a more tractable sense of ‘this, not that’. The use of gestural cues may reduce cognitive load for the infant (Goldin-Meadow, 2000; McGregor et al., 2009; McNeil et al., 2000); the key difference in our study seemed to be between having either one choice of word-object mapping, or more than one – beyond this, the benefits of gestural cues may begin to decline. Caregivers appeared to be sensitive to this lack of discrimination between the two- and six-object conditions, as there was no significant difference in their behaviour. A switch to laborious counting during the six-object condition, rather than being able to immediately perceive the number of items in the one- and two-object conditions (Cowan, 2001; Xu & Chun, 2009), may have affected how the caregiver then packaged information for their child. This could potentially lead to the treatment of the two- and six-objects as analogous by the caregiver, and thus the child. Similarly, gestural cues did not have a large effect on speed of learning in the six-object condition in the computational model compared to the one- and two-object conditions.

Studies of how children acquire representations of number additionally indicate that children around 20-months-old are not able to comprehend more than three or four objects (Feigenson et al., 2004; Le Corre & Carey, 2007; Wynn, 1990), which could also render performance across the two- and six-object conditions in our study somewhat analogous. Despite this, being able to distinguish only a limited number of stimuli may also help constrain word-referent mappings. Head-mounted cameras during toy exploration laboratory studies show that, despite having multiple objects in front of them, 20-month-olds tend to hold single objects in view at a time (L.B. Smith et al., 2011) and learn the names of objects that dominate their view simultaneously with label utterance (Pereira et al., 2014).

However, we did not test incremental increases in referential uncertainty, opting instead for no ambiguity, some ambiguity, and high ambiguity. An interesting avenue for future research would be to investigate whether there is a precise ‘tipping point’ in the number of potential referents at which caregivers cease to offer more gestural cues to their children and whether this then affects children’s learning – although the similar performance between the two-object and six-object conditions may suggest that anomalies in behaviour and learning are unlikely to occur with intermediate ambiguity between two and six objects.

Although cues are useful for supporting learning, they are also individually highly variable within naturalistic environments. Caregivers may not gesture towards intended referents on 85% of occasions (Iverson et al., 1999), articles may precede adjectives rather than nouns (Monaghan et al., 2007), and prosodic cues also are not always consistent (Fernald, 1991). The computational MIM simulations also found that the most robust learning occurred when gestural cues were present some of the time, rather than when they were exclusively present or absent. Why is this? Firstly, it has been established that a system that relies on perfectly reliable cues learns quickly, but learning is brittle when those cues are no longer reliable (Monaghan, 2017). Secondly, when identifying a target from amongst different competitors, the occasional lack of a cue may make the presence of one more salient, avoiding potential habituation effects (Veale et al., 2011) and preventing inhibition of other useful information (Kamin blocking effect; Shanks, 1985). A system where gestural cues vary may then have a higher degree of sensitivity to those cues than one where gestural cues are either always there, or always absent. Thus, variability of cues is not only more similar to real-world settings, but also benefits learning. This raises the intriguing possibility that the variability of cues when children are acquiring vocabulary may not be an accident of a noisy environment, but rather the stochasticity of adults' use of cues may be by design.

In our experimental study, we did not find any effect of training response variables on testing data – inclusion of caregiver gesture and speech use did not predict child accuracy after controlling for condition. If referential uncertainty and the cues in response to it are so vital to learning, why did this not manifest in our data? This may be partly due to our sample of mid-to-high SES families who had actively expressed interest in developmental research. Families from higher SES backgrounds have been found to use gesture more than those from lower SES backgrounds, with an increase in parental gesture correlating with increased child gesture and later vocabulary skill (Rowe & Goldin-Meadow, 2009). Our participants may well have been at a ceiling level of caregiver input, resulting in gesture adding very little. Gesture may be particularly beneficial to language development in environments with limited resources and a diminished quality of parental input (Kirk et al., 2013), and may be useful as part of language interventions in low income families (Vallotton, 2012). Consequently, we recommend that caution should be exercised when generalising our conclusions across different SES backgrounds.

Additionally, as our sample inclusion criteria precluded developmental delay, our findings may not extend to these populations (Hartley et al., 2019, 2020). Our results confirmed that caregivers appeared to be sensitive to task demands, and models predicting speech with gesture during training were improved with the addition of CDI subscales. Although these estimates were very small, the impact of child vocabulary could be more prominent in a language delayed sample (Wray & Norbury, 2018).

An alternative explanation concerning why caregiver behaviour did not predict children's behaviour relates to our sample's age (20-months-old on average). Previous literature links caregiver gesture and early child gesture use at 10–14-months-old (Liszkowski et al., 2012; Liszkowski & Tomasello, 2011) and caregiver gesture use in Rowe et al. (2008) predicted early child gesture use at 14-months-old, but not expressive vocabulary at 42-months-old. Caregiver gesture use appears relatively stable over time, whereas child gesture use may take a supportive role to speech once verbal ability is established (Goldin-Meadow, 2007; Iverson et al., 1999; Rowe et al., 2008). Subsequently, children in our study may have been at a stage where verbal input is weighted more heavily than gesture input. Although we examined some of these factors, our primary focus was deictic gestures. Future research could consider speech input in greater depth, including Mean Length of Utterance and temporal relations of naming events with gesture.

We also found a higher level of child dropout when testing trials commenced, reducing power for GEE analysis (which could also reduce the effect of caregiver gesture on child behaviour; Liszkowski et al., 2012; Liszkowski & Tomasello, 2011). Although we observed no significant differences between children that completed testing and those that did not, child fussiness may have been caused by objects being out of reach during training, resulting in frustration by the time testing commenced. This may mean that differences in temperament and attention could be present that were not accounted for. Additionally, whereas previous studies enabled children to freely explore an environment, we constrained the objects in our study to be out of reach to control for exposure times and interaction with the objects. This could have resulted in less gesture, particularly by children, who had no immediate receipt of the objects to which they gestured. Studies that compare objects within reach across varying environmental referential uncertainty, and that measure broader child traits, will usefully address these points. To isolate any effect of referential uncertainty itself from caregiver behaviour, future studies could also test children's word learning across referential uncertainty without caregiver interaction.

In conclusion, we found variability in gesture cue availability combined with referential ambiguity produced optimal learning in a computational model of word learning. This was supported by an experimental study that demonstrated that: (a) caregivers gestured according to the presence, rather than degree, of referential uncertainty, and (b) children learnt best in the presence, rather than absence, of referential uncertainty. These results advance understanding of communicative exchange during word learning, indicating that caregivers contingently adapt their gesture use according to the presence of referential uncertainty.

3 Chapter 3: Better early than late, and better late than never: The temporal dynamics of gesture cues in cross-situational word learning

3.1 Chapter introduction

Gesture cues support word learning and often occur in conjunction with spoken words, such as pointing at a referent for a novel label. Work by Trueswell et al. (Cartmill et al., 2013; Trueswell et al., 2016) and L. B. Smith et al. (Pereira et al., 2014; L. B. Smith et al., 2010) indicates that the timing of *when* a referent is visually highlighted in conjunction with hearing an auditory label is important for encoding word-referent pairs. This is consistent with how attention may be allocated as a result of endogenous cues, such as arrows (Brignani et al., 2009; Yoshida & Burlington, 2012), but also with how sustained attention to objects predicts vocabulary size (Yu et al., 2019).

However, we do not know (a) how gesture cues influence referential ambiguity, (b) how manipulating the timing of a gesture cue might affect word learning accuracy or (c) how the process of highlighting visual referents with gesture cues relative to auditory labels temporally unfolds with respect to attention. Addressing these knowledge gaps enables us to gain not only a better understanding of gesture cues in cross-situational word learning, but may also offer insight into why gestures occur before speech under naturalistic settings.

In this paper, we examine the influence of a pointing gesture cue in adult word learners by firstly altering the amount of referential ambiguity, and then by altering the timing of the cue, tracking the progress of learners using an eyetracker.

Author contribution for Chapter 3: *Rachael W Cheung:* design, data collection, analysis, writing, review. *Calum Hartley:* design, review. *Padraic Monaghan:* design, review

3.2 Abstract

Gesture cues provide substantial support for language, and gesture frequently accompanies child directed speech. How gesture cues integrate temporally with speech information during word learning is not yet clear. Across three pre-registered experiments, we investigated how the timing of gesture cues interplay with referential ambiguity during cross-situational word learning in adults. Experiment 1 showed that referential ambiguity can be reduced with a gesture cue to the same level as unambiguous conditions. Experiment 2 tested presentation of a gesture cue 1 second before or after a label with referential ambiguity, and showed that gesture preceding the label promoted the most accurate learning – although the presence of a late gesture aided learning more than having no gesture cue. Finally, Experiment 3 investigated the time course of learning with gesture cues before and after the label using eye tracking. We showed the learning advantage afforded by early gestures was due to how participants' attention was directed during label utterance, and that this advantage was apparent even at initial exposures of word-referent pairs. Our findings show gesture cues support word learning by reducing referential ambiguity as the learning process unfolds, allowing time-coupled integration of visual and auditory information that aids encoding of word-referent pairs.

3.3 Introduction

The environment surrounding language learners is busy, with multiple variable sources of information present (Holler & Levinson, 2019). In this environment, a learner must accurately assign unknown, novel words to the correct objects, concepts, or actions (*referent selection*) in order to acquire language, and further encode these pairings for later retrieval (*retention*). Referent selection poses the referential ambiguity problem (Markman, 1989; Quine, 1960; L. B. Smith & Yu, 2008): given the many sources of input, and the multiple possible pairings between words and referents, how does a language learner successfully arrive at the correct word-referent pair?

A number of learner-based constraints have been proposed to address the referential ambiguity problem. These include the novel name-nameless category (e.g. when presented with a familiar object with a known name and an unfamiliar object with an unknown name, the latter must be the referent of a novel word; Golinkoff et al., 1992), and explicitly contrasting an unfamiliar named referent with a familiar named foil (e.g. asking a child to retrieve a 'chromium tray, not the blue one'; Carey & Bartlett, 1978). Another example is mutual exclusivity (Markman & Wachtel, 1988), where learners reject one object in favour of another by process-of-elimination (disjunctive syllogism*;* Halberda, 2006): when presented with two objects, and one object is already known to be a 'cup', then the 'terb' must be the other object. The use of a contrast (Clark, 1987) in these instances aids referent selection by providing a means of prior reference – the familiar object – which reduces referential ambiguity. However, these strategies cannot be applied by learners in situations where all objects are novel and there are no familiar objects to disambiguate from.

Other prominent accounts have considered how the environment itself can contribute to more general learning processes, such as via the availability and use of cross-situational statistics (Siskind, 1996). Cross-situational learning refers to the aggregation of information and commonalities across several, rather than single, learning instances (Yu & Smith, 2007).

Typically, a cross-situational word learning paradigm presents a series of individually ambiguous trials involving two or more novel objects that co-occur with novel words, and the learner's task is to individuate as many correct word-object pairings as possible. These objects and labels are presented over many trials with different foils, but no additional information is provided to inform the learner which words refer to which objects. Thus, the learner must acquire novel label-object pairs simply by tracking the co-occurrence of particular words and objects across multiple exposures (e.g. Fitneva & Christiansen, 2011; Monaghan & Mattock, 2012; Roembke & McMurray, 2016; K. Smith et al., 2011; Yu & Smith, 2007; Yurovsky et al., 2013). Whilst internal bias accounts focus on disambiguating meaning within the context of an individual naming event, cross-situational word learning relies on the aggregation of knowledge across multiple naming events.

However, cross-situational statistics are only one source of information for how a learner might solve the mapping problem when faced with multiple unknown referents. Referential ambiguity may be reduced through other environmental cues. As a result, multiple cue models have explored how the use of additional cues such as gaze direction, prosody, or gesture (e.g. Hollich et al., 2000) can be combined with cross-situational word learning to facilitate mapping of word-referent pairs (Monaghan et al., 2017; Yu & Ballard, 2007).

Pointing gestures in particular have a high degree of accuracy when identifying a word's intended referent during parent-infant interactions (Cartmill et al., 2013; Frank et al., 2013) and also during adult cross-situational word learning (Monaghan, 2017; Monaghan et al., 2017). One possible explanation for their beneficial effect is that pointing gestures modulate the degree of referential ambiguity by leading participants to utilize single hypotheses about word-object pairs, reducing the formation of spurious word-object associations (similar to the effect of eye gaze; MacDonald et al., 2017). Therefore, pointing gestures may modulate referential ambiguity by reducing any potential conflicting associations between foils and words. However, whether or not external cues – such as pointing – can reduce uncertainty to the same level as learning instances with no referential ambiguity is currently unknown.

Understanding how external cues support cross-situational word learning also requires the consideration of temporal processing. The use of cues in learning and attentional orientation has been well-documented in the attention and memory literature (e.g. Hauer & Macleod, 2016), but remains under investigated in language research. Such studies distinguish the use of *endogenous* cues (e.g. arrows or eye-gaze), where attention is directed voluntarily to a target, as opposed to *exogenous* cues (e.g. flashing lights), where attention is directed automatically as a result of sudden salient stimuli (Jonides, 1980; Posner, 1981).

Naturalistic social cues during word learning, such as eye-gaze and pointing gestures, likely act as endogenous cues similar to those that are examined during attentional shifting experiments (Brignani et al., 2009). Pointing gestures may reduce referential ambiguity by orienting attention, thus strengthening the encoding of a highlighted word-referent pair. Similarly, changes in how endogenous and exogenous cues are weighted over the course of early development may inform how competition between different cues is resolved in language acquisition, with infants utilizing exogenous cues first, followed by endogenous cues as they mature (de Diego-Balaguer et al., 2016; Wu and Kirkham, 2010). However, despite calls for further examination of attentional cuing in word learning (L. B. Smith et al., 2010) and studies that examine infant word learning through attentional mechanisms in multi-modal environments (e.g. N. Kirkham et al., 2019; Yoshida & Hanania, 2007), the means by which attentional cues interact with cross-situational statistical information is not yet fully understood.

Studies that examine use of endogenous cues suggest there is temporal sensitivity to the role of these cues. Whereas exogenous cues quickly shift focused attention between a cue and a target at 50ms, shifts of focused attention due to endogenous cues may take up to 500ms (Berger et al., 2005; Shepherd & Müller, 1989). This indicates that the timing of a cue in relation to a label utterance appears crucial to word-referent mapping and to how attention is directed: cues may be more useful if they occur *before* the label in order to allow attention to be shifted early, thus allowing attention to be focused on the referent *during* label utterance.

This effect is reflected in naturalistic studies of gesture and labelling. In the Human Simulation Paradigm (HSP; Gillette et al., 1999), adult participants guess 'missing' words from parent-child interaction videos, where the target word is obscured by an auditory 'beep' (e.g. 'where's the [obscured target word]?'). Scoring participants' accuracy of guess provides a measure of how informative any surrounding cues are when identifying the target word. Trueswell et al. (2016) found that the gestures made by parents within parent-child interaction videos could be used to predict the accuracy of participants' guesses relative to target word onset. Gestures were time-locked to word utterance in their ability to reduce referential ambiguity. Vitally, shifting the obscuring 'beep' 2 – 4 seconds away from actual word occurrence significantly reduced guessers' accuracy in identifying the target referent. This indicates that the relative timing of gesture and speech events is crucial to identifying correct word-referent pairs, where even minimal disruption to the time course can yield reductions in accuracy of word-referent pairs. Evidence from head-mounted cameras also suggests that the timing of a referent's appearance is tightly linked to word utterance. When parents teach infants novel words, successfully learned referents tend to appear centrally in view from -6 seconds to +5 seconds from naming events (Pereira et al., 2014).

However, there has as yet been no direct demonstration of the learning effects of gestural cues at different onsets with respect to word-referent pairings, meaning the early cueing effect of gestures relative to labels has not been tested experimentally. Directly manipulating different onsets for gestural cues influencing word-referent mappings enables us to unpack why temporal processing matters in word learning, as well as allowing us to establish how the timing of gesture interacts with competing potential word-referent pairs within cross-situational word learning.

Finally, although various sources of information may aid accurate referent selection, this may not necessarily reflect long-term learning. Initially accurate referent selection under referential ambiguity may reflect ‘fast’, in-moment problem-solving by the learner, whereas retention of novel words may occur as a ‘slow’ and gradual process, during which multiple exposures are used to strengthen or weaken word-referent pairs over time *(*as in the *dynamic associative model;* McMurray et al., 2012). Whereas mapping novel word-referent pairs via cross-situational statistics is possible from 12 months (L. B. Smith & Yu, 2008), retaining and retrieving them may occur on a developmentally slower scale (Hartley et al., 2020; Horst & Samuelson, 2008; Vlach & DeBrock, 2019). However, we do not yet know how in the presence of endogenous cues, such as gesture, during cross-situational word learning might interact with retention.

Advancing the extant literature, our study addresses a series of questions concerning how a learner can identify a word-referent pairing amongst noise by using environmental cues to reduce referential ambiguity, and how this might affect the subsequent retention of novel words: 1) Can gestural cues compensate for multiple potential referents in the environment, and to what degree? 2) How might cues compensate for referential ambiguity as the learning process unfolds temporally? 3) Do the influences of gesture on referent selection also apply to retained word-learning?

In a series of three experiments, we address each of these questions in turn. In Experiment 1, we investigated how referential ambiguity might be overcome by manipulating the presence of a pointing gesture during conditions that had one object (ambiguity absent) versus those that had two (ambiguity present). Experiments 2 and 3 investigated the temporal process of how gesture cues are integrated with co-existing auditory and visual information to support accurate cross-situational word learning by manipulating the timing of a pointing gesture cue (Experiment 2) and employing an eye-tracker to uncover the dynamics of attention and relations to word learning (Experiment 3). Finally, in each of our experiments, we also tested how our manipulations might affect both immediate recall and retention (after a delay) of novel word-referent mappings.

All pre-registrations, data, viewing of experimental conditions and testing trials, and code for analyses of all experiments in this paper are available on the Open Science Framework (OSF): https://osf.io/2m9pe/?view_only=9d64688d03d84704aa5f2e8f8eb34dc9

3.4 Experiment 1: How do gesture cues interact with referential ambiguity during word learning?

In Experiment 1, we tested whether the presence of a gesture cue can reduce referential ambiguity sufficiently to benefit both immediate recall and retention accuracy. There were four conditions: 1) one object (the target; no referential ambiguity) with a gesture cue; 2) one object without a gesture cue; 3) two objects (one target and one unlabeled foil; referential ambiguity present), with a gesture cue, and 4) two objects without a gesture cue.

If cues aid word learning primarily by reducing referential ambiguity, then participants should not exhibit a difference in performance in one-object conditions regardless of whether a gesture is present – when there is no referential ambiguity, the information afforded by the gesture cue is redundant. However, in the two-object conditions, we hypothesised that participants would perform more accurately when trained with a gesture cue, as the cue would diminish referential ambiguity, leading to greater intake of highly accurate statistical input. If a gestural cue in the two-object condition is sufficient to reduce referential ambiguity, then learners should perform on par with the one-object conditions with and without gesture cues. Furthermore, if the presence of a gesture cue enables participants to benefit from contrasting target and foil objects, the two-object condition with a gesture cue might even yield more accurate performance at test than either of the one-object conditions. Finally, we also hypothesised that immediate test accuracy would predict retention accuracy, and that retention accuracy would be boosted by the availability of gesture cues in the two-object condition.

3.4.1 Method

Twenty monolingual English participants (*M* age = 21.0, *SD* = 1.53; 5 males, 15 females) without any sensory deficits were recruited via leaflets and the Lancaster University research participation system, which allows all members of the University community to partake in research. Informed, written consent was obtained from all individuals prior to participation. Participants were either paid £3.50 or received course credit for taking part. The number of participants was specified in the pre-registration and based on previous studies that test cross-situational word learning using a similar paradigm (e.g. Monaghan et al., 2015; Monaghan & Mattock, 2012).

Materials

All stimuli used can be found in Supporting Information (Appendix B). Thirty-two novel objects and 32 novel two-syllable words were taken from the NOUN database (Horst & Hout, 2016). Sound files for each word were made using the Serena system voice (Macintosh computer, OS 10.13). Each object and word were paired randomly for each participant to produce 32 word-object mappings per participant. Pictures and audio were presented on a Macintosh computer (OS 10.13, 21.5-inch monitor, 1920 x 1080 resolution) using PsychoPy3 (Pierce & MacAskill, 2018). Participants used closed cup headphones.

Procedure

Testing took place in a quiet room. All experiments included two training and test conditions and were run using a similar procedure. Participants completed a warm-up with two familiar objects and words presented as they would be during training. During the first condition, participants undertook the first training block with one set of 16 word-referent pairs, followed by an immediate testing block, then a five-minute distractor task (colouring in a geometric picture), before completing a retention testing block. They then repeated this process with another set of 16 word- referent pairs for the second condition.

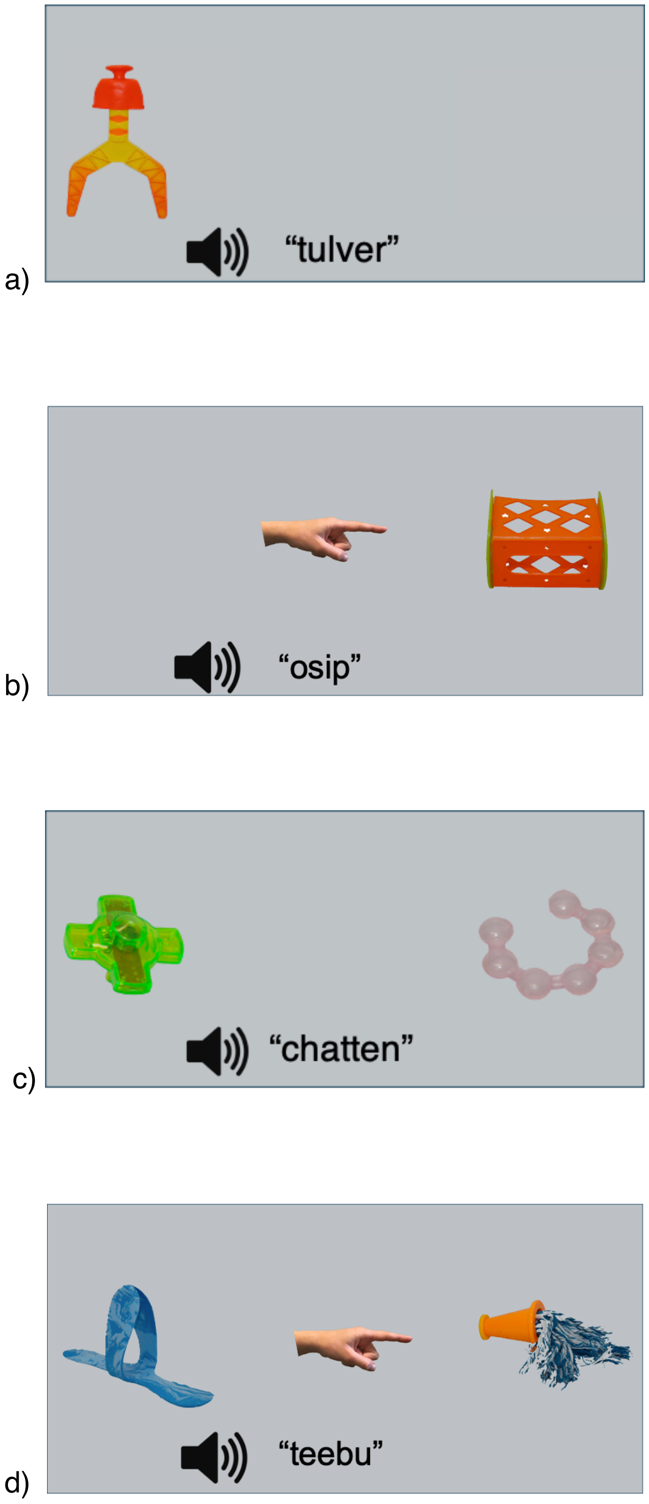
Each correct word-referent pairing appeared four times per training condition, with 16 word-referent pairings to be learnt per condition. Screen position of the objects was pseudo-randomised so that the target appeared an equal number of times on the left and on the right. The order of trials within training blocks was pseudo-randomised with the constraint that referents appeared no more than twice in a row. The order of conditions was counter-balanced across all participants. For all experiments involving two objects, target objects also acted as foils for their non-associated words and were pseudo-randomised with the constraint of appearing an equal number of times across all trials. To ensure participants could disambiguate words and referents based on cross-situational information, co-occurrences of the same targets and foils were minimised across trials.

Training blocks

Participants were randomly allocated to either a one-object or a two-object group. Within each group, they were exposed to both a no-gesture condition and a with-gesture condition, where a picture of a hand pointing to the target appeared simultaneously with the referent (Figure 1). The target word in both conditions was played 500 milliseconds after referent presentation.

Figure 1.

Experiment 1 training trial examples: a) one object, no cue; b) one object, with cue; c) two objects, no cue; d) two objects, with cue (to view experiments on OSF: https://osf.io/2m9pe/?view_only=9d64688d03d84704aa5f2e8f8eb34dc9).

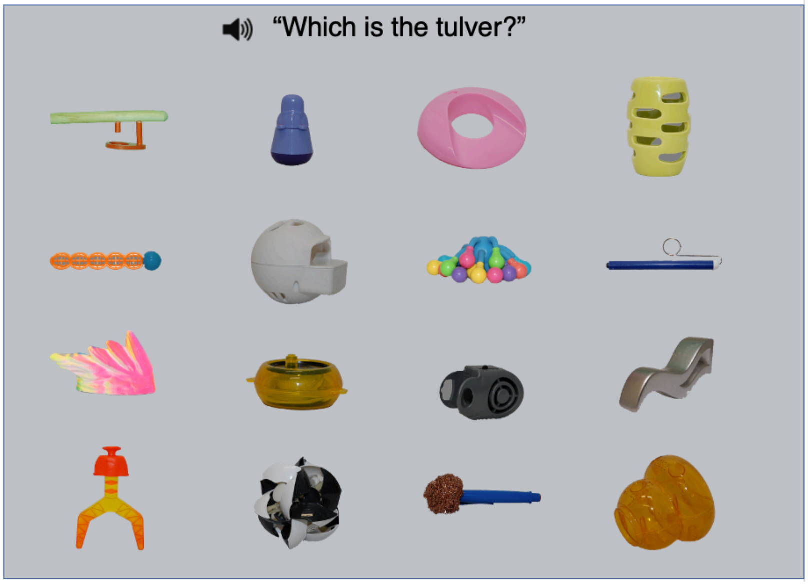


Testing blocks

In order to test learning accuracy for the word-referent pairs, participants were administered two testing blocks: *immediate*, which occurred immediately after training, and *retention*, which occurred after a five-minute distractor task (colouring in a complex picture). Each word was tested on one immediate trial and on one retention trial. During test trials, all 16 referent objects were presented simultaneously on-screen and the learner was asked to click on the correct referent for each target word, requested in a random order (‘*which is the [target word]?*’; chance level = 0.0625; Figure 2). The on-screen positions of the referents differed for immediate and retention trials.

Figure 2.

Example of testing trials for all Experiments: participants see all 16 referents for given condition, and are asked to click on the corresponding object for novel words (to view experiments on OSF: https://osf.io/2m9pe/?view_only=9d64688d03d84704aa5f2e8f8eb34dc9).



“Which is the tulver?”

3.4.2 Results and Discussion

Accuracy of correct word-referent pairs was scored as either 1 (correct) or 0 (incorrect) and entered into a series of general linear mixed effects models (GLMEs; using *glmer* in R, v1.1.463) as the dependent variable. Separate analyses were conducted for immediate testing blocks, retention testing blocks, and all testing blocks combined (i.e. immediate and retention testing blocks). This enabled direct comparison between trial types, reflecting the discrete word learning processes that may underlie immediate referent selection and retention of novel words after a delay. All model sequences began with a baseline model that contained only random effects. Subsequent models were then built progressively by adding individual fixed effects and comparing each model to the previously best-fitting model using log-likelihood comparisons (Barr et al., 2013).

For models predicting immediate testing accuracy, fixed effects of object condition (one object or two objects) and gesture condition (present or absent) were included. For models predicting retention accuracy, a fixed effect of accuracy for each word on immediate testing trials (1 for 'correct' or 0 for 'incorrect') was added to the fixed effects of object and gesture condition. For models predicting overall accuracy, fixed effects of object condition (one object or two objects), gesture condition (present or absent), and trial type (immediate or retention) were included. For all models, random effects of participant, target word, and target object, and test order (one object or two object condition first) were included and random slopes of condition were fitted, unless this prevented the model from converging.

The final best-fitting models to the data, and results for all three analyses, are presented in Table 1 and Figure 3. For all outputs and all model comparisons tested, please see OSF (https://osf.io/2m9pe/?view_only=9d64688d03d84704aa5f2e8f8eb34dc9).

Participants performed above chance in both trial types. For immediate testing trials, the best model fit demonstrated a fixed effect of gesture cue condition ($\chi^2(1) = 9.80$, $p = .002$). Participants were significantly more likely to achieve higher accuracy at test when a gesture cue was present during training (*p* < .001). There was no fixed effect of object condition.

For retention trials, a model with fixed effects of immediate testing accuracy, object condition, gesture condition, and an interaction between gesture cue and object condition provided the best fit ($\chi^2(1) = 8.26$, *p* = .016). This model demonstrated that if participants were correct on their immediate testing trial for a given word-referent mapping, they were significantly more likely to respond correctly on the corresponding retention trial (*p* < .001). The model also demonstrated a reduction in accuracy at test in the two-object training condition overall (*p* = .026). The interaction between gesture cue condition and object condition demonstrated that, in the two-object condition, the presence of a gesture cue during training significantly increased retention accuracy (*p* = .012).

Notably, the interaction effect between gesture cue condition and object condition was only apparent in the retention data. The GLME model for overall accuracy (immediate and retention testing trials together) demonstrated a fixed effect of gesture condition only: across both trial types, participants were significantly more likely to respond correctly when tested on words that were learnt in conjunction with a gesture cue (*p* < .001; model fit, $\chi^2(1) = 9.27$, *p* = .002). Overall, participants scored similarly across two objects with a cue (*M* = 0.72), one object with a cue (*M* = 0.67), and one object without a cue (*M* = 0.63), but scored least accurately when there were two objects without a cue (*M* = 0.45).

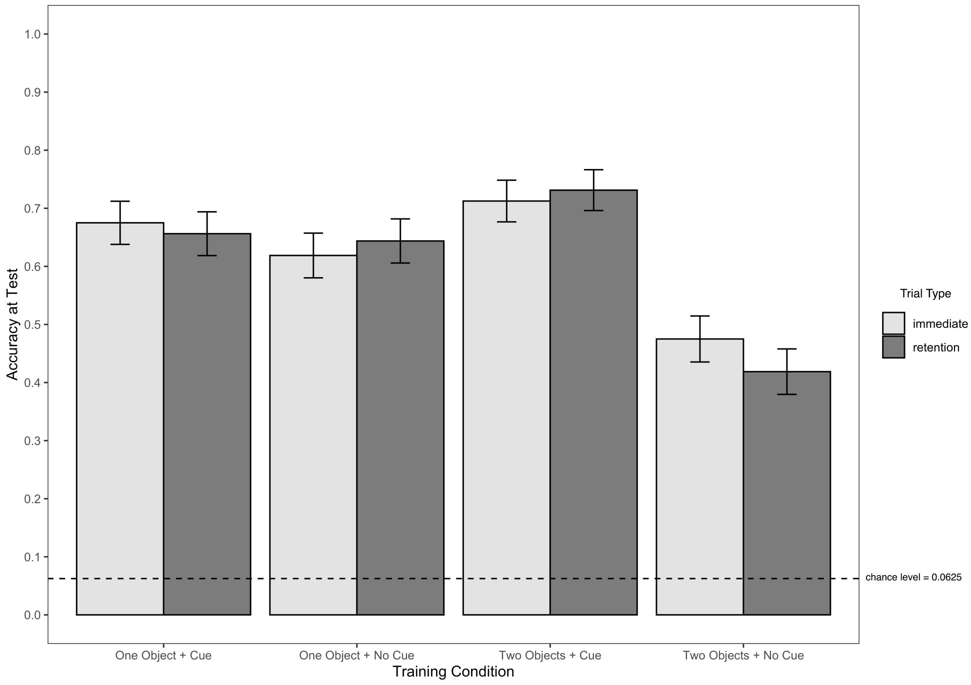
Table 1.

Experiment 1: general linear mixed effects model results predicting immediate and retention trial accuracy and overall accuracy (both immediate and retention trials) by gesture condition (cue or no cue) and object condition (one or two) during training.

Immediate accuracy				
Fixed effects	*estimate*	*SE*	*z-value*	*p-value*
(*intercept*)	0.28	0.37	0.75	.045
Gesture condition (cue)	1.19	0.36	3.35	<.001
Retention accuracy				
(*intercept*)	-0.77	0.40	-1.92	.054
Immediate accuracy (correct)	2.81	0.27	10.36	<.001
Gesture condition (cue)	0.10	0.43	0.24	.810
Object condition (two)	-1.11	0.50	-2.23	.026
Gesture (cue): object condition (two)	1.40	0.56	2.51	.012
Overall accuracy				
(*intercept*)	0.30	0.44	0.68	.50
Gesture condition (cue)	1.36	0.40	3.39	<.001

Figure 3.

Experiment 1: mean accuracy at test and standard error bars in immediate and retention trials across object condition (one or two objects) and gesture cue condition (cue or no cue).



These results show that, although the two-object condition contained substantially more information per trial, the addition of a gesture cue during training resulted in the same degree of learning as when words were learnt in the one-object conditions where there was no referential ambiguity. According to associative learning accounts (e.g., MacWhinney, 2005; McMurray et al., 2012), under conditions of referential ambiguity, learners ordinarily form both target-label and foil-label associations and use this information to gradually narrow down the correct word-referent pairings. Thus, the benefit of gesture as an endogenous cue thus appears to be in the reduction of this referential ambiguity, potentially minimising accidental label-foil associations and strengthening label-target associations.

The results of Experiment 1 thus indicated: (1) immediate testing accuracy is a reliable predictor of later retention of novel word-referent mappings, and the benefit of a gesture cue to retention appears secondary to processes that occur during learning; (2) higher referential ambiguity produces lower accuracy at test, extending the findings of previous studies where greater numbers of referents decreased accuracy under conditions where all referents were named (K. Smith et al., 2009; Yu & Smith, 2007), potentially due to an increased number of spurious associations being made between labels and non-target foils; and (3) the addition of a gesture cue during training may prevent spurious word-referent associations from being formed, improving accuracy on immediate test and subsequent retention trials to the same level as if there was only one referent present during training.

3.5 Experiment 2: when are gesture cues in word learning most useful?

Whilst Experiment 1 demonstrated that the addition of a pointing gesture during training can reduce referential ambiguity to improve learning accuracy, it did not unpack how gesture might be temporally integrated with a novel label during cross-situational word learning. As endogenous cues appear to induce slower attention shifts than exogenous cues (Shepherd & Müller, 1989), gesture cues that occur sometime before, rather than after, a label may be critical to encoding robust label-target associations and minimising spurious label-foil associations. Experiment 2 thus aimed to identify whether cue timing effects apply to the use of pointing gestures in cross-situational word learning.

Experiment 2 fixed the presence of referential ambiguity (two objects) during training and manipulated the timing of gesture cues relative to label utterance across two conditions: one where the pointing hand gesture appeared before the verbal label, and one where it appeared after the label. In the HSP, Trueswell et al. (2016) found that shifting an obscured word 2 seconds earlier than the word's original position was sufficient to reduce the accuracy score of those guessing the missing word from ~ 60% to ~ 43%. Shifting the word onset earlier, rather than after, actual word onset also had a pronounced effect: if the obscuring ‘beep’ was moved too early, guessers did not relate the visual event to the missing word, as they were perceived as too temporally discontinuous to be related to one another. However, shifting attention during word learning between potential referents can happen very quickly. In Halberda’s (2006) mutual exclusivity task that assessed how learners ‘double-check’ their novel word-referent mappings, participants shifted their attention from a known, distractor object to an unknown, target object within 225 milliseconds. We therefore examined the effect of presenting gesture cues by just 1 second before and after a novel label, to see whether sensitivity to cue timing can be observed occurred in a smaller window than tested by Trueswell et al. (2016) in the HSP.

Experiment 1 showed that gesture cues were useful because they reduced referential ambiguity, and the endogenous cuing literature indicates that attention cued before, rather than after, a label will likely be of most use in reducing referential ambiguity to strengthen word-referent mappings. We thus hypothesised that participants would respond more accurately on both immediate and retention trials when tested on words trained in the early gesture condition compared to the late condition. If cues are most useful during learning when they occur early, rather than late, this suggests that cues may best support cross-situational word learning by highlighting the target prior to (or at) label utterance, reducing spurious associations between the label and non-target foils. Late gesture cues may thus be less useful for word-referent mappings as any attentional shift that occurs due to the pointing gesture cue will be after the crucial information (the label) has been uttered, reducing the chance to reconcile the auditory label and the visual referent together and robustly encode the association. In line with Experiment 1, we also tested participants on both referent selection immediately after training, and on retention of novel words tested after a delay of 5 minutes.

3.5.1 Method

Participants

Participants were twenty monolingual English adults without any sensory deficits who had not partaken in any of the other experiments (age *M* = 20.9 years, SD = 5.16, 5 male, 15 female), as specified in the pre-registration. They were recruited and reimbursed as per the procedures outlined in Experiment 1.

Materials and Procedure

The Materials were the same as Experiment 1, and the Procedure was the same except for the following changes.

Training blocks

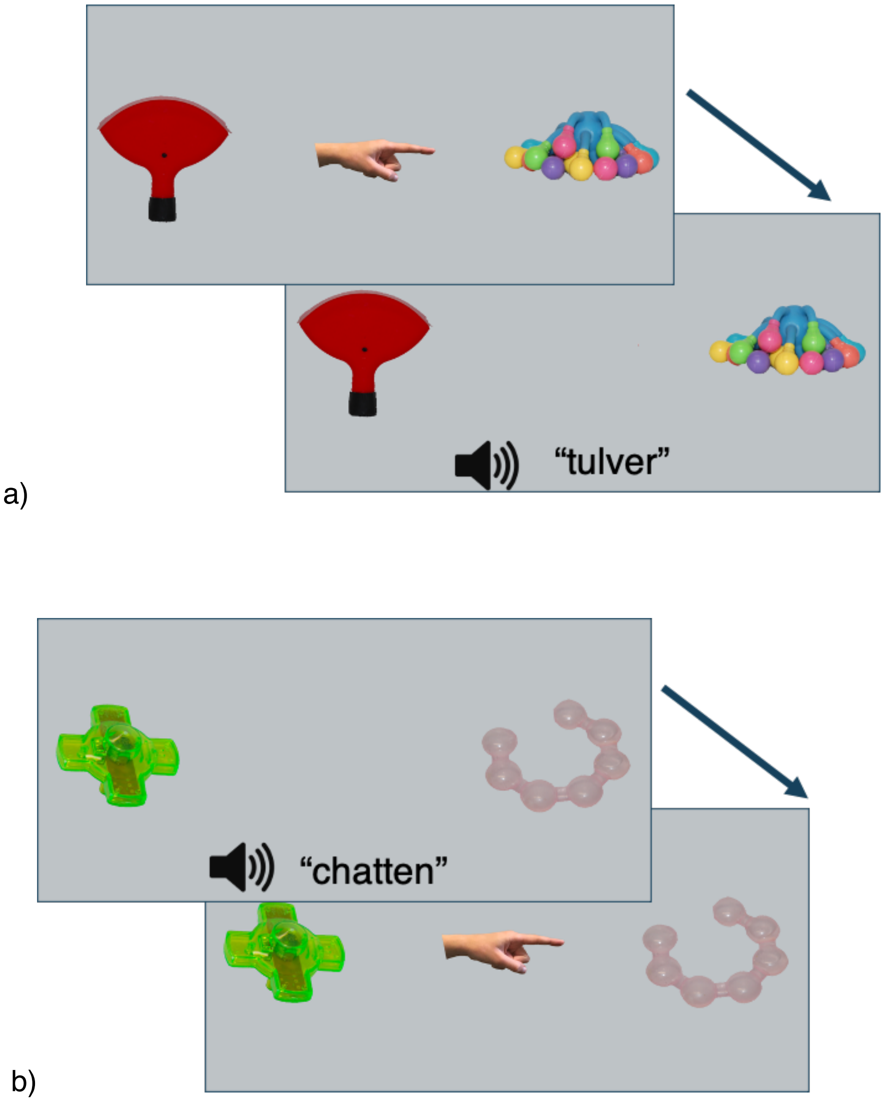
Training followed the same procedure as Experiment 1 with the following changes: at all times, participants saw two objects on screen, and the timing of the gesture cue with the novel label was adjusted to ensure an equal amount of time before and after label utterance in both conditions. In the 'early condition', participants saw the gesture cue 1 second *before* word utterance. In the 'late condition', the gesture cue appeared 1 second *after* word utterance. In both conditions, the two referents appeared for the duration of the trial (3 seconds), label utterance occurred at the same time at the 2 second mark after the referents had first appeared, and the cue lasted for 1 second (Figure 4).

Testing blocks

These were the same as in Experiment 1 (Figure 2), except that, in addition, participants were asked at debrief whether they noticed any differences between the two training blocks. If they answered "no", they were probed specifically about whether they had noticed any difference in the gesture cue.

Figure 4.

Experiment 2 and 3: Training trial examples, a) early gesture condition, b) late gesture condition. See OSF to view experiments (https://osf.io/2m9pe/?view_only=9d64688d03d84704aa5f2e8f8eb34dc9).



3.5.2 Results and discussion

GLMEs were constructed in the same way as described for Experiment 1; only the fixed effects of condition differed. For each model, a fixed effect of gesture condition (early vs. late) was included. The results of all three analyses for Experiment 2 are presented in

Table 2 and Figure 5. Participants performed above-chance in both conditions on both immediate and retention trials.

The best-fitting model for immediate testing trials demonstrated a fixed effect of gesture condition ($\chi^2(1) = 4.21$, $p = .040$). Participants were more likely to respond accurately in the early compared to the late gesture condition ($p = .029$). The best fitting model for retention trials demonstrated a fixed effect of immediate accuracy ($\chi^2(1) = 142.11$, $p < .001$). In line with Experiment 1, if participants responded correctly on immediate test trials for a word-referent pair, they were more likely answer correctly on the corresponding retention trial ($p <.001$).

Overall, participants had higher accuracy (immediate and retention test trials) in the early condition ($M = 0.69$) compared to the late gesture condition ($M = 0.60$). For overall accuracy, the best-fitting model contained fixed effects of both gesture condition and trial type ($\chi^2(1) = 4.47$, $p = .034$), indicating that participants were more likely to respond correctly when tested on words learnt in the early gesture condition compared to the late gesture condition ($p = .008$), and had reduced accuracy in retention test trials overall ($p = .032$).

At debrief, only four of the 20 participants reported noticing a difference between conditions related to the gesture cue. This was unexpected, as the conditions were split into two distinct training blocks and the timing differences between words and gestures spanned a 1 second interval, which we envisaged would be easily detected.

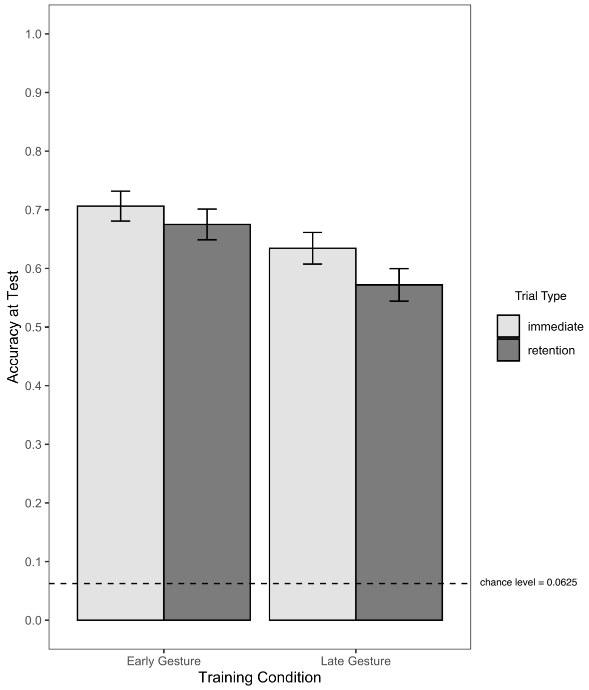
Table 2.

Experiment 2: general linear mixed effects model results predicting immediate and retention trial accuracy and overall accuracy (both immediate and retention trials) by training gesture condition (early or late).

Immediate accuracy				
Fixed effect	*estimate*	*SE*	*z-value*	*p-value*
(*intercept*)	0.79	0.43	1.83	.067
Gesture condition (early)	0.56	0.26	2.18	.029
Retention accuracy				
(*intercept*)	-1.19	0.35	-3.44	<.001
Immediate accuracy (correct)	2.92	0.28	10.59	<.001
Overall accuracy				
(*intercept*)	0.78	0.45	1.75	.081
Gesture condition (early)	0.78	0.29	2.67	.008
Trial type (retention)	-0.30	0.14	-2.14	.032

Figure 5.

Experiment 2: mean accuracy at test and standard error bars in immediate and retention trials across gesture cue condition (early or late).



The results of Experiment 2 indicate that the temporal ordering of cues with word utterance is important when initially establishing word-referent pairs, consistent with the cued attention literature (Hauer & Macleod, 2016; Yoshida and Burling, 2012). Our results not only confirm the importance of cue timing to referent identification identified by Trueswell et al. (2016), but also indicate that the effect of temporal continuity is even more fine-grained than -2 to + 2 seconds under certain circumstances. Gesture cues during training that occur just 1 second before the label utterance significantly improved accuracy at test when compared to those that occurred 1 second after word utterance. Further, the effect of temporal contiguity of gesture and spoken label during referent selection as demonstrated by Trueswell et al. (2016) was also shown to apply to retention of word learning in our cross-situational paradigm. Thus, gesture timing affects learning as well as identification of the target referent in the moment.

Associative models of word learning (MacWhinney, 2005; McMurray et al., 2012; Yu & Smith, 2012) indicate that a learner builds up weights on associations between a label and both targets and foils. We show that directing attention to the target with a pointing gesture cue prior to the word being spoken may prevent the learner from making false associations between a foil and the label, limiting any competing associations. However, cues that occur after the word is spoken do not appear to prevent some competing false label-foil associations from being formed, resulting in reduced accuracy at test relative to the early gesture condition. Applying a cue to indicate the target referent after the label has been spoken, does not provide the same quality of information as when attention is already drawn to the target referent prior to the label being spoken. Therefore, the presence of cues is not the only factor that promotes optimal learning – the contiguity of those cues must also be effective.

Interestingly, only four of the 20 participants noticed that the gesture cue appeared at different time points within trials across the two conditions. This suggests that the temporal synchrony of gestural and spoken information was not consciously available to the majority of participants, meaning that strategic use of information was not driving performance, and indicating that the difference in accuracy at test between the early and late conditions was not due to a conscious manipulation of attention by the learner themselves.

To summarise, Experiment 2 showed that: 1) immediate referent selection accuracy is a predictor of retention accuracy as per Experiment 1; and 2) providing a gesture prior to label utterance yielded superior accuracy in comparison to a late gesture. These results, however, do not yet indicate precisely how learners' attention to objects is affected by the timing of a gesture cue.

We hypothesised that the advantage of early gesture cues over late cues was due to where attention was allocated during rather than following label utterance. We therefore repeated the procedure of Experiment 2, to replicate the behavioural effects, but also with the addition of an eye tracker to monitor participants' gaze during training trials, allowing us to pinpoint where attention is directed during label utterance. We predicted that participants would, as for Experiment 2, perform more accurately when tested on words trained in the early gesture condition, compared to the late gesture condition, and that immediate testing trial performance would significantly predict retention testing trial performance. We made two additional predictions relating to the eyetracking data: (1) if the early gesture cue promotes attention to the target over the referent, then for the early gesture condition (relative to the late gesture condition), participants would have increased overall relative looking time to the target compared to the foil during the training trials, particularly during the spoken label, and (2) if the early gesture cue advantage for learning is due to where attention is located when the word is spoken, then higher accuracy would relate to higher fixations to the target during and immediately after the spoken label, but not prior to the spoken label.

3.6 Experiment 3: how do early gestures support more accurate word learning than late gestures?

3.6.1 Method

Participants

Participants were twenty monolingual English adults without any sensory deficits who had not partaken in any of the other experiments (*M* age = 19.9, SD = 4.15, 5 male, 15 female), as specified in the pre-registration. They were recruited and reimbursed as per the procedures outlined in Experiment 1.

Materials

The materials remained the same as in Experiment 2 with the following exceptions: a Tobii Pro X3-120 eye tracker was used (sampling rate 120Hz) in conjunction with a Windows computer (17-inch monitor, screen resolution 1600 x 900) to track binocular participant gaze throughout training trials. Participants were seated at a distance of approximately 60cm from the eye tracker.

Procedure

Participants' eye positions were calibrated using the Tobii Eye Tracker Manager five-point calibration system before the experiment. The rest of the procedure followed that of Experiment 2.

Processing eye-tracking data

An average of binocular data from the left and right eye was taken to give a single (x, y) co-ordinate for each gaze point. Where data from one eye was missing, the data from the other eye was taken. If data from both eyes were missing, linear interpolation within-participant and within-trial was used to smooth the data.

The data were split into time bins of 250 milliseconds, and three distinct areas of interest (AOIs; cue, foil, and target object) were identified. Fixations within these AOIs were detected using the *saccades* package in R, allowing for isolation of fixations whilst disregarding artifacts such as blinks. All processing code is available on OSF (https://osf.io/2m9pe/?view_only=9d64688d03d84704aa5f2e8f8eb34dc9).

3.6.2 Results and discussion

Accuracy at test (behavioural results)

We first constructed GLMEs analyses in the same way as for Experiment 2 with behavioural response data at test only. The results are presented in Table 3 and Figure 6 and replicated those of Experiment 2. Participants again performed above-chance in all conditions.

In immediate test trials, there was again a significant effect of condition ($\chi^2(1) = 4.28$, $p = .038$) – participants performed more accurately at test on words learnt during the early gesture condition compared to the late gesture condition ($p = .045$). Experiment 3 also demonstrated an additional effect of condition in retention trials where Experiment 2 did not.

A model that included fixed effects of condition and immediate accuracy provided the best fit for retention test trial data ($\chi^2(1) = 5.90$, p =.015). Participants achieved higher accuracy on retention trials for words learned in the early gesture condition ($p = .006$) and, as per Experiments 1 and 2, words that were correctly identified in immediate test trials ($p < .001$).

Overall, participants had higher accuracy (immediate and retention test trials) in the early condition ($M = 0.69$) compared to the late gesture condition ($M = 0.60$). A model predicting overall accuracy with fixed effects of gesture condition, trial type, and an interaction between gesture condition and trial type provided the best fit ($\chi^2(1) = 4.87$, $p = .027$). This model showed that participants performed less accurately in retention test trials ($p < .001$), and the interaction demonstrated that learners were more likely to respond accurately in retention trials for words learnt in the early compared to late gesture condition ($p = .026$). As for Experiment 2, only three of the 20 participants reported noticing a difference between conditions related to the gesture cue timing at debrief.

These behavioural results were very similar to those obtained in Experiment 2. Participants were more accurate at test on words learnt when the gesture cue occurred 1 second before label utterance (rather than 1 second after), they performed worse in retention trials in the late gesture condition, and learners were largely unaware of the difference in timing of the gesture cue – again demonstrating the difference in performance appeared to be independent of any conscious manipulation of attention.

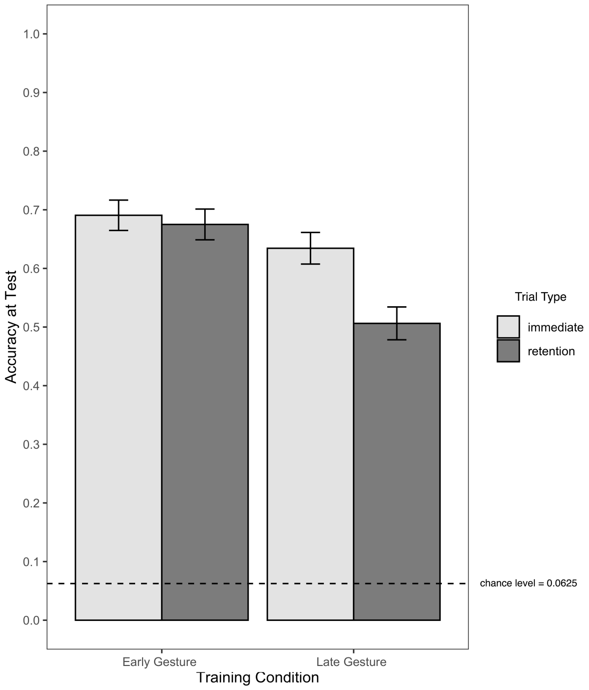
Table 3.

Experiment 3: general linear mixed effects model results predicting immediate and retention trial accuracy and overall accuracy (both immediate and retention trials) by training gesture condition (early or late).

Immediate accuracy				
Fixed effect	*estimate*	*SE*	*z-value*	*p-value*
(intercept)	0.77	0.34	2.29	.022
Gesture condition (early)	0.51	0.26	2.00	.045
Retention accuracy				
(intercept)	-1.62	0.43	-3.74	<.001
Immediate accuracy (correct)	2.58	0.27	9.73	<.001
Gesture condition (early)	0.94	0.34	2.75	.006
Overall accuracy				
(intercept)	0.87	0.40	2.19	.029
Gesture condition (early)	0.46	0.29	1.60	.109
Trial type (retention)	-0.73	0.19	-3.83	<.001
Gesture (early): trial type (retention)	0.62	0.28	2.23	.026

Figure 6.

Experiment 3: mean accuracy at test and standard error bars in immediate and retention trials across gesture cue condition (early or late).

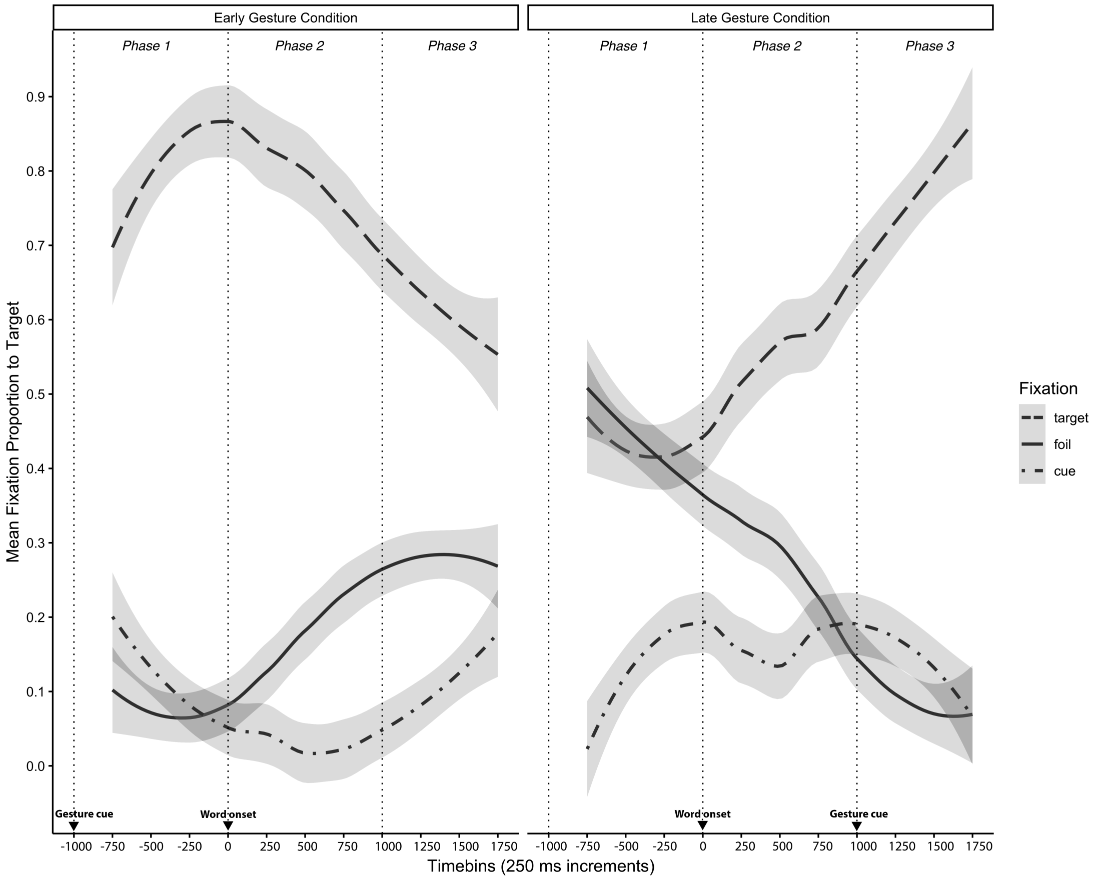


Target fixation proportion during training

The time course of eyetracking data over training trials is illustrated in Figure 7, which shows how mean fixation proportion to target, foil, and cue alters across trial time by condition (calculated using the function *geom_smooth* in the *ggplot2* package in R[v1.1.463], utilising local polynomial regression fitting). In the early gesture condition, participants looked predominantly at the target with a peak around word utterance, but began to look at the foil towards the end of the trial. In the late gesture condition, fixations at the beginning of the trial were split roughly equally between target and foil, but participants began to discriminate between target and foil after word utterance, with fixation to target rising after the gesture cue.

Figure 7.

Experiment 3: eye-tracking data time course during training trials, showing mean fixation proportion to target, foil, and cue during training by trial time at each 250ms time bin, separated by gesture condition (early and late), aggregated across all participants and trials. *Phase 1* = after gesture cue in early condition and before word occurrence in both conditions; *Phase 2* = after word onset; *Phase 3* = after gesture in late condition. [2]



To examine the effect of gesture cue timing on the learning process during training, we employed growth curve analysis (GCA) to analyse target fixation proportion across conditions. GCA allows for modelling of differences between participants whilst allowing for

[2] Due to technical issues with the eyetracking equipment, some data from the beginning of trials was lost; values on the x-axis indicate time bin interval start time.

within-participant differences across time (Mirman et al., 2008). We used the best-fitting orthogonal polynomials for the time form function, testing up to cubic polynomials.

GCAs were fitted according to Mirman (2014) using *lme4* in R (v1.1.463). A baseline model was constructed that predicted mean fixation proportion to target with fixed effects of all time terms, and random slopes of all time terms per participant, and random slopes of time terms for each participant for each condition. These models failed to converge despite applying techniques to retain maximal random effects structure (Barr et al., 2013; Mirman, 2014), resulting in a baseline model of all time terms with random effects of all time terms per participant. Subsequent models were then built up by adding a fixed effect of gesture timing condition (early or late) to the intercept only, and then adding a fixed effect of gesture timing condition to all time terms. Each model was compared to a baseline model, or previous best-fitting model, using log-likelihood comparisons. For all models, the early gesture training condition was used as the reference level.

The GCA model and data fits are shown in Figure 8,[3] with Table 5 showing fixed effect parameter estimates and standard error (*p*-values estimated using normal approximation for *t*-values). The overall time course of mean target fixations was best captured with a third-order (cubic) orthogonal polynomial ($\chi^2(1) = 20.22$, $p < .001$). The effect of condition on the intercept improved model fit on the intercept and all time terms (all $p < .001$). The GCA analysis indicated that target fixation proportion was significantly different between the two conditions, with participants exhibiting a mirrored effect (Figure 9): participants in the early condition looked longer at the target at the beginning of trials and decreased their fixation over the duration of trials, whilst participants in the late condition looked less at the target at the beginning of trials and increased their fixation over the duration of trials. To further test where differences between the early and late condition were significant, a series of post-hoc independent samples two-tailed *t*-tests for each time bin

[3] The drop in fixation proportion to target at time bin 8 (2000 ms) in the late condition was most likely due to the appearance of the cue, but this was not captured by a quartic orthogonal polynomial.

were carried out. These reflected the same pattern as the GCAs; the *t*-tests demonstrated a significant difference at all time bins except the time bin at 1750ms (8 out of 11 time bin differences were *p* <.001; Table 6).

Figure 8.

Experiment 3: growth curve analysis fitting a third-order orthogonal polynomial to mean fixation proportion to target by trial time at each 250ms timebin, separated by gesture condition (early and late), aggregated across all participants and trials. Data points indicate mean and standard error bars for target fixation proportion; lines indicate model fit.

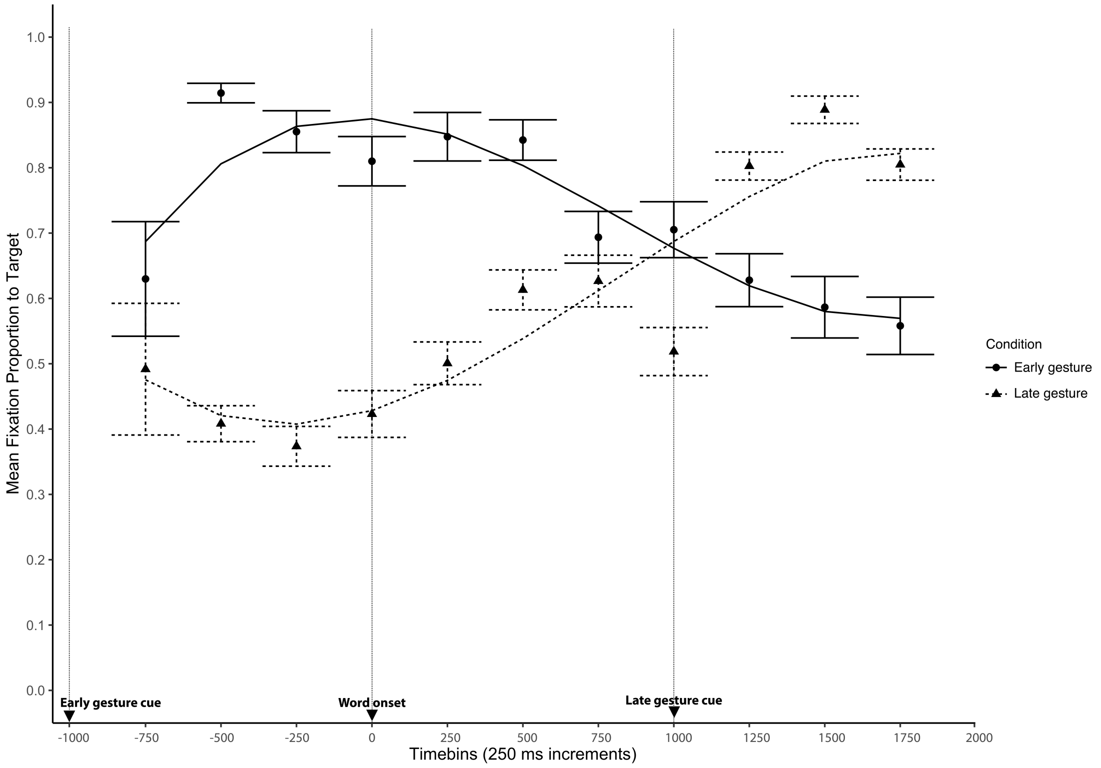


Table 5.

Experiment 3: results of growth curve analysis of mean target fixation proportion. Estimates of time terms between gesture condition (early or late) and model comparison.

	Early gesture condition				Late gesture condition			
Term	*estimate*	*SE*	*t-value*	*p-value*	*estimate*	SE	t-value	*p-value*
(intercept)	0.73	0.02	43.74	<.001	-0.15	0.02	-9.16	<.001
Linear	-0.26	0.06	-4.37	<.001	0.75	0.05	13.95	<.001
Quadratic	-0.20	0.05	-4.19	<.001	0.34	0.05	6.39	<.001
Cubic	0.14	0.05	2.99	<.001	-0.24	0.05	-4.56	<.001

Model comparisons	χ^2*(df)*	*p-value*
(intercept)	45.49(1)	<.001
Linear	137.74(1)	<.001
Quadratic	36.38(1)	<.001
Cubic (full)	20.22(1)	<.001

Table 6.

Experiment 3: post-hoc t-tests comparing mean fixation proportion to target at each 250ms time bin across all trials between conditions.

	Early		Late		Comparison		
Time bin, ms	*Mean*	*SE*	*Mean*	*SE*	*t-value(df)*	*95% CI*	*p-value*
-750	0.75	0.07	0.82	0.07	-0.72 (25.57)	-0.28, 0.13	.048
-500	0.91	0.01	0.41	0.02	16.16 (29.32)	0.44, 0.57	<.001
-250	0.86	0.03	0.37	0.03	10.89 (37.90)	0.39, 0.57	<.001
0 (word onset)	0.81	0.04	0.42	0.04	7.43 (37.89)	0.29, 0.49	<.001
250	0.85	0.04	0.50	0.03	7.01 (37.39)	0.25, 0.45	<.001
500	0.84	0.03	0.61	0.03	5.28 (38.00)	0.14, 0.32	<.001
750	0.69	0.04	0.63	0.04	1.20 (38.00)	-0.05, 0.18	.238
1000	0.71	0.04	0.52	0.04	3.31 (37.16)	0.07, 0.30	.002
1250	0.63	0.04	0.80	0.02	-3.81 (28.90)	-0.27, -0.08	<.001
1500	0.58	0.04	0.89	0.02	-5.87 (26.25)	-0.41, -0.20	<.001
1750	0.56	0.04	0.80	0.02	-4.93 (29.4)	-0.35, -0.14	<.001

In line with our hypothesis, participants were more likely to fixate on the target before and during the word utterance in the early compared to the late condition. However, the increase in target fixation prior to cue onset over trials in the late gesture condition demonstrates that, over multiple exposures to word-referent pairs, participants could identify the correct target prior to the cue's appearance. The cue in the late condition thus appeared to act as a confirmation of a referent, whereas in the early gesture condition, the cue appeared to act as a predictor of the referent prior to label occurrence. We then assessed how these patterns during training might have affected participants' performance at test.

Predicting accuracy at test by target fixation proportion during training

To examine the effect of looking behaviour during training on word learning accuracy and address our hypothesis that the early gesture cue advantage was secondary to where attention was directed during label utterance, two analyses were performed. The first assessed *when* target fixation during training trials might be the most crucial predictor of selecting correct word-referent pairs at test, and the second was to assess whether target fixation differed across multiple exposures to the word (4 exposures in total per word-referent pair) during training. An added fixed effect of condition was not included due to a high variance inflation factor between condition and target fixation proportion (>3; Zuur et al., 2010).

Analysis 1: When does target fixation during training predict word learning accuracy?

Figure 8 shows diverging fixation proportion to the target across the early and late gesture conditions, and we sought to identify when looking behaviour during training trials had the biggest effect on accuracy at test. To achieve this, target fixation data were split into three distinct training phases that matched specific events within the training trial, each comprising four time bins (Figure 8):

a) Phase 1: before the verbal label in both conditions, and after cue occurrence in the early gesture condition (-1000 – 0 milliseconds)
b) Phase 2: after the verbal label in both conditions (0 – 1000 milliseconds)
c) Phase 3: after the occurrence of the gesture in the late condition (1000–2000 milliseconds)

The first set of GLMEs were constructed with fixed effects of eye-tracking behaviour for each of these time periods and built in the same format as for all other experiments. Only the fixed effects differed; instead of a fixed effect of condition, average fixation proportion to target for each of the training Phases (per word and per participant; coded as Phase 1, 2, and 3) was used. Interactions between time periods were not tested due to high VIF values within interaction models.

There was a significant effect of fixation proportion to target in Phase 2 (Table 7) on immediate trial accuracy ($\chi^2(1) = 6.10$, *p* = .014), retention trial accuracy ($\chi^2(1) = 4.17$, *p* = .041), and overall accuracy ($\chi^2(1) = 8.47$, *p* = .004). These results indicate that the more participants looked to the target immediately after labelling, the more likely they were to correctly identify the word-referent relationship when tested immediately after training and following a 5-min delay. GLMEs fitted for Phases 1 and 3 did not demonstrate a significant effect of fixation proportion to target on accuracy in any of the test trials, indicating that looking behaviour during training before the word occurred and after the cue occurred in the late gesture condition did not influence performance at test.

Table 7.

Experiment 3: general linear mixed effects model results predicting immediate and retention trial accuracy and overall accuracy (both immediate and retention trials), with fixed effects of fixation proportion to target during training. Only fixation proportion during training Phase 2 (after the label utterance) was a significant predictor of accuracy.

Immediate accuracy				
Fixed effect	*estimate*	*SE*	*z-value*	*p-value*
(intercept)	0.09	0.42	0.23	.822
Target fixation proportion (Phase 2)	1.13	0.45	2.54	.011
Retention accuracy				
(intercept)	-1.68	0.55	-3.05	.002
Immediate accuracy (correct)	2.56	0.27	9.64	<.001
Target fixation proportion (Phase 2)	1.13	0.55	2.05	.041
Overall accuracy				
(intercept)	-0.10	0.49	-0.20	.084
Target fixation proportion (Phase 2)	1.07	0.36	2.97	.003

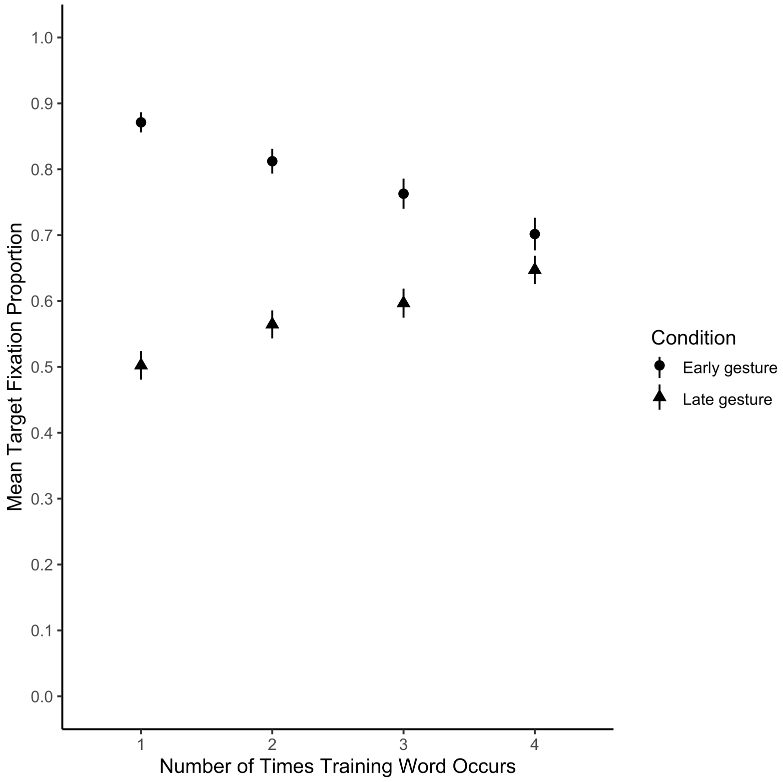
Phase 2 = after label utterance.

Analysis 2: Does word-referent exposure influence accuracy?

Having identified Phase 2 as the crucial time period in training, we next examined whether there was any effect of word-referent exposure. We conducted a further analysis of fixation proportion to target during Phase 2, taking into account the number of times participants had been exposed to the word-referent pair. Each word-referent pairing had four exposures during training, and the expectation of cross-situational word learning is that participants successfully learn word-referent pairs after multiple exposures. This was reflected in the GCA analysis of eye-tracking data; in the late condition, participants fixated on the target over multiple trials after word occurrence even before the cue appeared (Figure 8). Figure 9 also illustrates how participants looked less at the target during label utterance in the early gesture condition as word-referent exposure increased, whereas they exhibited the opposite pattern in the late gesture condition, looking more at the target at label utterance with multiple exposures. These profiles likely reflect different learning strategies over time between the two conditions.

Figure 9.

Experiment 3: Mean fixation proportion to target and standard error bars during label utterance (Phase 2; Figure 7) by word-referent exposure (the number of times participants were exposed to novel word-referent pair), separated by gesture condition (early and late), aggregated across all participants, all words, and all trials.



Fixation proportion data were split into the first and second times the word-referent pairing occurred (first exposures) and the third and fourth times the word-referent occurred (last exposures) during training. Models were constructed using the same processes as described previously, with fixed effects of fixation proportion to target during first and last word-referent exposures (as separate fixed effects), immediate accuracy (coded as '1' for correct or '0' for incorrect) for retention trial analysis, and trial type (immediate or retention) for overall accuracy.

For immediate accuracy, the best fitting model included a fixed effect of average target fixation proportion during first word-referent exposures ($\chi^2(1) = 6.19$ $p = .013$; Table 8), indicating that the more participants looked at the target during first exposures to word-referent pairs, the more likely they were to correctly identify word-referent pairs during immediate trials ($p = .012$). However, there was no significant effect of last exposures, or added effect of condition.

For retention data, a model with fixed effects of immediate accuracy and average fixation proportion to target during the first exposure to word-referent pairs, proved to be the best fit ($\chi^2(1) = 4.82$, $p = 0.28$; Table 8). This indicated that participants were more likely to be accurate in retention test trials if they had been correct on the corresponding immediate test trial ($p < .001$), and fixated longer on the target during the first two exposures to the word-referent pair ($p = .026$).

For target fixation data predicting overall accuracy, fixed effects of trial type and average target fixation proportion during first word-referent exposures were found ($\chi^2(1) = 8.54$, $p = .003$, Table 8). Participants responded less accurately on retention trials than immediate trials ($p < .001$), and more accurately overall if they fixated longer on the target during the first two exposures to the word-referent pairs (p = .003).

Table 8.

Experiment 3: general linear mixed effects model results predicting immediate and retention trial accuracy, and overall accuracy (both immediate and retention trials) at test with fixed effects of average fixation proportion to target during Phase 2 (after label utterance) categorised by word-referent exposure (first exposures or last exposures).

Immediate accuracy				
Fixed effect	*estimate*	*SE*	*z-value*	*p-value*
(intercept)	0.43	0.43	1.00	.317
Target fixation proportion (first exposures)	0.92	0.36	2.52	.012
Retention accuracy				
(intercept)	-1.63	0.53	-3.07	.002
Target fixation proportion (first exposures)	1.06	0.48	2.23	.026
Immediate acc. (correct)	2.61	0.30	8.81	<.001
Overall accuracy				
(intercept)	0.31	0.54	0.58	.559
Target fixation proportion (first exposures)	0.97	0.32	3.01	.003
Trial type (retention)	-0.52	0.15	-3.42	<.001

First exposures = first and second exposure to correct word-referent pair

Together with the GCA analysis, these results indicate that participants learned words more accurately when the gesture occurred 1 second before the word, rather than 1 second after, primarily because they exhibited higher target fixation during the period surrounding label utterance. Furthermore, on the first exposures to word-referent pairs, participants already demonstrated higher target fixation proportion during label utterance in the early gesture training condition (Figure 10), which predicted higher accuracy at test.

Overall, Experiment 3 demonstrated several key findings: 1) we replicated the results of Experiment 2, showing that participants again performed more accurately in a condition where gesture occurs before, rather than after, word utterance, 2) participant fixation to target immediately after label utterance – within 1 second – resulted in the most accurate word learning, 3) participants were more likely to be fixating upon the target at this crucial time in the early, rather than late, gesture condition during training, and 4) these effects are already apparent during the first exposures to novel words, providing a boost to word acquisition at the point where the identity of the referent is uncertain (given the low number of occurrences of cross-situational correspondences at the beginning of training).

3.7 General Discussion

The role that cross-situational statistics can play in word learning is well documented. However, the mechanisms through which environmental cues facilitate cross-situational word learning are not well understood. In this paper, we showed how studies of cue use in language acquisition are aligned with the long-standing tradition of studies of visual attentional cueing. We highlighted how the effectiveness of gesture use in language learning is matched to the timing of endogenous cue reorientation, potentially tailored to exploit the coordination of attention at the moment that the speaker provides the label in order to optimise word learning.

Experiment 1 demonstrated that providing an informative gesture cue can effectively eliminate referential ambiguity when learning from two objects, leading to performance on par with – but not exceeding – the single object conditions. Experiment 2 demonstrated that early gesture cues under referential ambiguity yield superior learning to late gesture cues, indicating that *when* cues occur in relation to label utterance has a direct influence on the learner’s accurate mapping of word-object associations. Experiment 3 replicated these gesture timing results, but also confirmed that this superior learning was due to the cue directing attention to the target referent during label utterance. Finally, all experiments demonstrated that immediate referent selection accuracy was a predictor of later retention accuracy, and that this effect was a stronger predictor of retention than any manipulation of gesture condition – indicating that the dynamics of the referent selection process is vital to retention later on (McMurray et al. 2012; Yu and Smith, 2012).

These results are consistent with studies that examine the time course of how and when endogenous cues orient attention to objects (Berger et al., 2005; Yoshida & Burling, 2012). However, these effects have not previously been merged with studies of word learning, and our study investigating cross-situational word learning with different temporal arrangements of gesture cues shows how endogenous cueing by a speaker can interplay with speech to support learning.

Studies that examine gesture cues under naturalistic settings also indicate different effects of temporal order for cued attention during word learning. When analysing discourse during semi-naturalistic mother-infant interactions, Frank et al. (2013) found that pointing gestures were used to introduce new topics and tended to be largely used at the beginning of discourses about objects. Further, children in this study also looked less at an object as it was talked about more, mirroring the pattern of target fixation behaviour in the early gesture condition of Experiment 3 (Figures 9 and 10). Relatedly, Griffin and Bock (2000) found that words tend to occur 1 second after speakers look at a target object in naturalistic settings, and gestures also appear more frequently before, rather than after, speech in these naturalistic settings (Bergmann et al., 2011). Coupled with this, novel words are learnt by infants most accurately when they are centred in view and largest in size during label utterance (Pereira et al., 2014), and children's attention to referents is highest during, and just after, label utterance in naturalistic mother-infant interaction videos (Trueswell et al., 2016). Taken together, there is sufficient evidence to suggest that the benefit of an early gesture cue is consistent with the literature concerning word learning in naturalistic settings, and the endogenous cue literature indicates that this dovetails with how attention is directed during the time course of word learning.

Overall, the benefit of gesture cues to word learning in this context to be mediated by quality rather than quantity: it may matter more *when*, rather than *how much,* a learner fixates upon a target referent. As Experiment 3 demonstrated, simply looking at a target prior to label utterance is not sufficient to improve learning. Our analysis of target fixation prior to word occurrence during training (Experiment 3, Phase 1) did not predict accuracy at test, despite participants in the early gesture condition having more time to fixate on the target before label utterance. Rather, the predictive value of early gesture cues leads to a learner fixating upon the correct referent as label utterance occurs from the very first exposures to novel word-referent relationships, and this may confer an advantage in word-referent mapping at test. This difference is apparent even by varying the relative timing of the gesture cue to label utterance by only 1 second, as participants performed significantly less accurately in the late gesture cue condition across both Experiments 2 and 3. Consistent with these findings, MacDonald et al. (2017) found that adult learners still tracked a single hypothesis and spent less time on alternative word-referent pairs when a gaze cue to a target object was present (as opposed to absent) even after being given the same amount of time to visually inspect the objects during cross-situational training in both conditions. The authors suggested this was because gaze increased opportunity to maintain attention on the target referent.

When examining adult cross-situational word learning, Yu et al. (2012) found that strong and weak learners exhibited a pattern of looking behaviour that only began to differ around the middle stages of their training, likely due to the gradual aggregation of statistical co-occurrences over time. This is consistent with our results in Experiment 3, where participants in the late gesture condition increasingly fixated on the target over trials with increased word-referent exposure (Figure 10). However, during the early gesture condition, participants began trials by fixating upon the target because they were cued towards it. In Yu et al. (2012), strong learners had increased attention to the referent towards the end of trials, rather than the beginning. With an early gesture cue, learners in Experiments 2 and 3 may have been provided with a shortcut that enabled them to direct their attention towards the target from the very first exposure, resulting in more accurate performance at test. This is in line with the eye tracking data that showed that fixations to target in the first exposures to word-referent pairs, rather than the last exposures, were predictive of word learning accuracy.

Of further note is that learners performed more accurately with a cue during training even when it occurred in the late condition (yielding mean immediate test trial accuracy of 0.63; Experiments 2 and 3, late gesture condition) than under cross-situational word learning without a cue (mean immediate test trial accuracy of 0.48; Experiment 1, two object condition, no cue condition) – confirming that additional cues are better than no cues at all.

Looking to the referent when the label is uttered may provide an advantage through longer looking time and attention to the referent. This will increase the strength of association between the label and target referent, which builds up gradually over multiple learning situations. Additionally, the reduction in looking to the foil object will likely reduce occurrence of spurious associations between labels and foils, which can have a further beneficial effect in supporting learning the precise word-referent mapping intended by the speaker (e.g., Yu & Ballard, 2007; McMurray et al., 2012).

A further benefit for learning may result from differences in prediction resulting from the gesture cue occurring before the label. Ramscar et al. (2010) manipulated the ordering of objects and labels during word learning in adults and children (24–30-month-olds) and found that, when objects were presented prior to labels, learning was more accurate than when labels preceded objects. This was thought to be due to the informativeness of whether a label or an object acts as a conditioning cue; the occurrence of objects prior to labels enables learners to process object features as distinctive cues that competed for relevance when predicting the label, whereas being exposed to the label first provides a far more constrained source of information to predict objects from. Consistent with this, learners in our study appeared to use the early gesture cue as a predictor of the referent, whereas in the late gesture condition, the gesture may have simply served as a confirmatory check on the participant's assumption, resulting in a weaker prediction for the learner.

Curiously, only seven out of the 40 participants tested across Experiments 2 and 3 noticed that the cue differed in timing. This suggests that learners have little meta-awareness of the context surrounding word learning itself, and consequently likely have little explicit control over how gesture, labels, and referents are sequenced in communicative situations. Other studies also report that learners are surprised at how they perform at test (e.g. Yu & Smith, 2007), and explicit awareness of learning during cross-situational word learning can boost performance at test relative to implicit awareness (Kachergis et al., 2014; Monaghan et al., 2021). Further studies could further examine meta-awareness in relation to different types of cues to assess whether this affects performance. More practically, it is an open question as to whether instructing caregivers to gesture before labelling could effectively alter their behaviour, potentially creating more optimal learning conditions for their child.

Limitations and future directions

There are a number of limitations to this study. Firstly, although trial lengths were the same across conditions in Experiment 2 and 3, and objects appeared on screen for the same amount of time in both conditions, the in-trial duration that learners could make use of the gesture was not identical. Due to the nature of the paradigm, learners had the duration of the trial to study the correct referent on screen during the early gesture condition, whereas learners had far less time during the late gesture condition (where gesture occurred after the label utterance). This may have led to more accurate performance in the early gesture condition due to the extra time spent examining the target referent. However, our eye tracking analysis indicated the time period around label utterance during training to be the crucial point relative to performance at test, suggesting that the quality of time spent (i.e. focusing directly on the target when label utterance occurs) is more important than quantity. Furthermore, allowing participants to fixate upon target referents for equal amounts of time across a cued and non-cued conditions has not influenced word learning accuracy elsewhere (MacDonald et al., 2017).

In our experiments, we used a cutout photograph of a pointing finger and hand to act as a gestural cue (Figures 3 and 4). It is possibly debatable whether this could be truly considered a pointing cue, with anything approaching a realistic level of social and pragmatic engagement. Future studies could address this issue by further examining the role of social versus non-social cues under the same conditions of referential ambiguity, or even weigh different types of social cues, such as dynamic eye gaze (head turns with accompanying eye gaze; e.g. MacDonald et al., 2017) against one another. Similarly, whilst it is possible that an arrow might yield the same results, the advantage of using a pointing gesture cue over an arrow is simply that they play a more prominent role in naturalistic language acquisition. However, whether visual attention grabbers such as lights and arrows outweigh those of social cues such as head turn and eye gaze are currently addressed elsewhere (e.g. see Axelsson et al., 2012; Hartley et al., 2020; Wu & Kirkham, 2010). By using a pointing gesture cue, we are assuming that the learner takes for granted that the cue itself is *meant* to be used, regardless of its origin. Humans as learners are adaptive and resourceful, and are likely able to make use of multiple cues in multiple ways. Many models of word learning fundamentally rely upon this notion, with research indicating that children and adults adapt and weigh different cues according to the task at hand (Hollich et al., 2000; MacDonald et al., 2017; Monaghan, 2017; Yu and Ballard, 2007).

Of note is that, despite pointing cues being reliable indicators of referents, they nonetheless occur relatively rarely in naturalistic learning environments. Frank et al. (2013) report that in their semi-naturalistic mother-infant video corpus they had a recall value of 0.10/1, whereas maternal eye gaze had a recall value of 0.36/1. In Trueswell et al., (2016), highly informative vignettes that contained maternal gesture were rare, and in Iverson et al. (1999), mothers only used pointing cues during word learning 15% of the time. Pointing cues then, as useful as they are, are but one of several cues that can supplement cross-situational word learning.

Finally, we did not test during training, opting instead to test all word-referent pairs at the end of training. Although testing trial-by-trial would have provided a direct measure of choice by learners throughout training, this might have encouraged learners to make trial-by-trial hypotheses (Kachergis et al., 2014), and it would not have allowed participants to make word-referent selections when faced with all objects at test for the first time. Testing during training may also increase learning simply by way of forcing participants to choose an object after each trial.

Conclusion

In conclusion, this series of experiments offers multiple insights into how cues can facilitate disambiguation of meaning when the learner is faced with referential ambiguity. The value of gesture cues appears to be in compensating for referential ambiguity by providing accurate information about referents. Gesture cues are particularly useful when highlighting referents prior to labels; when a perfectly disambiguating gesture cue occurs before a novel word is spoken, this provides a superior benefit to the learner than when a gesture occurs after a novel word – although a late gesture provides more benefit than no gesture at all. These temporal effects are consistent with how gestural cues interoperate with speech in naturalistic studies, and show how the attention literature around endogenous cues is also applicable across cross-situational word learning. The experiments presented here provide a controlled setting that demonstrates how and when gesture can support cross-situational statistical learning, and furthermore, translate well-investigated attention and memory effects into effects of cueing during word learning.

‘By 12 months, Mia had not said a word; no mama or dada, just high-pitched sounds. Nibby, Mia's mother, was told all children were different. But Mia was different: she could not speak.

At two, Mia was referred to a specialist communication disorder team, but it was a year before an appointment was issued. By this time, Nibby remembers Mia as "a screaming bundle of frustration".

"At nights sometimes she would sob in my arms," says Nibby, of Pateley Bridge, North Yorkshire. "We were lost in a nightmare with a child with terrible difficulties, no support and nowhere to go facing a really hard homelife of two-hour-long tantrums, sometimes longer (...)

A few days ago, Nibby asked Mia a question. "She answered yes, and it made me feel like I've won the lottery," said Nibby. But bad days, Mia hides her mouth behind her hands and screams the place down.’

- Czernik, A., 2013, Apr 11. Mia, 7, wants to learn and play. *The Guardian.*

4 Chapter 4: Word learning in atypical language development: Late talking children

4.1 Why study late talking children?

Thus far, this thesis has described word learning in typical development. The previous two studies focused on the initial stages of word learning; in the following section word learning is approached from the opposite direction by categorising learners through their output and vocabulary knowledge, and determining what processes and mechanisms contribute to differences in production.

Late talking (LT) children exhibit a delay in talking without concurrent sensory, neurological or cognitive disorders. Identified at approximately 2-years-old, LT children in the research literature fall at the 10th percentile or below in expressive vocabulary compared to typically developing (TD) peers (Desmarais et al., 2008). They are a heterogenous group (Rescorla, 2011) and have a range of receptive vocabulary (Fisher, 2017). For example, Wake et al. (2011) report a range of 73 – 103 on the auditory Preschool Language Scale-3 (standardised score, norm = 100) for 283 LT children aged 2-years-old (identified as LT at 18-months-old, <20th percentile, expressive CDI). Many LT children appear to recover (Rescorla, 2011), but a significant subset are subsequently diagnosed with Developmental Language Disorder, a persistent delay in language without biomedical cause (previously known as Specific Language Impairment, SLI; Bishop et al., 2017). Early intervention for children at risk of developmental delay has shown evidence of improved outcomes (Conyers et al., 2003; Winter & Kelley, 2008). As early language difficulties are linked to poor social and academic outcomes (Law et al., 2009), employing a ‘wait and see’ approach to LT children may run the risk of missing a key intervention period to the detriment of individuals later on (Singleton, 2018; Collisson et al., 2016).

Distinguishing between children who will continue to struggle with language development, and those who will not, is an on-going challenge for researchers. Potential risk factors explain only a small proportion of the variance in outcomes and have been inconsistent predictors of language ability at best (H. D. Nelson et al., 2006; Reilly et al., 2010). Furthermore, although some outcome studies have found that most LT children reach the normal range of vocabulary by school age, LT children continue to perform worse than their TD peers on reading and language tasks, suggesting that the difference between the two groups persists. Research investigating mechanisms of LT is still emerging, but suggests LT children may rely on different strategies to TD children during word learning (Moyle et al., 2007; Weismer et al., 2013). Thus, delineating the mechanisms that underpin differences between LT and TD children early on is key to understanding firstly, why this difference appears to transcend certain risk factors (such as socioeconomic status (SES) or family history of language delay), and secondly, how to identify LT children with persisting severe deficits early on with the hope of improving their later outcomes.

4.2 Epidemiology

Accurate prevalence is difficult to ascertain, partly due to the different measures used to determine LT status. Within research, parent-reported checklists of productive vocabulary are the most common. These includes the *MacArthur Communicative Development Inventories: Words and Sentences (CDI)*, where LT children fall at or below the 10th percentile of the CDI (92/680 words for females and 63/680 for males; Fenson et al., 2007), and the *Language Development Survey (LDS),* where LT children produce fewer than 50 words or no word combinations (Rescorla, 1989). Although most studies use the 10th percentile as a cut-off for language impairment using the CDI (Fisher, 2017), it is worth noting that cut-offs can range between the 5th – 30th percentile (Colunga & Sims, 2012; Girolametto et al., 2001; MacRoy-Higgins et al., 2013), introducing further variation in LT classification. Other measures such as the *Reynell Developmental Language Scales* (RDLS; Clegg et al., 2015) and the *Ages-and-Stages Questionnaire* (ASQ; Zubrick et al., 2007) are also sometimes used. Age at identification and assessment in LT studies ranges from 18–35 months (Fisher, 2017; Rescorla, 2011), largely due to the inherent individual differences present in early language development (although most studies use 24 months of age as a benchmark for defining LT status; Fisher, 2017).

Large-scale studies, such as the Early Language in Victoria Study (ELVS; Reilly et al., 2007, 2010, 2018), offer the best estimates of LT prevalence. In this Australian community sample (*N* = 1741), 19.7% of children were classified as LT at 24 months using the CDI criteria of ‘less than or equal to the 10th percentile’ for expressive vocabulary. Similarly, the UK-based Avon Longitudinal Study of Parents and Children (ALSPAC; Roulstone et al., 2002) found that 18.5% of 1118 children used either single words or babbled only at 25 months. However, as Roulstone et al. (2002) did not use any formal measure of expressive language skills, instead opting for parent-reported utterances of 3–4 words, two words together, single words or babble only, it is unclear how many of these children would be classed as LT children. Rates elsewhere range from 12.6–24.7% (see Table 1). This wide range of prevalence rates, is at odds with the CDI criteria itself – one would expect LT prevalence to be closer to 10.0% if LT children are those that fall at the 10th percentile of the population in expressive vocabulary. This may suggest that the norms used are too high and over-estimate what children can say, thus producing higher numbers of those classed as LT children, or that the sample surveyed in certain studies was not representative of the larger population. Norms also vary by country – for example, children in the UK show lower receptive and expressive vocabulary than American children at the same age (Hamilton et al., 2000).

Variance in prevalence rates may also be attributed to the difficulty in classifying individuals within large epidemiological cohort studies. Rescorla (2011) notes that in small-scale studies, LT children tend to be a well-defined group, but in large-scale studies such as the ELVS, they are more heterogeneous. Subsequently, filtering out those who have concomitant developmental conditions that compound language delay is difficult. For example, in the Zubrick et al. (2007) study, children who were classed as language delayed were also more likely to be delayed in motor and social skills compared to those without language delay, suggesting their sample may have included those with developmental disorders, such as autism spectrum disorder (ASD). Nonetheless, the most conservative prevalence rate (9.6%) comprises a significant proportion of 2-year-olds in the population, making the outcomes of LT children of significant concern.

Table 1.

Variability across prevalence rates in large population-based studies of language ability in children aged 18–30 months old.

Study	*Proportion of study population classed as LT (%)*	*Total population sample*	*Country*	*Criteria used*	*Age at classification (months)*
(E. S. Armstrong et al., 2007)	24.7	689	USA	≤10th percentile of CDI	24
Collisson et al. (2016)	12.6	1023	Canada	≤10th percentile CDI	24–30
(Dale et al., 2003)	9.6%	802	UK	≤15th percentile CDI	24
Henrichs et al. (2011)	14.8	3759	Netherlands	≤10th percentile of CDI	18
Horwitz et al. (2003)	13.5	1189	USA	≤10th percentile of CDI	18–23
(Rescorla et al., 1993)	13.0–15.0	200	USA	<50 words using LDS	24
Zubrick et al. (2007)	13.0	1766	Australia	ASQ Communication	24

ASQ = Ages-and-Stages Questionnaire; CDI = Communicative Development Inventories; LDS = Language Development Survey; LT = late talking

4.3 Outcomes

Many longitudinal outcome studies suggest that LT children score within the average range for their age group by the age of 5 – 7 years (Rescorla, 2011). Despite this, as a group, LT children consistently score lower than TD children on many language measures, even after recovery. Rescorla (2002, 2005, 2009; Rescorla et al., 1997; Rescorla et al., 2000) conducted a longitudinal study following 34 LT children from 2–19 years old, comparing their progress to a TD control group. Between the ages of 2–4 years, approximately 50% of the late talking sample with persistent delays reached typical expressive language skills using mean length of utterance. Between the ages of 5–9 years, LT children were scoring within normal ranges for their age groups on reading, vocabulary, and phonological assessments, but their group averages for these assessments were significantly lower than the control group. This difference persisted at the age of 13, and by the age of 17, LT children (*n* = 26 of original group) again performed within the average range on language and reading tasks, but significantly worse than TDs on vocabulary and grammar tasks (Cohen's *d* = 0.92) and verbal memory (Cohen's *d* = 0.93), but not on reading and writing tasks. Other studies have found similar results – despite scoring within the normal range for their age group, LT children show reduced performance compared to TD children in a number of language tasks including non-repetition tasks (Thal et al., 2005), general language ability, syntax, morphosyntax, speech production (M. L. Rice et al., 2008) and sentence complexity during conversation (Domsch et al., 2012).

Further complicating the matter are LT children who do not recover, and experience worsening deficits. The prevalence of language impairment without sensory, neurological or intelligence-quotient impairment in the population is roughly 7% (Leonard, 2014). These individuals are more recently referred to as having DLD, although DLD itself now has a broader criteria of persistent language delay without biomedical cause and without reference to intelligence-quotient (Bishop, 2017).

The discrepancy between outcome studies suggesting that the majority of LT children recover, and those that suggest otherwise, is highlighted when considering children who have their language impairment identified later in life. Leonard (2014) notes that if these children are originally LT children, then outcome studies should reveal a much higher proportion of LT children with poorer outcomes. He argues this discrepancy is a result of studies using small homogenous samples that are not representative of larger populations and that filter out the most impaired children who meet DLD criteria. Two major population-based cohorts lend credence to this reasoning. Dale et al. (2003) examined 8,386 twins at 24-months-old and compared LT with TD children at 4-years-old. By age 4, 40.2% of the LT group met the criteria for persistent language difficulties. In the ELVS sample (Reilly et al., 2010), 19.7% of the original sample (*n* = 1741) were LT children; by age 4 years, 17.2% of all children remaining in the study (*n* = 1596) met the criteria for SLI and 20.6% had low language status on at least one expressive and receptive composite score (Clinical Evaluation of Language Fundamentals subscales, CELF; Wiig et al., 2006). The relatively high percentages of LT children who have persistent language delays in these studies appears to be at odds with small-scale outcome studies that suggest LT children reach the normal range for their age group. Of note, however, is that the twin study and the ELVS followed children up to the age of 4 years – whereas Rescorla (2002, 2005, 2009; Rescorla et al., 1997; Rescorla et al., 2000) previously mentioned found that LT children reach the average range of language abilities from the age of 5 years and above. This means it is possible that the children in Dale et al. (2003) and Reilly et al. (2010) were not yet at a point where recovery to typical age ranges of language measures was visible.

However, in a smaller study (*n* = 44), Armstrong et al. (2017) found that 49% of their sample who were LT children at age 2 years had persistent language delays at 10-years-old. As with prevalence rates, the difference in populations studied in small-scale and large-scale outcome studies likely accounts for some of this discrepancy – smaller scale outcome studies of LT children may not capture the sheer heterogeneity found in the wider population.

What is clear from these outcome studies is that a proportion of LT children do not recover, and it is extremely difficult to predict who these individuals will be. Rescorla (2009) in particular advocates for a spectrum of early language delay, where LT children are further divided into those with a less severe delay who largely improve (sometimes called 'late bloomers'), and those who have persistent delays. Armstrong et al. (2007) divided a sample of 689 American children into three groups: 1) late talkers with persistent delays in expressive language from ages 2–4 years (using the CDI and RDLS); 2) so-called late bloomers with initial language delay at 24 months old that resolved to reach the average range by the age of 5, and; 3) TD children. LT children performed within the average range on vocabulary and verbal memory tasks (picture vocabulary, letter-word identification, memory for sentences) but were the worst of the three groups, with gaps between the three groups persisting to the age of 10–11 years.

A dimensional approach thus appears sensible, but to be meaningful for intervention, any differences among these groups that are visible from the onset of LT status early on must be identified. A number of outcome studies have looked at predictors for language delay in the hopes that consistent risk factors in LT children who do not recover might be identifiable.

4.4 Predictors of outcomes

Potential predictors of language delay are diverse, and include male sex (Collisson et al., 2016; Hammer et al., 2017; Henrichs et al., 2011; Reilly et al., 2007, 2010; Schjølberg et al., 2011), family history of language delay (Collisson et al., 2016; Dale et al., 2003; Lyytinen et al., 2005; Reilly et al., 2007, 2010), low SES (Fisher, 2017; Hammer et al., 2017; Hartas, 2011; Rescorla et al., 2007), and maternal health factors and birth history (Hammer et al., 2017; Henrichs et al., 2011; Schjølberg et al., 2011), as well as more specific language-related measures, such as low expressive or receptive vocabulary at 18 months old (Armstrong et al., 2017; Fisher, 2017). These risk factors are often found to have significant but small effects in predicting LT at or before the age of 24-months-old, as well as in predicting outcomes later on. Studies that have examined later outcomes of language delay are reviewed below, with study characteristics given in the table below for ease of reference (Table 2).

Table 2.

Characteristics of language outcome predictor studies.

Study authors	*Proportion of study population classed as LT at intake (%)*	*Total population sample (N)*	*Language measures used and age of administration*
Armstrong et al. (2017)	11.5%	783	2-yo: LDS 10-yo: CELF-3
Bishop et al., (2003); Dale et al. (2003)	9.6%	8,386 twins	2-, 3-, 4-yo: short-form CDI
Henrichs et al. (2011)	14.9%	3,759	18-mo: short-form CDI 30-mo: LDS
Horwitz et al. (2003)	13.5%	870	CDI administered once between 18–39 months old
Lyytinen et al., (2001, 2005)	17.0%	200	2-yo: CDI and BSID expressive score 2.5-yo: RDLS 3.5-yo: BNT and PPVT-R
Reilly et al. (2007, 2010, 2018)	19.7%	1,741	8-, 12-, 24-mos: CDI 4-yo: CELF-PS

ASQ = Ages-and-Stages Questionnaire; BNP = Boston Naming Test; BSID = Bayleys Scales of Infant Development; *CELF = Clinical Evaluation of Language Fundamentals (PS = Preschool; 3 = Third Edition); CDI = MacArthur-Bates Communicative Development Inventories; LDS = Language Development Survey; LT = late talking; mo = months-old; NJ = New Jersey; NY = New York; PPVT-R = Peabody Picture Vocabulary Test-Revised; yo = years-old*

Male sex

Reilly et al. (2010) identified male sex as a significant predictor of low language status (>1.25 SD below the mean on the CELF) at 4-years (expressive delay, *OR* = 1.90; receptive delay *OR* = 2.29) and for SLI (now DLD; expressive delay *OR* = 1.43, receptive delay *OR* =: 2.20) when combined with other risk factors including birth history, SES, and maternal vocabulary. Similarly, Hammer et al. (2017) found male sex a significant predictor of low receptive vocabulary at 48-months-old (OR = 1.36) when combined with similar risk factors.

Family history of language delay

Lyytinen found that LT children with a family history of dyslexia have significantly lower receptive and expressive language at age 3;6 years (Lyytinen et al., 2001) and at age 5.5 years (Lyytinen et al., 2005) compared to LT children without a familial risk for dyslexia. However, these studies had small subgroups of LT children with and without familial risk (*n* = 10-12), making any conclusions about heritability difficult to generalise to a larger population. Larger studies have found family history to be a small but significant risk factor for later language delay (Reilly et al., 2007, 2010; Zubrick et al., 2007).

Bishop et al. (2003) found the group heritability of language delay at 24-months-old in a large twin study was significant but small regardless of outcome (h^2_g=0.240; where h^2_g is an index of the extent to which the mean difference between groups is due to genetic factors). In particular, heritability was found to be significantly higher in LT children with persistent delay when parental concern at 3 years (h^2_g=0.41) or professional involvement at 4 years (h^2_g=0.41) was used as an outcome, as opposed to language-based outcomes such as parent-reported expressive vocabulary. However, Fernald and Marchman (2012) suggest a drawback of twin studies is that they tend to focus on families with high SES with more resources and support, and that SES might well be a moderating factor in the heritability of language delay.

Socioeconomic status

Low SES has been found to have a negative effect on language outcome more generally (Fernald et al., 2013; Golinkoff et al., 2019; Hartas, 2011). Reilly et al. (2010) also found low SES predicted worse language outcomes at age 4-years in addition to LT status at 2-years-old, but this combined with other predictors (maternal health, materal vocabulary, family history of language delay, male gender, child birth history) only moderately discriminated between children with and without adverse language outcomes. Hammer et al. (2017) identified LT status at 24-months-old as a significant predictor of low vocabulary at 48-months-old (*OR* = 2.92), with additional effects of low SES (*OR* = 3.14). When adding in several factors that overlap with SES, such as childcare availability and maternal health, this partially explained the effect of SES as a variable in the model, indicating that factors overlapping with SES should also be accounted for. Other studies however have found low SES to be a non-significant predictor for low language outcomes (Clegg et al., 2015; Horwitz et al., 2003).

Other factors

Horwitz et al. (2003) found that bilingualism (*OR* = 2.78), high parental worry about language (*OR* = 5.13), low family expressiveness (direct expression of feelings, *OR* = 1.95) and low social competence (*OR* = 5.13) all increased the risk of language delay. Bishop et al. (2003) also found parental concern to be a significantly positive predictor of language outcomes, but deemed it insufficient as a risk factor alone, as it cannot be used to ascertain severity of language delay. Additionally, it is worth noting that bilingualism can lead to perceived, rather than actual, language delay if only vocabulary for one language is assessed (Hoff et al., 2012) and determining directionality between other factors found to be significant is problematic, making their use as predictors somewhat limited.

In a sample of 90 LT children, Armstrong et al. (2017) found that maternal smoking in pregnancy increased the risk threefold of persistent language delay in LT children versus LT children who recovered (*OR* = 3.34). Interestingly, of the wider sample who had no language delay at age 2, 26% (182/693) subsequently were found to have low language skills on the CELF at age 10 – suggesting their language problems were not adequately detected at age 2. In this latter group, low SES, lower parental education level, and male gender were significant risk factors for lower CELF-3 scores at age 10.

Can a model of risk factors be used to determine the need for intervention?

Some clinicians have called for a model of risk factors to help identify LT children at risk of further language delay where the more risk factors an individual has, the lower the threshold for intervention (Collisson et al., 2016). Determining an appropriate cut-off at which the number of risk factors is deemed sufficiently high will be the next stage to this model, and has already proved highly difficult, as the effects of risk factors are small (Collisson et al., 2016; Reilly et al., 2010), and only explain a small proportion of variance in language outcomes. Subsequently, any model of risk factors based purely on epidemiological data to determine intervention may not be precise enough to be useful to clinicians.

In a large Netherlands-based population cohort, Henrichs et al. (2011) split their sample using expressive vocabulary outcomes into a reference group without delay, 'late bloomers' (LT children at 18 months who have normal vocabulary at 30 months), late onset delay (TD children with normal vocabulary at 18 months but delayed at 30 months), and persistent delay. Receptive vocabulary delay at 18 months old conferred the highest risk of language delay – these children were 4 times more likely to be a late bloomer (*OR* = 4.25), almost 4 times as likely to have late onset delay (*OR* = 3.92) and 9 times more likely to have persistent delay (*OR* = 9.09). Higher maternal age increased the risk of being a late bloomer (*OR* = 1.05). Children from homes with low maternal education were two times more likely to have a persistent delay (*OR* = 2.13). Other studies have also found receptive vocabulary to be helpful in distinguishing between LT groups and their outcomes (Lyytinen et al., 2005). Overall, in Henrichs et al.'s study, expressive vocabulary at 18 months was the strongest predictor of expressive vocabulary at 30 months, explaining 11.0% of the variance, with receptive vocabulary at 18 months explaining an additional 0.5%. Additional demographic (gender, age, ethnicity), maternal (parenting stress, educational, age, income) and perinatal (birth weight, gestational age, prematurity) factors explained only 6.2% of the variance at both 18 and 30 months. Another large population study that looked at child factors such as birth order, gender, and birthweight, alongside maternal emotional health and maternal education found these factors only explained 4.1 – 6.3% of the variance in a model predicting language outcome at 18 months (Schjølberg et al., 2011).

Similar results were found by the ELVS (Reilly et al., 2007, 2010). The ELVS examined at potential risk factors for language delay, including gender, SES, perinatal factors (such as maternal parity, prematurity and birth weight), non-native English-speaking households, and family history of language delay in 1720 infants recruited at 8 months of age and followed up at 12- and 24-months-old, and later at 4 years. However, these risk factors at 8-months-old and 12-months-old explained only 7% of the variance at 24 months, consistent with findings elsewhere (Henrichs et al., 2011; Zubrick et al., 2007). By 4 years old, baseline predictors accounted for 18.9% and 20.9% of the variance in receptive and expressive language performance respectively. The addition of LT status at 2 years as a predictor increased this to 23.6% and 30.4% respectively, but the authors still concluded that this was insufficient to predict language delay in children.

Fisher (2017) conducted a meta-analysis of predictors of expressive vocabulary in LT children aged 18-28 months to establish an overall effect size across studies. All studies were prospective and had at least 5 months' worth of follow-up after LT status was designated. A total of 20 studies were identified (*n* = 2134) with a range of follow-up from 5 to 28 months. Expressive vocabulary size (most often assessed by parent-report) receptive language, phrase speech, SES, gender, and family history were examined as predictors. Of these, only expressive vocabulary size, receptive language and SES yielded significant effect sizes. Receptive language yielded the largest effect size (Pearson's $r = 0.34$), whereas expressive vocabulary (Pearson's $r = 0.25$) and SES (Pearson's $r = 0.11$) yielded smaller effect sizes.

The studies discussed above indicate that risk factors associated with parents, children, and low expressive and receptive vocabulary contribute as risk factors to language outcomes, but not to a sufficient degree to enable accurate prediction of these outcomes, nor distinguish between LT children with persistent delay and those who recover. A systematic review by Law et al., (2000) designed to evaluate whether the UK needed universal screening for speech and language delay concluded that there was insufficient evidence to do so. A total of 21 prevalence studies published between 1967–1997 were reviewed. The results suggested that the progression of speech and language disorders, including late talking, was too heterogeneous to clearly identify risk factors that could be used to identify those suitable for intervention. A more recent major US taskforce (H. D. Nelson et al., 2006; Siu, 2015) also concluded that the evidence for routine screening of children for speech and language delay was insufficient due to the lack of reliable risk factors as predictors of language outcomes.

While it is clear that late talking affects a significant number of children and carries a risk of further language delay; what is less clear is what to do with that information. Without distinguishing between those who will struggle from those who will not, it remains difficult to identify what may cause continued language difficulties in LT children. Subsequently, further research is required to elucidate whether LT children are qualitatively, rather than just quantitatively, different, and to examine the benefits of taking a dimensional approach to language delay that accomodates individual differences. This necessitates an examination of how LT children actually learn words, and whether their performance on word learning tasks has any predictive value in determining outcomes.

4.5 Word learning in late talking children

This thesis focuses on three candidate word learning mechanisms: nonword repetition, fast mapping, and cross-situational word learning. To avoid repetition, previous studies that have investigated word learning mechanisms in LT children are reviewed in the following chapter (Chapter 5). However, as previously described, LT children may have a wide range of receptive vocabulary (Fisher, 2017; Henrichs et al., 2011), and are grouped on their expressive output. As a result, understanding the relationship between the stages of word learning and expressive vocabulary in particular is key to identifying how word learning mechanisms might falter in LT children. The links between word learning and expressive vocabulary are thus reviewed here.

Recall that McMurray et al.'s (2012) model assumes a cognitive approach to word learning, where referent selection is an in-the-moment competitive process between visual referents and auditory word-forms through a lexical concept layer (decision-making), and retention is a Hebbian process that strengthens and prunes word-referent associations based upon co-occurrences over time (learning). Applied to LT children and word learning, the research question becomes bidirectional: do LT children apply different constraints to the decision-making process that affects how new words are later retained in the vocabulary, and how do those learnt associations within the vocabulary influence decision-making when faced with novel words?

How do small expressive vocabularies relate to word learning?

A promising explanation for LT concerns the relationship between phonology and vocabulary. Mirak and Rescorla (1998) analysed speech samples from 35 LT children (Reynell Expressive Language score <6 months below chronological age) during free play with toys and during the administration of standardised tests. They found that LT children used fewer consonants than TD children, but that expressive vocabulary did not predict

Mean Length of Utterance (MLU) or syntactic complexity at age 3 years. They concluded that LT children have delayed, rather than different, phonetic abilities at the time of LT.

Studies that make use of the phonological elements of words such as phonotactic probability (PP) and neighbourhood density (ND) in LT children offer valuable insight into how concurrent phonological delays might impact word acquisition. PP refers to how likely it is that a sequence of phonemes will occur, with high PP being a high likelihood of co-occurrence (e.g. *mp* in 'bump') and low PP being a low likelihood (e.g. *mt* in 'dreamt'), whereas ND is determined by the number of words generated by deleting, adding, or substituting a single sound within a given word (Luce & Pisoni, 1998; Vitevitch & Luce, 1999). Words with high ND have many phonemic neighbours, e.g. 'sit' has 36 (including 'hit', 'lit') and words with low ND have few neighbours, e.g. 'these'. In TD, children learn words with high PP more rapidly (Storkel, 2001) and are also more likely to have correct phoneme awareness for words with high ND (Hogan et al., 2011).

Using PP and ND, Edwards et al. (2004) proposed that children with smaller vocabularies have less support to rely upon for phonological representations of words. They tested production and fluency in nonword repetition with low- and high PP words, and also measured vocabulary in children aged 3 – 8-years-old. Higher accuracy for low-PP words was significantly correlated with higher expressive vocabulary (R^2 = .30), and higher accuracy for high-PP words significantly correlated with expressive vocabulary, albeit to a lesser degree (R^2 = .21). In particular, they showed that the effect of PP was largest in children aged 3–4-years-old who had less developed lexicons. They suggested that when exposed to a novel word with low PP, there was less access to similar words in the lexicon that could support the phonological representations involved in producing the new word. Crucially, this process was described as dynamic: more experience in phonological patterns leads to more generalising over lexical knowledge, and vice versa. Reduced expressive vocabulary in LT children would thus feed into reduced ability to perceive and then produce novel words.

Stokes (2014) took a slightly different approach, examining the structure of LT children's expressive vocabularies. They proposed that LT children struggle to activate the correct word form for words that are low in ND. When investigating children's expressive and receptive vocabulary at 1;6- and 2;0-years-old, Stokes identified that smaller lexicons had higher ND values than large lexicons for expressive vocabulary, but not receptive vocabulary. Under this proposal, low ND words have weaker lexical representations, and thus a low ND word can be activated for receptive processing, but not expressive. In contrast, those that have high ND have stronger lexical representations, as they are heard and used more frequently, and thus more easily accessed when requested to produce them, reducing demands on working memory. LT children may then have more fragile phonological representations that limit word production, equating to higher ND values of words that LT children understand and produce, whereas TD children are able to produce low and high ND words.

How does phonology relate to vocabulary and learning new words during fast mapping in LT children?

Studies that examine PP and ND of stimulus words learnt during fast mapping tasks demonstrate that LT children may be extracting phonological information in novel words differently to TD children. These test referent selection, and also word production, immediately following exposure to word-referent stimuli. For example, Weismer et al. (2013) found that LT toddlers (30-month-olds ≤15th percentile on CDI; *n* = 30) showed no advantage for low PP and ND when learning new words, whereas TD toddlers did. Novel word production in the fast mapping task also correlated with expressive vocabulary at 30 months of age in LT children (Spearman's Rank ρ=0.44) and, at 5;6 years , their novel word comprehension correlated with receptive language skills (Spearman's rank ρ=0.43). In TD children however, receptive vocabulary correlated with novel word production (Spearman's rank ρ=0.39). These findings suggest that LT children rely on inter-domain processes during fast mapping, whereas TD children are able to make cross-domain links.

MacRoy Higgins et al. (2013) also found that 24-month-old LT children (<15th percentile on CDI; *n* = 12) exhibited no difference in performance for high or low PP/ND during fast mapping, but in contrast to Weismer et al. (2013), TD children showed better production and more sensitivity to errors for high over low PP/ND words. They did not report concurrent relationships with receptive and expressive vocabulary. However, the latter study used a preferential looking task at 24 months of age that deliberately probed responses to high PP/ND and low PP/ND sequences comprising 12 words (six low PP/ND words, six high PP/ND), whereas Weismer et al.'s (2013) study focused on fast mapping with post-hoc tests of PP/ND consisting of only two words (one low PP/ND, one high PP/ND). Regardless, both studies suggest that LT children may be making use of PP and ND differently to TD children.

Overall, these studies are consistent with the theory that LT children may have less robust phonological representations, secondary to having smaller vocabularies and less practice in using them for production. This is also consistent with being less able to accurately produce novel words immediately after hearing them in nonword repetition tasks. However, as only a handful of studies have examined fast mapping in LT children, how LT children perform in fast mapping tasks and links with their nonword repetition performance requires further examination. Furthermore, how these mechanisms relate to LT children's expressive vocabulary over time has yet to be fully established.

How do phonology and fast mapping relate to retention and longer term learning in LT children?

Edwards et al. (2004) and Stokes (2010, 2014) propose that smaller concurrent expressive vocabularies limit the amount of abstraction that can take place for the phonological make-up of novel words based on existing words in the vocabulary. The results of Weismer et al. (2013) and MacRoy Higgins et al. (2013) suggest that during fast mapping referent selection tasks, LT children perform less accurately than TD children as a result of extracting different information from high and low PP words. However, this only covers referent selection, and does not address mechanisms that are indicative of longer term learning, such as retention.

LT children in MacRoy Higgins et al. (2013) required more exposures to high PP words than TD children in order to achieve the same level of accuracy when tested immediately after training, raising the possibility that LT children may require a higher number of exposures of words in order to make up for deficits in phonological representation. Alternatively, M. L. Rice et al. (1994) proposed that in DLD, children rely on frequency of input for a longer period of time than TD children. When producing their first words, high frequency of input appears to be particularly important for children; however, once a sufficient lexicon has been accumulated, children rely on this input much less to produce words (Hart, 1991). Continued reliance on input as a result of having smaller lexicons may also explain why repeated exposures to words may be more effective for LT children.

Stokes (2010; Stokes et al., 2012) also suggests that LT children are relying on input, rather than on existing vocabulary, again in relation to phonological input. LT children may rely on statistical learning mechanisms at first to extract relevant auditory information from the input, but whereas TD children broaden attunement to statistical regularities of less common words (those with low ND), LT children fail to do the same, resulting in reduced vocabulary expansion. As a result, Stokes (2014) also suggests that LT children may benefit from greater repetition of target words that have high ND to aid longer term learning.

Repetition of information is inherent in CSWL and in longer term learning of words (McMurray et al., 2012). In order to learn a word beyond producing a novel label, or selecting the correct referent for a novel label, children must extract information around the co-occurrences of words and referents over time to store word-referent mappings for later use. However, word learning studies in LT children tend to test immediate comprehension and production of novel words, rather than retention after a delay, and no studies to date have tested CSWL in LT children.

Thus, identifying whether or not LT children are able to retain word-referent mappings following referent selection may help determine whether they have difficulties encoding novel words sufficiently for longer term learning following single exposures. Furthermore, testing LT children on referent selection and retention performance on CSWL tasks would not only reveal their ability extract statistical information around word-referent mapping, but also test their ability to process this information sufficiently for longer term learning.

Word learning and outcomes

Studies that examine differences in word learning abilities may also help to predict later vocabulary outcomes. For example, Weismer (2007) found novel word comprehension and production on a fast mapping task in LT children at 30-months-old (n = 30, <10th percentile CDI) predicted their MLU 12 months later. Fernald and Marchman (2012) tested LT (n = 32, <20th percentile CDI) and TD children's performance on a word learning task that required the recognition of familiar word-referent pairs. Children's processing efficiency at 18 months predicted individual vocabulary growth over the following year. Task performance accounted for 4–17% variance in addition to LT status in predicting vocabulary growth trajectories and allowed better prediction of those with persistent delay at 30 months (39% of original LT group) – LT children who were quicker and more accurate on the task had steeper, faster growth trajectories than those who performed slower and less accurately. However, studies that use word learning mechanisms to predict outcomes are far less plentiful than studies that investigate epidemiological risk factors in late talking. Subsequently further research is necessary to identify the predictive value of word learning in LT children.

Summary

Understanding how LT children learn words may, firstly, highlight differences in comparison to TD children and, secondly, help identify mechanisms that predict later language outcomes. This in turn may help predict which LT children will develop more severe and persistent language deficits in the future. In order to test this framework. studies that examine word learning in LT children over time must relate word learning performance to both early and concurrent expressive vocabulary *and* include measures of short- and longer-term word learning mechanisms. This includes phonological perception and articulation following single exposures as in nonword repetition, the mapping of referents to words and subsequent retention following single exposures in fast mapping, and finally, mechanisms such as cross-situational word learning that make use of repeated exposures over time to produce longer term learning. If LT children are impaired at each step of this outlined process, then this may suggest that they represent a discrete group of children with generalised word learning difficulties. If, however, they are impaired on some, but not all, mechanisms, then this would indicate a narrower range of difficulties that may be targeted for both identification of outcomes and intervention.

5 Chapter 5: The mechanisms of word learning in early development: A longitudinal study of late talking and typically developing children

5.1 Chapter introduction

Studying LT children over time offers the chance to identify, firstly, whether or not LT children learn words differently to TD children, and secondly, how later language outcomes differ as a result of these mechanisms. Previous research suggests that LT children may have delayed phonological abilities that influence how they build, and rely on, their expressive vocabularies during word learning (Edwards et al., 2004; Mirak & Rescorla, 1998; Stokes, 2010, 2014; Stokes et al., 2012). However, any impairment in phonological ability must be related not only to processes that test the immediate mapping of words to referents, such as in fast mapping tasks, but also to processes that underlie longer term learning, such as the extraction of statistical, associative information.

This Chapter reports a longitudinal study of LT and TD children followed over 18 months. Children were seen at 2;0 – 2;5-years-old (T1), 3;0 – 3;5-years-old (T2), and 3;6 – 3;11-years-old (T3). T3 data collection was interrupted by the COVID-19 pandemic; subsequently, a further measure was administered remotely at 4;0 – 4;5-years-old. Consistent with the majority of the literature (Fisher, 2017), LT children were identified using the Oxford CDI (Hamilton et al., 2000) with criterion of <10th percentile on expressive vocabulary at 2;0 – 2;5-years-old. Three tasks that examine different aspects of word learning (word repetition, fast mapping, and cross-situational word learning) were then tested at 3;0 – 3;5-years-old, and at 3;6 – 3;11-years-old, and related to both early and later expressive vocabulary.

Author contribution for Chapter 5: *Rachael W Cheung:* design, data collection, analysis, writing, review. *Calum Hartley:* design, review. *Padraic Monaghan:* design, review

5.2 Abstract

Late talking (LT) children are a heterogenous group characterised by developmentally delayed expressive language at 2;0-years-old in the absence of any other delay. Variability in word learning mechanisms in LT children may contribute to linguistic abilities and explain why some recover, whilst others do not. In a longitudinal study from age 2;0 – 3;11 years, we tested a cohort of TD (*n* = 40) and LT (*n* = 21) children across a series of tasks designed to isolate different mechanisms involved in word learning: encoding and producing spoken forms of words (using a nonword repetition task), identifying referents for words (using a fast mapping task), and learning associations between words and referents (using a cross-situational word learning task).

We found that LT children had lower accuracy on nonword repetition than TD children, despite most reaching TD ranges for expressive vocabulary. We found no between-group differences in fast mapping and retention accuracy, although both were predicted by concurrent expressive vocabulary. LT children performed less accurately than TD children on cross-situational word learning retention trials, despite showing no between-group differences during referent selection training trials.

These results indicate that LT children continue to have deficits in phonological representation that impact their word learning ability and expressive language abilities, but do not show difficulties in fast mapping novel words. They also raise the possibility that LT children struggle to retain associative information about word-referent mappings. LT children may thus use some, but not all, word learning mechanisms differently than TD children.

5.4 Introduction

Late talking (LT) children fall at or below the 10th percentile for expressive vocabulary compared to typically developing (TD) children at around 2-years-old, despite the absence of concurrent developmental delays or sensory disorders (Desmarais et al., 2008; Fisher, 2017). Although the majority catch up to their TD peers by school age (Rescorla, 2011), LT children are at increased risk of Developmental Language Disorder (DLD; Leonard, 2014; Reilly et al., 2010). However, there are few consistent factors across LT children that enable practitioners to reliably predict who is at risk of DLD. Expressive vocabulary alone is not a clinically useful predictor of language delay (Law & Roy, 2008; Leonard, 2009), and demographic predictors such as socioeconomic status, family history, and male gender explain only a small amount of variance in outcomes (Dale et al., 2003; Fisher, 2017; Hammer et al., 2017; Hartas, 2011; Henrichs et al., 2011; Lyytinen et al., 2005; Reilly et al., 2010; Rescorla et al., 2007). Furthermore, although most LT children recover, they appear to score at the lower end of the TD range across language measures once they reach school (Rescorla, 2009; Rescorla, 2002, 2005; Rescorla et al., 2000). Employing a wait-and-see approach may thus risk missing out on a key early intervention period (Singleton, 2018; Collisson et al., 2016)

In order to understand how and why outcomes in LT children differ, we must consider whether or not they learn words in a qualitatively different way to TD children. When learning a word for the first time, children must perceive the phonological elements that make up that word and articulate them (as measured by nonword repetition tasks). They must also map the word to its correct referents (referent selection, as measured by in-the-moment processes during fast mapping tasks), and then develop and retain the word-referent association (retention, as measured in cross-situational word learning, CSWL, tasks). These processes enable children to be able to both understand (receptive vocabulary) and produce (expressive vocabulary) words later on. As children build a lexicon, their existing knowledge of words may also impact on how they learn novel ones (Edwards et al., 2004; Stokes, 2010, 2014). Thus, any examination of word learning tasks in LT children requires relating performance on tasks to both early and later vocabulary.

We examined LT and TD children's performances on tasks that probe these different mechanisms involved in language learning in a longitudinal study. This allowed us to determine which aspects of processing are impacted in LT children as their language skills develop over time. We next summarise studies of nonword repetition, fast mapping, and CSWL in LT and TD children.

Nonword repetition tasks in LT children

LT children have shown deficits in nonword repetition at the time of identification. Nonword repetition tasks require children to repeat a list of novel words immediately after a speaker produces them (Coady & Evans, 2008). Stokes and Klee (2009) found that children at or below the 16th percentile on the expressive CDI at 2 – 2;6-years-old could be identified based on their nonword repetition. Marini et al. (2017) also reported that LT children aged ~2;6-years-old had impaired nonword repetition performance compared to TD children. These studies indicate that LT children are characterised by concurrent delays in the immediate perception, storage, and articulation of novel words.

However, if children with a history of LT continue to show reduced accuracy in nonword repetition after reaching typical vocabulary, this would suggest that early expressive delay may also have an enduring impact on children's ability to encode the phonology of novel words, even after they have reached typical levels of vocabulary. For example, Marini et al. (2017) found that nonword repetition at ~2;6-years-old correlated with articulation ($r = .47$), naming ($r = .33$), semantic fluency ($r = .34$) and lexical comprehension ($r = .27$) approximately 11 months later. Studies that test nonword repetition at older ages have found that LT children identified at 2-years-old are also impaired at ~ 2;6- and ~3-years-old (Rujas et al., 2017), as well as at 4 – 6-years-old (D'odorico et al., 2007). Broader assessments suggest that children with a history of LT still struggle to produce speech when tested on their articulation and phonological abilities at 4 – 5-years-old, showing reduced accuracy and more errors on standardised speech assessments than TD children (Neam et al., 2020). However, none of these studies tested concurrent expressive vocabulary, meaning it is impossible to know whether LT children had reached the TD range. Conversely, others have found no differences on nonword repetition tasks between TD and recovered LT children at 3-years-old (MacRoy-Higgins & Dalton, 2015) and 5-years-old (Petruccelli et al., 2012).

Thus, whilst LT children may have impaired nonword repetition, whether they continue to have difficulties once they have recovered remains less certain. Edwards et al. (2004) proposed that a smaller expressive vocabulary leads to more fragile phonological representations, and a limited ability to abstract over existing knowledge to support novel word encoding and articulation. Despite this, in a review of the literature, Coady and Evans (2008) reported that only receptive vocabulary correlates with nonword repetition. Conversely, studies by Munson et al. (2005) and Chiat and Roy (2007) involving speech and language therapy clinic samples reported that both expressive and receptive vocabulary correlate with nonword repetition task performance. Thus, further research is necessary to determine how expressive vocabulary correlates with nonword repetition over time, and also how nonword repetition ability may tie into other mechanisms of word learning, such as referent selection and retention.

Fast mapping in LT children

Fast mapping refers to the early process of word learning where children encounter a novel word and its referring object for the first time, and are able to disambiguate the novel word correctly (Carey & Bartlett, 1978). Fast mapping tasks assess the ability of children to comprehend or produce these novel words immediately after single exposures, but unlike nonword repetition, also require accurate referent selection (selecting the correct object that matches the word).

Proposed strategies for fast mapping include mutual exclusivity, which assumes that each object has only one label (Markman & Wachtel, 1988). Thus, when faced with two objects – one familiar with a known label and one unfamiliar – children infer that a novel label must refer to the unfamiliar object. TD children are able to use this principle to constrain their referent selection for novel words (Markman et al., 2003) and respond accurately on fast-mapping tasks at around 2-years-old (Bion et al., 2013). However, it is not yet clear whether LT children apply the same strategies as TD children when fast mapping unfamiliar words.

Referent selection can be conceptualised as a competitive process between potential word-referent pairs, where fast mapping is driven by cognitive constraints like mutual exclusivity, rather than by existing associations between known words and lexical concepts (Halberda, 2006; McMurray et al., 2012). Based on their nonword repetition task performance, we would expect LT children to be less accurate at *producing* novel words. However, if LT children are able to perform above-chance and equivalent to TD children when tested purely on *comprehension* of fast mapped words, this would suggest that the initial competitive process involved in referent selection is intact, and that referent selection is not necessarily related to early expressive delay. If, however, LT children show reduced accuracy compared to TD children, this would suggest that early expressive delay may be related to receptive fast mapping abilities during referent selection.

Only a few studies to date have examined fast mapping in LT children, yielding evidence for reduced performance in both comprehension and production of novel words. Weismer et al. (2013) found that LT children aged 2;6-years seemed to score above-chance (25%) at test for comprehension of familiar and novel words. However, in comparison to TD controls, LT children responded less accurately on production of familiar and novel words and on comprehension of novel words, but were equally able to comprehend familiar words. In a similar task, MacRoy-Higgins et al. (2013) also found that LT children aged 24-months-old performed less accurately than TD children on comprehension and production of novel words. Rujas et al. (2019) reported that LT children struggled to fast map and extend novel words (comprehension) when tested at three timepoints (~2;2-years-old, ~2;9-years-old and ~3;4-years-old), compared to TD children. However, Rujas et al. did not measure concurrent vocabulary throughout their study, meaning it is unclear whether their later timepoints included non-recovered and recovered LT children.

These results hint that referent selection may be related to expressive delay, but further investigation is necessary given the scarcity of studies. In addition, fast mapping does not necessarily indicate longer term learning and retention of words – TD children aged 2-years-old who have high accuracy during fast mapping referent selection, show low accuracy when tested on retention of the same words just 5 minutes later (Horst & Samuelson, 2008). Language acquisition is thus thought to result from the interaction between fast mapping processes that identify referent selection during online learning, and slower, longer term learning where word-referent associations are gradually strengthened and pruned over time (McMurray et al., 2012). Thus far, no prior studies of fast mapping in LT children have tested retention after a delay. As TD children show limited ability to retain words from fast mapping even at 4-years-old (Vlach & Sandhofer, 2012), we would also expect to observe limited retention in LT children. However, if LT children show less accurate retention in comparison to TD children, this might indicate that the processes underlying retention after fast mapping are related to early expressive delay.

Cross-situational word learning in LT children

Statistical learning refers to the ability to extract information from the environment and then discern patterns from that information (Romberg & Saffran, 2010). In a typical cross-situational word learning (CSWL) task, learners must use statistical information to correctly pair words and referents from trials that contain ambiguous visual objects and auditory labels, by noting when labels and objects co-occur (Yu & Smith, 2007). Infants (Smith & Yu, 2008) and children (Bunce & Scott, 2017; Vlach & DeBrock, 2019) are able to identify correct word-referent pairs during CSWL tasks. A key feature of CSWL is the repetition of information across trials that leads to accurate referent selection. Over development, the retention of novel mappings through repeated exposure and associative learning may contribute to longer term learning (McMurray et al., 2012).

Some studies have found that LT children require more exposures to learn words. When testing fast mapping comprehension, MacRoy-Higgins and Dalton (2015) found that children with a history of LT benefitted from more exposures to words with high phonotactic probabilities than TD children. Children with DLD also appear to require more exposures than TD children to learn words (Gray, 2004, 2006; Kan & Windsor, 2010; Rice et al., 1994). Thus, if both LT children and those with DLD require repeated exposures during word learning, their novel word learning may be increasingly dependent on repeated statistical information than lexical principles that constrain referent selection.

If LT children rely on statistical information and repeated exposures to word-referent mappings, this might result in performance on par with TD children during CSWL, but not in fast mapping tasks, where they only have one exposure to a novel word. Studies of CSWL in children with autism spectrum disorder (ASD), a population with significant language difficulties (Eigsti et al., 2011), have not found any differences in CSWL when compared with TD children matched on receptive vocabulary (Hartley et al., 2020; Venker, 2019). These findings indicate that there were no qualitative differences between the populations in how they utilised statistical information – rather, their language difficulties stemmed from elsewhere.

In the only study to our knowledge that assesses task-based CSWL in DLD, Ahufinger et al. (2021) found that although both bilingual children with DLD and TD children

(8-years-old) performed above chance at test, the DLD sample scored significantly less accurately than TD children. However, they tested word-referent mappings immediately after training, rather than referent selection during training, and did not test the retention of words after a delay, meaning it is not possible to distinguish between referent selection and retention ability for these children.

No studies to our knowledge have tested task based CSWL in LT children, nor do CSWL studies typically relate vocabulary to CSWL task performance. If CSWL reflects a general cognitive learning mechanism (the ability to extract statistical information and to use process-of-elimination), rather than a language specific mechanism (McMurray et al., 2012; Yu & Smith, 2012), the processes involved in CSWL may also be less dependent on existing vocabulary than nonword repetition performance. However, if LT children are impaired in CSWL as a function of limited ability to extract statistical information, this may help characterise why these children appear to have difficulties adding words to their lexicon over time.

Overall, these three tasks – nonword repetition, fast mapping, and CSWL – reflect separate mechanisms that apply to word learning. However, they may apply differently to children according to their individual language abilities, and to how these change over time. The trajectory of language development is a key part of understanding how LT children may differ to TD children as they develop, particularly as LT children are a heterogenous population (Reilly et al., 2010). For example, Weismer (2007) found combining non-verbal IQ, expressive language, and a test of novel word comprehension correctly identified 90% of LT children identified at 2-years-old (N = 40), who reached the TD range 12 months later, and correctly rejected 91% who did not reach this range. Considering how both early and later vocabulary relates to the distinct mechanisms outlined during word learning is thus important, and may help highlight LT children who continue to have impairments in vocabulary production.

The present study

The extant literature regarding LT children leaves a series of open questions concerning their word learning abilities, and where in the process they may struggle. Firstly, although research suggests that LT children are impaired concurrently on non-word repetition (Marini et al., 2017; Stokes & Klee 2009), the literature reports mixed results on whether these children continue to show impairments once recovered (D'Odorico et al., 2007; Neam et al., 2020). Secondly, studies that examine fast mapping in LT children indicate potential deficits in rapid comprehension and production (Weismer et al., 2013). However, these are few, do not test retention, and do not always report relationships with expressive vocabulary, making it difficult to identify whether LT children continue to struggle once reaching the TD vocabulary range. Thirdly, children with DLD show impairments in CSWL (Ahufinger et al., 2021), which may also be found in LT children, but this has not yet been tested. Finally, despite the heterogeneity of LT children, studies do not always account for the trajectory of vocabulary development over time.

We used a longitudinal design to study a cohort of TD children and LT children recruited at 2 – 2;5-years-old and followed up at 3 – 3;5-years-old and 3;6 – 3;11-years-old. We investigated whether LT children make use of the same strategies as TD children during different stages of the word learning process, examining how both LT status and concurrent expressive vocabulary relate to these stages. Using a repetition task that assesses production of real words as well as nonwords (PSRep Test; Chiat & Roy, 2007), we tested whether LT children show prolonged deficits in their ability to encode and reproduce novel phonological information (nonwords), as well as assessing how intact their phonological representations are for familiar information (real words). Using a fast mapping task that measures comprehension, we tested whether LT children show intact use of mutual exclusivity during referent selection. Using a CSWL task, we tested children's ability to track co-occurrences between words and objects over multiple individually-ambiguous exposures in order to disambiguate correct word-referent associations. We also tested retention following both fast mapping and CSWL, allowing us to identify whether LT children show deficits in the acquisition of novel word-referent pairs after a short delay.

We hypothesised that LT children would demonstrate lower accuracy across all tasks in comparison to TD children at all time points. We also hypothesised that higher expressive vocabulary would correlate with more accurate performance across all tasks.[4] By relating past and present vocabulary to tasks that test different stages of word learning, we highlight which processes relating to word learning in LT children may be atypical, and how the trajectory of children's expressive language development may also be affected by these mechanisms.

5.5 Method

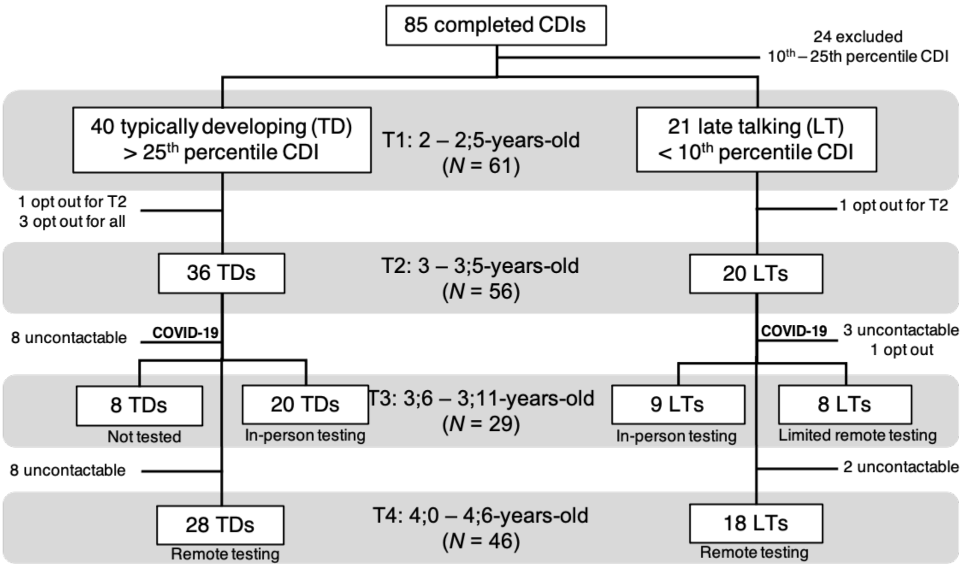
Participants

Participants were recruited as part of a longitudinal study that followed-up LT and TD children between the ages of 2 – 2;5-years-old to 3;6 – 3;11-years-old. Participants were recruited using flyers from Lancaster Babylab, via health visitors, and through nurseries in the Lancashire area. Once consent to contact was obtained, parents completed the Oxford-CDI (Hamilton et al., 2000) and were included if they met one of the following criteria for the two groups: TD with an expressive vocabulary score ≥ 25th percentile, or LT with an expressive vocabulary score ≤ 10th 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 English, with no history of developmental or sensory delays or disorders.

[4] 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 all but two LT children recovered at T2 and 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 and T3.

A total of 85 families completed the CDI; of these, 24 children were excluded due to the aforementioned criteria. A total of 61 children (40 TD and 21 LT) comprised the final cohort. Visits occurred at 2 – 2;5-years-old -years-old (baseline T1), 12 months from baseline at 3 – 3;5-years-old (T2), and 18 months from baseline at 3;6 – 3;11-years-old (T3). As a result of the COVID-19 pandemic, data collection during the third timepoint was interrupted. For the remaining cohort that had not been tested, the remaining LT children (8) were tested online on only expressive and receptive vocabulary questionnaires. An additional timepoint (T4) at 4 – 4;6-years-old was added that could be administered remotely to gain extra information about the cohort. The progression of the study and sample sizes can be seen in Figure 1.

Figure 1. Study diagram showing the progression of longitudinal study and sample sizes across timepoints.



Questionnaires

The *Oxford-CDI* (Hamilton et al., 2000) was used at T1 to confirm participants' allocation to either the LT or TD group. This questionnaire asks parents to indicate all of the words that their child says and understands (i.e. estimating their total expressive and receptive vocabulary sizes).

The *Expressive and Receptive One Word Picture Vocabulary 4th Edition* (EOWPVT-4/ROWPVT-4; (Martin & Brownell, 2011) were used as measures of expressive and receptive vocabulary at T2 and T3, administered by the experimenter. For the EOWPVT-4, children are shown a picture of an object and asked to name it, and for the ROWPVT-4, children are shown four pictures at a time and asked to point to the picture that shows the specific word asked for.

The *Leiter-3* non-verbal Cognitive Battery (Figure Ground, Form Completion, Classification Analogies, Sequential Order; Roid et al., 2013) was used as a measure of non-verbal IQ at T3.

The *Vineland-3 Domain General Parent-Report* questionnaire (Sparrow et al., 2016) was used at T4 as a measure of general functioning.

In addition, at all timepoints, general information concerning access to speech and language therapy, sensory or developmental diagnoses, parental concern surrounding speech and language skills, and parental socioeconomic status was also recorded.

Testing session set-up

The tasks and questionnaires that were administered at each timepoint can be viewed in Table 1. We utilised a mobile testing set-up to maximise retention of participants in the study, with data collection occurring within a room at Lancaster Babylab or at the participant's home. Where testing took place in the home setting, care was taken to ensure a quiet space and clear environment, with just the child and main caregiver present. During testing, the child was seated on one side of a 1 metre fold-out table on a small chair with the caregiver sitting next to them on the floor, and the experimenter was seated on the floor on the other side.

Table 1.

Measures administered at each timepoint.

Timepoint	*Measures*	*Tasks*
T1: 2 – 2;5-years-old (N = 61)	Oxford-CDI	
T2: 3 – 3;5-years-old (N = 56)	EOWPVT-4 ROWPVT-4	Fast-mapping and retention PSRep Test
T3: 3;6 – 3;11-years-old [a] (N = 29)	EOWPVT-4 ROWPVT-4 Leiter-3	Cross-situational word learning
T4: 4 – 4;6-years-old (N = 46; remote testing)	Vineland-3	

[a] An additional 8 LT children were also tested at this time point remotely on only the EOWPVT-4 and ROWPVT-4 during COVID-19

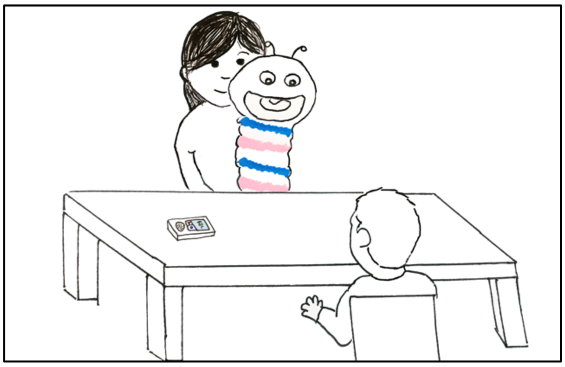
Nonword Repetition Task: The Preschool Repetition (PSRep) Test (Chiat & Roy, 2007)

Stimuli: The PSRep contains 18 word and 18 non-words of varying lengths. Accuracy of children's repetition of the stimuli was recorded.

Procedure: The PSRep Test is designed to maximise young children's participation in nonword repetition tasks using live presentation (Chiat & Roy, 2007). During the task, the experimenter delivered live presentation of real word and non-word stimuli with the use of a sock puppet that had a moveable mouth. The puppet was held in front of the experimenter, blocking the child's view, so the puppet 'said' the words (Figure 2). The experimenter began with four warm-up trials (two real words and two non-words) that were not coded before progressing to the test stimuli. The order of real words and non-words was counterbalanced across participants, with half of them receiving non-words first, and the other half receiving real words first. After the child has made an attempt at repeating the requested item, the experimenter progressed to the next trial. Where the child did not make a response, the item was repeated up to three times in total. Children's responses were recorded on a dictaphone and written in verbal form on the response sheet, and subsequently coded by the experimenter for total items correct (accuracy; for syllable loss, please see Supporting Information, Appendix C) according to the criterion set out by Chiat and Roy (2007). An independent second coder coded the responses from the PSRep Test, showing good inter-rater reliability (Cohen's *k* = .89).

Figure 2.

Preschool Repetition Test set-up. Stimuli are presented live. The puppet is held in front of the experimenter blocking the child's view of the experimenter, giving the illusion that the puppet is speaking.



Fast-mapping and retention task (Hartley et al., 2019)

Stimuli: The task was adapted from Hartley et al.'s (2019) fast mapping task. Participants had four one-syllable novel words to learn: *lep, darg, terb, yok* selected from the NOUN database (Horst & Hout, 2016). The novel words were randomly paired with four novel objects for each participant prior to the task beginning. Novel objects were all different colours and shapes, but approximately the same size, and familiar objects were common objects that were checked for familiarity with the parent beforehand. All object stimuli for this task are presented in Appendix C.

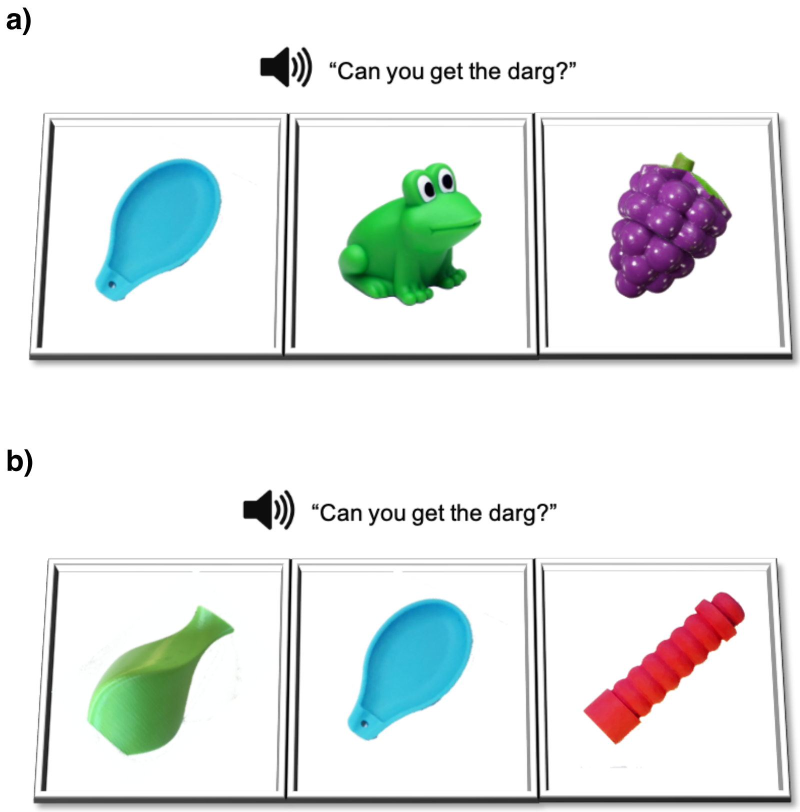
Procedure: Participants began with three warm-up trials where they were asked to select a familiar object from an array of three: 'Look! Can you get the [object name]?'. If they responded correctly, they were told: 'Great job! That is the [object name]!' If they responded incorrectly, they were given feedback: ‘Actually, this is the [object name]. Can you get the [object name]? Well done, you touched the [object name]!’.

Participants then completed eight referent selection trials (Figure 3a) – four Familiar and four Unfamiliar. For each trial, the experimenter would say: ‘Let’s look at some new things!’ and display a tray with three objects: two familiar and one novel. On Familiar trials, children were asked to select a familiar object (‘Can you get the [familiar object]?’. On Unfamiliar trials, children were asked to select the novel object (‘Can you get the [novel word]?’). Regardless of the selection made, the experimenter only said: ‘Thank you.’ The order in which objects were requested was pseudorandomised with the constraint that no more than two trial types of the same kind occurred in a row, and the position of the objects was counterbalanced using a 3x3 Latin Square across participants.

Participants were then given a five-minute break to play with a simple jigsaw puzzle. On return to testing, participants completed 8 retention trials (Figure 3b). For each trial, they were shown three of the novel objects they had learnt words for in the preceding referent selection trials. The experimenter said: ‘Look!’, and after 3 seconds, they requested one of the novel objects using the corresponding label for that object (‘Can you get the [novel name]?’). This repeated until all novel objects had been requested twice. The position of the three objects was pseudorandomised with the constraint that the target did not appear in the same position more than twice in a row. The order in which objects were asked for was counterbalanced across participants using a 3x3 Latin Square.

Figure 3.

Fast mapping and retention task: example of a) referent selection trials; b) retention trials.



Cross-situational word learning task (CSWL; Hartley et al., 2020)

Stimuli: The CSWL task was adapted from Hartley et al. (2020). Stimuli were presented on a Windows 10 SurfacePad Pro touchscreen. There were four two-syllable novel words to learn over 32 trials: *teebu, blicket, fiffin,* and *verdex* from the NOUN Database (Horst & Hout, 2016). For each participant, novel words were pseudo-randomly paired with one of four novel objects with different shapes and colours, but similar sizes (Appendix C).

Procedure: Participants began with three warm-up trials, that were not scored, where they saw two familiar objects and were asked: 'Which is the [familiar word]? Touch the [familiar word]'. The warm-up trials repeated until the participant identified the correct referent for each word, before proceeding to the training trials.

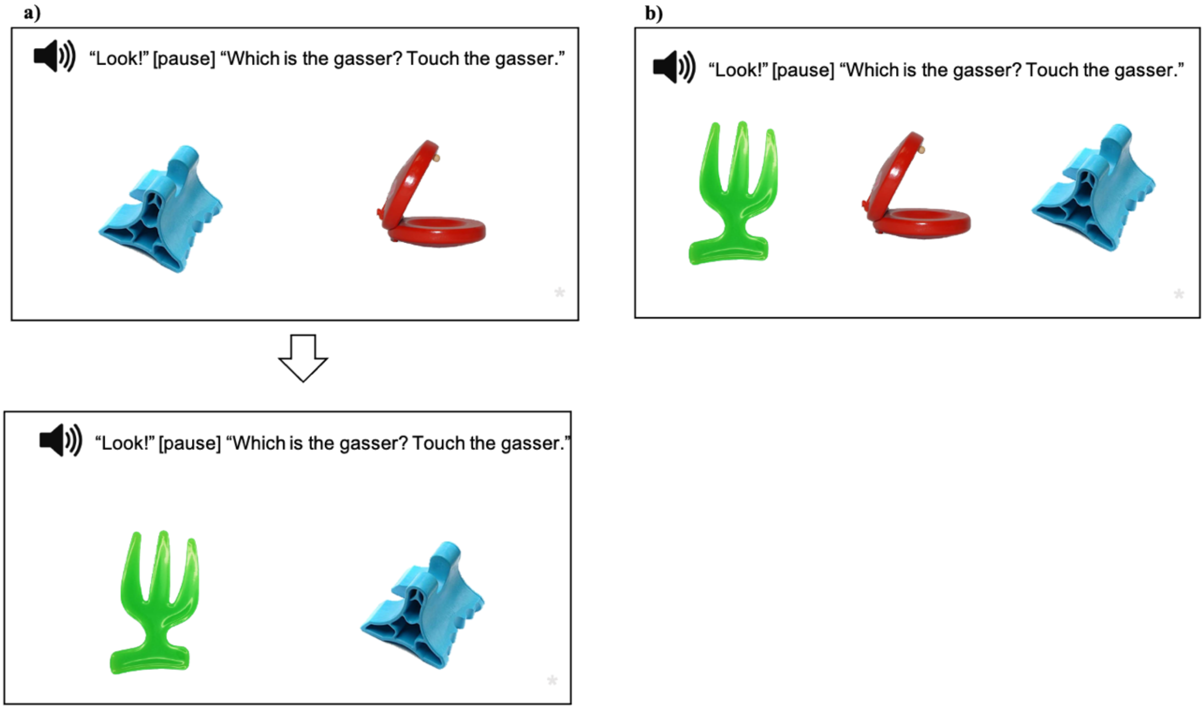
On each training trial, participants saw two objects on the screen. A female voice directed them to: ‘Look!’. After viewing the pictures for 2.5 seconds, the same voice asked: ‘Which is the [novel word]? Touch the [novel word]’ (Figure 4). Each of the four novel word-referent mappings were presented four times; there were 32 trials in total. Each object appeared four times as a target, and four times as a foil. The target appeared an equal number of times on the left and right of the screen, and the order of trials was randomised. When children made their choice by touching the screen, their choice was recorded and the task automatically advanced to the next trial. If they did not make a choice, the experimenter advanced the trial using a hidden asterisk button in the upper right-hand corner of the screen.

The children then had a five-minute break where they played with the examiner using a jigsaw puzzle, before commencing the retention trials. They began with three warm-up trials where they saw three familiar objects positioned on the left, middle, and right of the screen. The target appeared in each of the three possible locations on one trial, and the trial order was randomised so that children selected targets in each location before the testing trials began.

Children completed eight retention trials – each novel word-object pair was tested on two trials. Three of the objects from the training trials were presented on the screen at a time. After viewing the pictures for 2.5 seconds, the female voice asked: ‘Which is the [novel word]? Touch the [novel word]’ (Figure 4). All objects were used four times as foils across the eight trials. The position order was randomised per participant with the constraints that the target object appeared in each position at least twice, and never more than twice in a row. The testing order was also randomised per participant.

Figure 4.

Cross-situational word learning task: a) example of two training trials: the learner is able to infer that the gasser must be the blue object, based on co-occurrence across the trials; b) example of retention trial.


a)
"Look!" [pause] "Which is the gasser? Touch the gasser."
"Look!" [pause] "Which is the gasser? Touch the gasser."
b)
"Look!" [pause] "Which is the gasser? Touch the gasser."

5.6 Results

We first report the study sample characteristics for each timepoint. We then assess the extent to which the mechanisms tested by the nonword repetition test, the fast mapping referent selection and retention task, and the CSWL task, show different performance for the LT and TD groups. We then report how the trajectory of expressive vocabulary relates both predictively and concurrently to these mechanisms of word learning. Data and code from this experiment can be viewed on OSF (https://osf.io/feg6d/?view_only=26b5bcbe085f4822bbede23a88a87471), alongside pre-registrations and a document that explains project deviations due to COVID-19.

Sample characteristics

The final samples for each task can be seen in Tables 2 and 3. All families were from mid-socioeconomic status (SES) backgrounds, measured by parental education levels. At T1 (2 – 2;5-years-old), there were 61 participants (40 TD; 21 LT). Between T1 and T2, 3 TD families dropped out of the study permanently (1 family emergency, 2 uncontactable). One TD family and one LT family dropped out for T2 (both due to pregnancy), but returned to participate in T3.

At T2 (3 – 3;5-years-old), 56 children (36 TD; 20 LT) from the T1 sample participated (Table 2). All but two LT children were above the 10th percentile on the EOWPVT-4, indicating that most of our sample comprised recovered LT children.[5] As a result, all tasks were analysed with T1 expressive vocabulary as a function of population (TD or LT), and T2 and T3 expressive vocabulary as continuous variables, rather than comparing TD children against a LT group that homogenised recovered and non-recovered children. All 56 completed the fast-mapping and retention task at T2. A total of 53 were administered the PSRep Test (34 TD; 19 LT). Of these, 3 TD children refused to speak nonwords due to shyness and only completed real word stimuli. A total of 2 TD children and 1 LT child were excluded due to a high number of incomplete trials (completing less than half of the stimuli, as specified in the pre-registration). These numbers concerning exclusion and non-responders for nonwords were consistent with Chiat and Roy's (2007) results in the same age group.

At T3 (3;6 – 3;11-years-old), 29 participants (20 TD; 9 LT) were tested before the COVID-19 pandemic ceased face-to-face testing (Table 3). All were tested on the Leiter-3; Welch two-sample *t*-tests showed that the TD children and LT children did not differ significantly in non-verbal IQ (Table 3). One TD child and one LT child did not complete the CSWL task due to fussiness. A total of 27 children completed the training trials in the CSWL

[5] At T3, one child remained at the 10th percentile, and the other child reached above the 10th percentile.

task (19 TD and 8 LT), all of whom had completed the fast-mapping task and the PSRep Test at T2. A further 2 LT children did not complete the CSWL retention trials due to fatigue; 6 LT children and 19 TD children successfully completed the CSWL retention trials. A further 8 LT children completed the EOWPVT-4 and ROWPVT-4 online at T3; of these, all had completed the fast-mapping task and 7 had completed the PSRep Test at T2.

At T4 (4 – 4;6-years-old), 46 participants (28 TD; 18 LT) completed the Vineland-3 remotely via video-call or telephone-call during the COVID-19 pandemic. LT children scored significantly lower than TD children on the Vineland-3 Adaptive Behaviour Composite (ABC) scores, (*t*(32.39) = -2.17, *p* = .037; Table 3), but not within thresholds indicative of developmental delay (Sparrow et al., 2016). The ABC combines communication, daily living skills, and socialisation subscales; when these subscales were examined individually, there were no significant differences between groups using Welch two-sample *t*-tests. There were also no significant group differences on the motor subscale or in maladaptive behaviour.

Due to high VIF (>3; Zuur et al., 2010) for expressive and receptive vocabulary scores when entered into the same model, these were analysed separately. As we were interested in examining how early classification of LT related to performance in different mechanisms of word learning, we report here the predictive effect of expressive vocabulary only (see Supporting Information, Appendix C, for analyses of receptive vocabulary). Additionally, to allow for capturing the trajectory of vocabulary development over time, we also tested the relation between word learning performance with concurrent expressive ability.

Table 2.

PSRep Test and fast mapping task: sample demographics and vocabulary at time of testing. Note that unless otherwise specified, standardised scores were used.

Task	*Fast-mapping and retention mean (SD)*		*PSRep Test mean (SD)*	
	TD (n = 36)	*LT (n = 20)*	*TD (n = 34)*	*LT (n = 19)*
Age (decimal years)	3.20 (1.52)	3.18 (1.54)	3.21 (0.13)	3.19 (0.13)
Sex (m : f)	16 : 20	14 : 6	15 : 19	13 : 6
T1 CDI receptive [a]	381.0 (42.1)	266.0 (88.4)	381.0 (43.2)	273 (84.7)
T1 CDI expressive [a]	325.0 (79.6)	61.6 (50.3)	322.0 (80.6)	63.8 (50.7)
T2 ROWPVT-4	166.03 (10.48)	108.50 (10.16)	115.74 (10.67)	108.95 (10.32)
T2 EOWPVT-4	120.47 (9.21)	107.75 (13.80)	120.03 (9.29)	108.47 (13.79)
	TD (n = 19)	*LT (n = 9)*	*TD (n = 17)*	*LT (n = 9)*
T3 Non-verbal IQ (Leiter-3)	98.6 (6.86)	92.0 (12.8)	98.2 (6.86)	92.0 (12.8)
	TD (n = 25)	*LT (n = 17)*	*TD (n = 23)*	*LT (n = 17)*
T4 Vineland ABC [b]	100.56 (6.34)	94.82 (6.43)	99.87 (6.13)	94.82 (6.43)
T4 Vineland Com	144.16 (180.54)	103.47 (25.47)	146.70 (188.33)	103.47 (25.47)
T4 Vineland DLS	96.28 (5.43)	94.76 (7.87)	95.83 (5.41)	94.76 (7.87)
T4 Vineland Soc	98.48 (7.56)	92.82 (7.94)	97.83 (7.44)	92.82 (7.94)
T4 Vineland Mot	98.28 (6.01)	95.76 (8.99)	98.22 (5.66)	95.76 (8.99)
T4 Vineland MB [a]	6.00 (3.04)	7.12 (4.39)	6.0 (3.05)	7.12 (4.39)

ABC = Adaptive Behaviour Composite; Com = Communication subscale; DLS = Daily Living Score subscale; MB = Maladaptive Behaviour subscale; Mot = Motor subscale; LT = late talking; PSRep = Preschool Repetition; TD = typically developing; vocab = vocabulary

[a] *Raw scores used*

[b] *This is a composite of Communication, Daily Living Skills, and Socialisation subscales*

Table 3.

Cross-situational word learning task: sample demographics and vocabulary at time of testing. Unless otherwise specified, standardised scores were used.

Task	*Cross-situational word learning task mean (SD)*	
	TD (n = 19)	*LT (n = 8)*
Age (years)	3.76 (0.12)	3.71 (0.15)
Sex (m : f)	7 :12	6: 2
T1 CDI receptive [a]	394.32 (23.41)	275.38 (58.60)
T1 CDI expressive [a]	350.11 (67.95)	74.38 (57.95)
T3 ROWPVT-4	111.68 (5.39)	111.88 (7.68)
T3 EOPVT-4	122.53 (8.60)	108.38 (13.73)
T3 Non-verbal IQ (Leiter-3)	98.84 (6.74)	93.75 (12.49)
	TD (n = 15)	*LT (n = 8)*
T4 Vineland ABC [b]	102.07 (6.32)	97.63 (6.42)
T4 Vineland Com	110.2 (6.20)	113.0 (34.61)
T4 Vineland DLS	97.80 (4.54)	97.25 (6.36)
T4 Vineland Soc	98.27 (8.97)	96.25 (7.74)
T4 Vineland Mot	99.13 (6.31)	100.13 (9.37)
T4 Vineland MB [a]	6.00 (3.34)	6.25 (5.28)

ABC = Adaptive Behaviour Composite; Com = Communication; DLS = Daily Living Score; MB = Maladaptive Behaviour; Mot = Motor subscale; LT = late talking; PSRep = Preschool Repetition; TD = typically developing

[a] *Raw scores used*

[b] *This is a composite of Communication, Daily Living Skills, and Socialisation subscales*

Differences in word learning mechanisms between LT and TD children

For analyses between groups and examining the relationships between expressive vocabulary and task performance, general linear mixed effects (GLME) models were employed using the functions *glmer* from the package *lme4* in R [v1.1.463]. Across all models, we tested fixed effects of population at T1 to determine how LT status related to task accuracy, and also effects of concurrent vocabulary to determine how this relation might change with vocabulary development. These models were built up progressively, starting with a null model that contained random effects of participant and target word. Fixed effects were then added sequentially, with each model compared to the previous best-fitting model using log likelihood comparisons (Barr et al., 2013). Fixed effects tested are detailed underneath each task section.

Are LT children impaired on nonword repetition?

To examine whether children's performance on the PSRep Test differed according to expressive vocabulary, we predicted accuracy (item correct: incorrect = 0, correct = 1) using two GLME analyses, with: 1) population (determined at T1 using CDI; TD = 0, LT = 1), and 2) concurrent vocabulary as fixed effects. These models were tested alongside fixed effects of word length (number of syllables) and word type (word = 0, non-word = 1), with random effects of participant and target word. Random slopes of participant per word did not converge and so were omitted from the model.

The best-fitting model contained fixed effects of population and word length ($\chi^2(2)$ = 12.73, *p* = .003; Table 4): LT children scored significantly less accurately (*M* = 0.48, *SE* = 0.02) than TD children (*M* = 0.81, *SE* = 0.01; *p* < .001). All children scored less accurately as word length increased (2-syllables: *p* = .007; 3-syllables: *p* <.001). There was no interaction between population and word length, and no effect of word type.

There was also a predictive effect of concurrent expressive vocabulary (T2, EOWPVT-4) on task accuracy. The best-fitting model to the data contained fixed effects of concurrent expressive vocabulary and word length ($\chi^2(2) = 12.79$, $p = .002$; Table 4): accuracy increased with higher expressive vocabulary ($p < .001$), and all children scored less accurately as word length increased (2-syllables: $p = .007$; 3-syllables: $p < .001$). Again, there was no interaction between expressive and word length, and no effect of word type.

Table 4.

Preschool Repetition Test: general linear mixed effects model results predicting item correct by fixed effects of T1 and T2 expressive vocabulary.

Relation with early expressive vocabulary (measured at T1: 2 – 2;5-years-old)				
Fixed effect	*estimate*	*SE*	*z-value*	*p-value*
(intercept)[a]	2.90	0.39	7.39	< .001
T1 population (late talking)	-2.14	0.36	-6.03	< .001
2-syllable words	-1.22	0.46	-2.68	.007
3-syllable words	-1.73	0.46	-3.78	< .001
Relation with concurrent expressive vocabulary (measured at T2: 3 – 3;5-years-old)				
Fixed effect	*estimate*	*SE*	*z-value*	*p-value*
(intercept)	-7.55	1.79	-4.21	< .001
T2 expressive (EOWPVT-4)[a]	8.32	1.52	5.47	< .001
2-syllable words	-1.22	0.46	-2.68	.007
3-syllable words	-1.73	0.46	-3.79	< .001

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

<u>Are LT children impaired in fast mapping?</u> To examine whether children's performance on fast mapping differed according to expressive vocabulary, we predicted accuracy (item correct: incorrect = 0, correct = 1) on referent selection and then retention trials using GLME analyses. These models contained: 1) population (determined at T1 using CDI; TD = 0, LT = 1), and 2) concurrent vocabulary as fixed effects. For models that tested retention trial accuracy, we also added a fixed effect of referent selection accuracy, to assess whether accuracy on referent selection trials affected subsequent retention trials for the same item. Random effects of participant and target item were included. As all participants scored at ceiling on familiar trials, we tested only unfamiliar referent selection trials. A model with random slopes of participant per word did not converge and so were omitted from the final model.

There was no predictive effect of early expressive vocabulary (T1, CDI) on referent selection or retention trials. For referent selection, LT children (*M* = 0.83, *SE* = 0.09) scored on par with TD children (*M* = 0.87, *SE* = 0.07). For retention, LT children (*M* = 0.32, *SE* = 0.11) scored slightly less accurately than TD children (*M* = 0.40, *SE* = 0.08).

There was an effect of concurrent expressive vocabulary (T2 EOWPVT-4) on referent selection trials. A model with fixed effects of concurrent expressive vocabulary provided the best fit to the data ($\chi^2(1) = 15.53(1)$, *p*-value < .001; Table 5). This showed that participants' accuracy during referent selection trials for unfamiliar words increased with concurrent expressive vocabulary (*p* <.001).

There was also an effect of concurrent vocabulary for retention trials.[6] A model with fixed effects of concurrent expressive vocabulary, referent selection accuracy, and an interaction between expressive vocabulary and referent selection accuracy provided the best fit ($\chi^2(3) = 9.20(3)$, *p*-value = .027; Table 5). This model indicated that higher expressive vocabulary predicted higher accuracy (*p* = .023), and that responding accurately on a referent selection trial significantly increased the likelihood of responding correctly on the corresponding retention trial for the same word (*p* = .043). The interaction also indicated that children with higher concurrent expressive vocabulary were more likely to score accurately

[6] A possibility is that the difference in the predictive effect of expressive vocabulary at T1 and T2 was due to a difference in variable type, as T1 was discrete, and T2 was continuous. An additional analysis was run with T1 expressive vocabulary used as a continuous variable, which yielded the same results. Thus, this difference was not likely to be due to variable type.

even if they were incorrect during referent selection. This suggests that not only did higher concurrent expressive vocabulary predict higher accuracy on referent selection trials and subsequent retention trials, but that it may have also enabled children to map words to referents during retention trials even if they had been wrong previously – i.e. children with higher expressive vocabulary may have been able to 'correct' their previous errors actively during testing. However, the effect of this was not significant (*p* = .051), despite the model providing significantly better fit to the data with the interaction than without, so must be interpreted cautiously.

Table 5.

Fast mapping task: results of general linear mixed effects model predicting accuracy in referent selection and retention trials by concurrent expressive vocabulary.

Referent selection trial accuracy				
Fixed effect	*estimate*	*SE*	*z-value*	*p-value*
(Intercept)	-3.07	2.12	-1.45	.015
T2 expressive vocabulary (EOWPVT)[a]	6.36	1.66	3.83	<.001

Retention trial accuracy				
Fixed effect	*estimate*	*SE*	*z-value*	*p-value*
(intercept)	-8.14	3.23	-2.52	.012
T2 expressive vocabulary (EOWPVT)[a]	6.45	2.84	2.28	.023
Referent selection (correct)	6.96	3.43	2.03	.042
T2 expressive[a] * referent selection (correct)	-5.86	3.00	-1.96	.051

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

Do LT children show impairments in cross-situational word learning? To examine the effect of early and concurrent expressive vocabulary, we used GLMEs to predict task accuracy in training (referent selection) trials and then retention trials. We first tested the relation of accuracy to early expressive vocabulary (fixed effect: T1 population (TD or LT, determined by CDI), and then tested relations with concurrent expressive vocabulary (fixed effect: T3, EOWPVT-4), with a random effect of participant. Models with random effects effects of target item, and random slopes of participant per target item, failed to converge so were omitted.

There were no effects of early or concurrent expressive vocabulary on training trials. LT children (*M* = 0.62, *SE* = 0.19) scored on par with TD children (*M* = 0.63, *SE* = 0.12).

However, there was a significant effect of population on retention trial accuracy ($\chi^2(1) = 4.83$, *p* = .028; Table 6), with the best fitting model to the data containing a fixed effect of Population (LT or TD). LT children (*M* = 0.31, *SE* = 0.21) scored significantly less accurately than TD children (*M* = 0.52, *SE* = 0.12; *p* = .025).[7] This must be interpreted with caution as only 6 LT children completed this part of the task due to the COVID-19 restrictions limiting data collection. There was no effect of concurrent expressive vocabulary for retention trials.

Table 6.

Cross-situational word learning task: results of general linear mixed effects model predicting accuracy in retention trials with early expressive vocabulary (T1, CDI).

Fixed effect	*estimate*	*SE*	*z-value*	*p-value*
(intercept)	0.09	0.19	0.45	.650
T1 population (LT)	-0.90	0.40	-2.24	.025

Predicting early and later expressive vocabulary from combined mechanisms

As an exploratory analysis, we used linear models to test the extent to which performance across all tasks combined to relate to early (T1) and later (T3) vocabulary (*lm* base function in R). This enabled us to determine how the child's developing expressive vocabulary related to the mechanisms investigated in the tasks.

[7] An additional analysis using T1 as a continuous variable, as for the fast mapping task was not possible, as these models failed to converge.

Using data from all timepoints (T1, T2, T3) from children who completed all three tasks (*N* = 22; 6 LT, 16 TD), the model significantly predicted 32% of variance in children's T1 vocabulary at 2 – 2;5-years-old (Table 7; *adjusted* R^2 = 0.32; *F*(5, 16) = 2.04; *p* = .046). However, only the PSRep Test was a significant predictor of children's past T1 vocabulary at 2 – 2;5-years-old. When predicting future T3 vocabulary at 3;6 – 3;11-years-old, all three tasks combined predicted 45% of the variance (Table 7; *adjusted* R^2 = .45; *F*(5, 16) = 4.45; *p* = .010). Of the predictor variables, only the PSRep Test predicted children's future vocabulary.

Table 7.

Predicting early and later vocabulary by task performance (accuracy) using data from all timepoints (T1, T2, T3; *N* = 22).

Predicting early (T1; CDI) expressive vocabulary at 2;0 – 2;5-years-old				
Variance	*estimate*	*SE*	*t-value*	*p-value*
(intercept)	-44.56	211.70	-0.21	0.836
Preschool Repetition Test	4.41	1.49	2.96	.009
Fast mapping referent selection	-0.75	1.60	-0.47	.645
Fast mapping retention	0.90	1.47	0.61	.549
Cross-situational word learning referent selection	-0.14	3.15	-0.05	.964
Cross-situational word learning retention	0.86	1.48	0.58	.570
Predicting later (T3; EOWPVT-4) expressive vocabulary at 3;0 – 3;5-years-old				
Variance	*estimate*	*SE*	*t-value*	*p-value*
(intercept)	53.07	19.00	2.79	.013
Preschool Repetition Test	0.32	0.13	2.43	.027
Fast mapping referent selection	0.01	0.14	0.04	.968
Fast mapping retention	0.23	0.13	1.77	.096
Cross-situational word learning referent selection	0.52	0.28	1.85	.083
Cross-situational word learning retention	-0.03	0.13	-0.25	.080

As the analysis that contained all three timepoints was considerably smaller due to interruption of T3 data collection, we also conducted an additional analysis using data from children who completed all tasks at T1 and T2 and for whom we also had T3 data. The model (data: $N = 53$; 19 LT, 34 TD) predicted 40% of the variance in children's past T1 vocabulary at 2;0 – 2;5-years-old (Table 8; *adjusted* $R^2 = 0.40$; $F(3, 49) = 12.32$; $p < .001$). Only the PSRep Test was significant in relation to children's early vocabulary. When relating to later T3 vocabulary at 3;0 – 3;5-years-old, the model (data: $N = 33$; 16 LT, 17 TD) predicted 47% of the variance (Table 8; *adjusted* $R^2 = 0.47$; F(3, 29) = 10.28; $p < .001$). The PSRep Test and fast mapping retention accuracy predicted children's later vocabulary.

Table 8.

Predicting early and later expressive vocabulary by task performance (accuracy) using data from completed timepoints (T1, T2).

Predicting early expressive vocabulary at 2 – 2;5-years-old (T1, CDI; *N* = 53)				
Variance	*estimate*	*SE*	*t-value*	*p-value*
(intercept)	26.75	76.30	0.35	.727
Preschool Repetition Test	4.17	0.70	5.93	<.001
Fast mapping referent selection	-1.07	0.78	-1.37	.178
Fast mapping retention	0.14	0.80	0.18	.860

Predicting later expressive vocabulary at 3 – 3;5-years-old (T2, EOWPVT-4, *N* = 33)				
Variance	*estimate*	*SE*	*t-value*	*p-value*
(intercept)	85.92	7.92	10.84	<.001
Preschool Repetition Test	0.28	0.07	3.99	<.001
Fast mapping referent selection	0.04	0.09	0.48	.634
Fast mapping retention	0.22	0.09	2.53	.017

5.7 Discussion

Studies of word learning mechanisms in LT children offer the chance to unpack how early expressive language is delayed and relates to word learning over time. We identified three critical tasks that highlight key mechanisms involved in word learning: perception and production of phonology, selection and retention of referents, and acquisition of associations between words and referents. We further tested LT and TD children's vocabulary development during the study to investigate how vocabulary growth related to these word learning mechanisms.

LT children continue to show impairments in phonology, but are able to select referents accurately

LT children were impaired on the PSRep Test, consistent with the literature (e.g. Marini et al., 2017). However, unlike Weismer et al. (2013), LT children did not show impaired performance as compared to TD children during referent selection in fast mapping or in CSWL. This may have been as a result of our sample containing recovered LT children, whereas previous studies have tested non-recovered LT children at a younger age. LT children that reach the TD vocabulary range thus appear to be able to fast map unfamiliar words on par with TD children. LT children also scored at ceiling for comprehension of familiar items during fast mapping on par with TD children, but scored less accurately on the PSRep Test for real words as well as for nonwords. This demonstrated that although they were able to identify known objects without difficulty, LT children's ability to produce both familiar and unfamiliar words was compromised.

Recovered LT children show possible deficits in retaining statistical information from the environment

LT children showed evidence of impairment on CSWL retention trials, but not on fast mapping retention trials. This suggests that despite reaching TD ranges, LT children may have a weaker encoding of links between words and referents that is tapped by tasks which test retention from repeated exposures, such as CSWL, but not by tasks that test only single exposures and immediate referent selection, such as fast mapping. Although our CSWL results must be interpreted with caution, given the much smaller sample as a result of COVID-19, they do suggest fertile ground for future research for testing CSWL in language delay.

As CSWL performance did not relate to concurrent vocabulary across the sample, however, these results might also be secondary to more general processes that run parallel to vocabulary acquisition, such as working memory, which may also be implicated in nonword repetition (Marini et al. 2017). This may be attributed to the reduced sample sizes at T3 due to COVID-19, random variability, or could be task-related. Lab-based CSWL tasks may well test general purpose learning mechanisms that are of use for initial referent selection and competition as outlined by McMurray et al. (2012), but whether or not performance on these tasks correlates with children's longer-term vocabulary remains to be further investigated. For example, Vlach and DeBrock (2019) found that receptive vocabulary did not predict CSWL task performance in 3-year-olds.

Although we did not find differences in fast-mapping abilities or in initial CSWL referent selection trials, other studies that directly test online learning have found differences during the learning process itself, despite no differences in overall accuracy. For example, Ellis et al. (2015) tested novel word learning with an eye-tracker (looking-while-listening paradigm) in 18-month-olds and found that, although TD and LT children looked equivalently at the target, there were between-population differences in looking behaviour during testing. They proposed that LTs divided their attention between target and foil equally, being uncertain about the target, whereas TD children predominantly focussed on the target. This is consistent with Ahufinger et al. (2021), who found children with DLD showed more ambivalence when fixating between targets and competitors at test during CSWL, whereas TD children showed a rapid increase in looks to target over competitors. As we did not use eye tracking in our study, it is possible that LTs showed a similar pattern of uncertainty around the target that was not captured by referent selection, but was captured when testing retention trials, which test the robustness of learnt word-referent pairs. It is thus possible that even if accuracy between groups does not differ, strategies applied during word learning tasks might.

Understanding the trajectory of vocabulary development through word learning mechanisms

Our results also showed how, across the whole sample, children's expressive abilities may interact with word learning mechanisms as their vocabulary develops over time. Firstly, our analyses that showed the higher the concurrent expressive vocabulary of children, the more accurately they scored on not only the PSRep Test, but on both referent selection and retention fast mapping trials. Secondly, both PSRep Test and fast mapping retention predicted expressive vocabulary scores at the last time point, suggesting that children's ability to not only store phonological information, but also their ability to retain fast mapped word-referent pairs, appears to influence their ability to add words to their expressive vocabulary later on.

Expressive vocabulary may thus be the result of storing robust semantic *and* phonological representations, where phonological representations are both auditory and articulatory. Thus, although recovered LT children were able to recognise stimuli and activate semantic representations sufficiently during referent selection in order to comprehend novel words in both fast mapping and CSWL, they may have had weaker phonological representations stored in their expressive vocabulary as a secondary to their early language delay, resulting in a reduced ability to produce both words already in the lexicon (real words) and to utilise existing knowledge to produce novel words. Deficits in the CSWL retention trials also hint at possible additional deficits in retaining statistical information that may compound LT children's ability to add to their existing lexicon.

Overall, these results are consistent with Edwards and colleagues (Edwards et al., 2004; Munson et al., 2005) and Stokes (2010, 2014; Stokes et al., 2012) who suggest that, as a part of a dynamic system between phonology and the lexicon, smaller expressive vocabularies result in less support for storing, generating, and using phonological representations, which in turn feeds back into further development of the lexicon. Although both receptive and expressive vocabulary tests tap both phonological and semantic representations, expressive vocabulary places more weight on stored phonological representations that connect both auditory processes (involved in recognising words) and oromotor processes (involved in articulating words; Edwards et al., 2004). For comprehension tasks, phonological representations can be relatively weak – one only needs to recognise a given stimulus to activate semantic representation. For production, however, both phonological and semantic representations must be sufficiently strong to reproduce a stimulus faithfully enough to be recognised by someone else.

Our results also highlight the benefit of adopting individual differences as part of language acquisition studies, as opposed to grouping children into categories. Throughout our analyses, we used mixed effects models that allow for random effects of participant. For LT children in particular, embracing this heterogeneity may explain a large amount of the variance that has yet to be identified. Moves towards this have been made in LT (Fernald & Marchman, 2012; Perry & Kucker, 2019) and TD populations (Samuelson, 2021), but are yet to be widely adopted as standard. Future studies could also employ the use of mixed effects modelling, as well as testing a wide range of vocabulary ability, to better characterise LT populations and their subsequent outcomes.

Limitations and future directions

One major limitation towards the end of the study was the interruption of testing due to the COVID-19 pandemic. This meant that T3 data was incomplete, and non-verbal IQ data could not be collected for the whole sample. This also meant that only eight and six LT children took part in the CSWL training and retention trials respectively. Findings from the CSWL task must thus be interpreted tentatively, and require replication in a much larger sample. The fact that two LT children and one TD child could not tolerate the retention trials may also have reflected some individual differences in attention that were not controlled for.

We also did not test fast mapping production or generalisation, only comprehension. Weismer et al. (2013) for example showed that LT children's vocabulary scores and fast mapping performance were inter-domain (expressive vocabulary predicting production, receptive predicting comprehension) whereas TDs were cross-domain (both vocabulary scores predicting both tasks). This was because the third session was particularly long as a result of the Leiter-3, and pilot testing had shown children had trouble tolerating the session even with breaks. However, as expressive vocabulary predicted fast mapping comprehension across our sample, this suggests the LT children tested here were not limited to inter-domain relationships between task and vocabulary.

Another limitation is that our sample consisted of relatively similar families from mid-high SES backgrounds who had actively signed up for an 18-month longitudinal study on child development. However, although this means our findings may not generalise to samples with different demographic features, they do suggest that where similar family environments that have resources, time, and interest in child development, LT children may have a good chance of catching up to their peers in terms of vocabulary, as all but one of the children reached typically developing range by the last timepoint.

Conclusion

This study indicates that LT children are impaired across some, but not all, mechanisms involved in the different stages of word learning. Despite most LT children recovering at time of testing, they still exhibit significant differences in their ability to encode and repeat words – making more errors when repeating both real words and non-words– even when individual differences are taken into account. This result is also consistent with

LT children having weaker phonological representations in models that describe phonological and lexical development as dynamic processes that affect one another (Edwards et al., 2004; Stokes, 2010; 2014). Furthermore, although LT children do not show any impairment in the initial referent selection stage, as tested by fast mapping or CSWL tasks, they do show evidence to suggest they may also be less able to retain information learnt through CSWL. Overall, our results add to the evidence base surrounding word learning mechanisms in LT children by highlighting the interplay between expressive vocabulary and word learning mechanisms over time.

6 Chapter 6: Receptive and expressive language ability differentially support symbolic understanding over time: Picture comprehension in late talking and typically developing children

6.1 Chapter introduction

A primary focus of the longitudinal study was to investigate word learning in LT children, as covered in Chapter 5; however, an additional aim was to examine the wider effects of LT status on other areas of development, as the effects of early language delay on domains that are heavily tied to language, such as social ability and symbolic development, have not been well researched.

Symbols more generally, and the ability to use them, form a vital part of communication throughout life. In particular, because language scaffolds understanding of non-linguistic symbols such as pictures before the fourth year (Callaghan, 2000; J. Kirkham et al., 2013), any early deficit in language ability could have cascading effects on symbolic development more broadly. Although symbolic play has been found to be reduced in children with DLD (Casby, 1997; Rescorla & Goossens, 1992) and ASD (Hartley & Allen, 2015, 2014), and one study of dyslexia showed less symbolic play in a subset of LT children (*n* = 14) than TD controls (Lyytinen et al., 2001), no prior studies have examined symbolic development in LT children.

Symbols are described generally as a culturally scaffolded system (Callaghan et al., 2011), and as a result, heavily overlap with socio-cognitive development (Tomasello et al., 2005). However, social ability in LT children has been examined mostly only in relation to behavioural and emotional outcomes. For example, the ALSPAC study (Clegg et al., 2015) examined emotional and behavioural functioning at 6 years of age as an outcome variable with expressive vocabulary at 2 years. The results indicated that expressive vocabulary at the age of 2 had a mild, but significant, effect. Longobardi et al. (2016) examined the relationship between social ability and language in 268 children aged 18–35 months, and found that language ability predicted social competence. Impaired social and emotional functioning has also been found in some samples of LT children when compared to TD peers. Irwin et al. (2002) examined emotional and behavioural problems in a sample of 14 LT children (≤10th percentile CDI) and 14 TD matched controls. They found that the LT children were more likely to have problem behaviours, including depression/withdrawal, competence, compliance, and showed less interest in play. Conversely, Whitehouse et al. (2011) found, although LT children (*n* = 142) had higher concurrent rates of behavioural and emotional problems at 2-years-old, there was no association at later follow-up between 5–17 years of age.

Horwitz et al. (2003) followed up a sample of American children from 18-months-old to 39-months-old. At 24-29 months of age, children with language delay (*n* = 47; ≤10th percentile CDI) had lower social ability than TD children (*n* = 293). However, no significant differences between the groups in externalising, internalising, or dysregulation behaviours were found. At 30–39 months, those with language delay continued to have lower social ability and also showed significantly more externalizing behaviours than TD children. Despite these findings, it is worth noting that although Horwitz et al. asked parents to report developmental delay in their demographics questionnaires, they did not separate LT children from those with language delay resulting from developmental disorders (e.g. ASD) in their analyses. This methodological limitation is especially important as Rescorla et al. (2007) found that correlations between language and internalizing/externalizing behaviour in two US samples of children did not remain significant once those with neurodevelopmental delay and pervasive development disorder were excluded. They did, however, find that those with language delay had higher social withdrawal compared to TD children.

In sum, any relationship between language delay and social ability will be difficult to understand, particularly as these factors are likely to be bidirectional. Despite this, language acquisition works in tandem with socio-cognitive mechanisms from an early age (Hollich et al., 2000; Tomasello, 2003, 2010). Given this, differences in social ability related to LT and symbolic understanding require further investigation.

In the following paper, the wider effects of expressive language delay are examined by utilising a cross-domain approach, considering how receptive and expressive vocabulary alongside social ability can affect children's symbolic understanding of pictures.

Author contribution for Chapter 5: *Rachael W Cheung:* design, data collection, analysis, writing, review. *Calum Hartley:* design, review. *Padraic Monaghan:* design, review

Submitted for publication to: *Journal of Experimental Child Psychology* (Cheung, R.W., Hartley, C., & Monaghan, P. (minor revisions) Receptive and expressive language ability differentially support symbolic understanding over time: picture comprehension in late talking and typically developing children*.* Preprint: https://psyarxiv.com/tjw72/.

6.2 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;5 years and 3;6 – 3;11 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;0-years-old, and expressive vocabulary predicted performance at ~3;6-years-old. Task performance at 3;6-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;0-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.

6.3 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;6-

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; K. 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. J. 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; J. 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). The majority of LT children recover by approximately age 5 (Rescorla, 2011). However, 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 autism 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;6-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;6-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 (K. 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 TD 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; J. 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 TD 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;0 – 2;5-years-old (timepoint 1; T1) and 3;6 –

3;11-years-old (timepoint 2; T2). We also considered the effect of social ability measured at 2;0 – 2;5-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.[8] As an exploratory analysis, we also hypothesised that children with less sophisticated social ability would score lower on picture comprehension accuracy.

6.4 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;5-years-old to 3;6 – 3;11-years-old. The picture comprehension task was administered at the first and last time points.

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 ≥ 25th percentile, or LT with productive vocabulary score ≤ 10th 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.

[8] 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.

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;5-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;6 – 3;11-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 (Supporting Information, Appendix D for all stimuli).

Pictures: Sixteen black and white laminated cards were used that had a simple black pen drawing of one familiar or unfamiliar object.

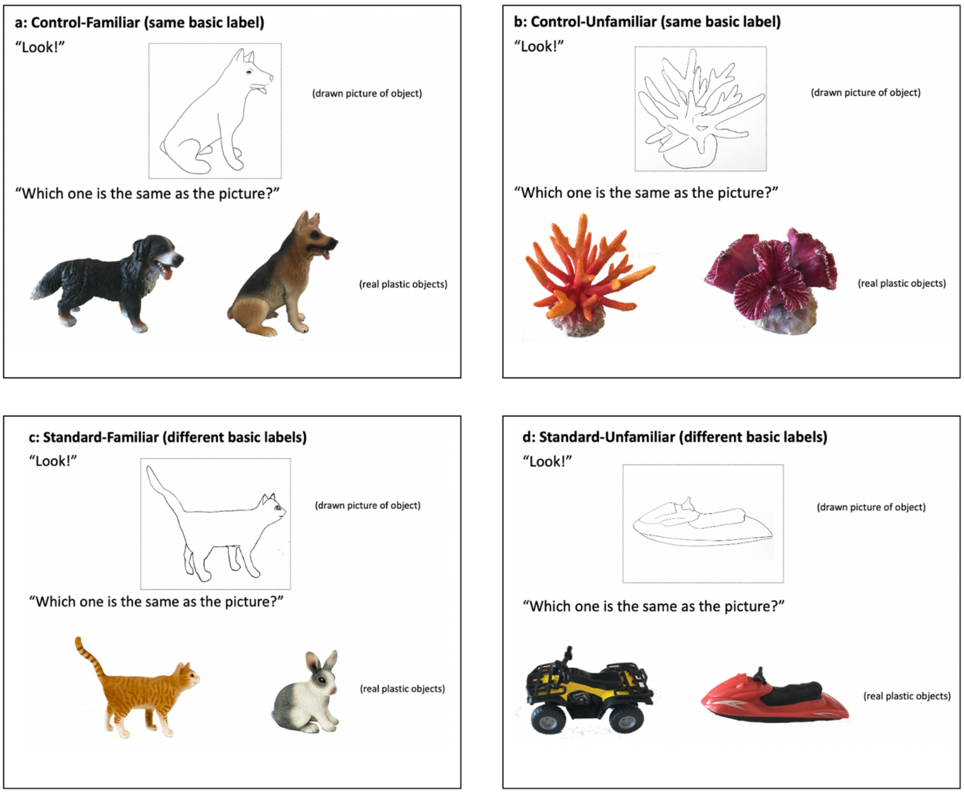
Display: Objects were placed on a tray with a deep lid that had a handle and a cut out at the back that allowed the experimenter to rearrange objects out of sight of the child. The objects remained hidden until the experimenter lifted the lid to reveal the two objects sitting on the tray.

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.


a: Control-Familiar (same basic label)
"Look!"
(drawn picture of object)
"Which one is the same as the picture?"
(real plastic objects)
b: Control-Unfamiliar (same basic label)
"Look!"
(drawn picture of object)
"Which one is the same as the picture?"
(real plastic objects)
c: Standard-Familiar (different basic labels)
"Look!"
(drawn picture of object)
"Which one is the same as the picture?"
(real plastic objects)
d: Standard-Unfamiliar (different basic labels)
"Look!"
(drawn picture of object)
"Which one is the same as the picture?"
(real plastic objects)

6.5 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;5-years-old (*N* = 59)		T2: 3;6 – 3;11 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 2;0 – 2;5-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;6 – 3;11-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; LT: *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.

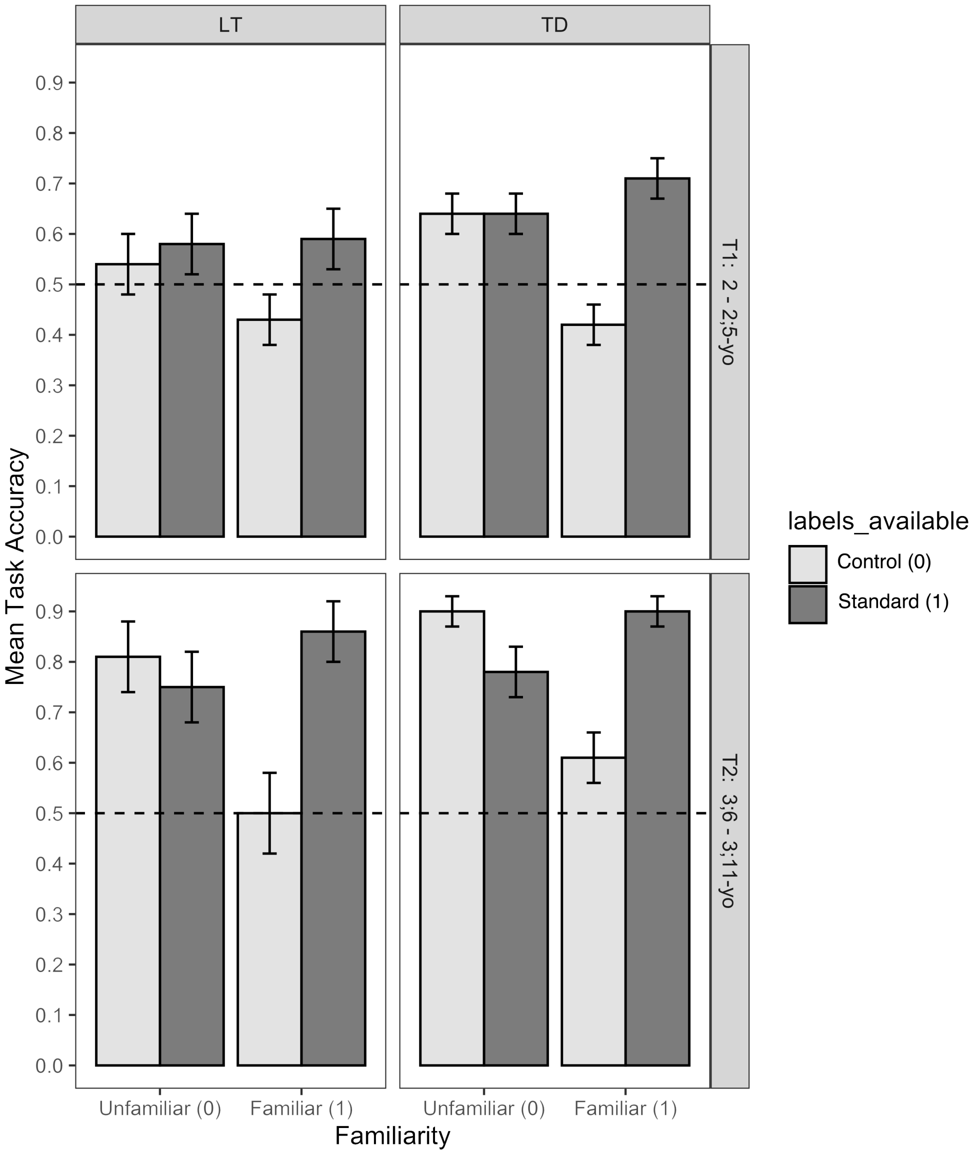
Mean accuracy and standard error at test at each timepoint per group. Trial types: *Control* = object pairs with the same global label and same basic label, inhibiting verbal scaffolding; *Standard* = object pairs with the same global label and different basic labels, allowing verbal scaffolding; *Familiar* = 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;5-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;6 – 3;11-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.[9] All models run can be viewed on the Open Science Framework (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 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?

[9] 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.

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;6– 3;11-years-old as compared to 2;0 – 2;5-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: general linear mixed effect model results predicting 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;6 – 3;11-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;5-years-old).*

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 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;0 – 2;5-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;5-years-old, performed more accurately on the picture comprehension task when they were 3;6 – 3;11-years-old ($p = 0.18$).

Table 4.

Cross-sectional analyses of predictive effect of receptive vocabulary on task accuracy. General linear mixed effect model results predicting T1 and T2 accuracy as dependent variables (cross-sectional) with fixed effects of trial type (familiarity of objects and availability of labels) and T1 receptive vocabulary.

T1: age 2;0 – 2;5-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;6 – 3;11-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;5-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;6 – 3;11-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: general linear mixed effect model results predicting 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;6 – 3;11-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

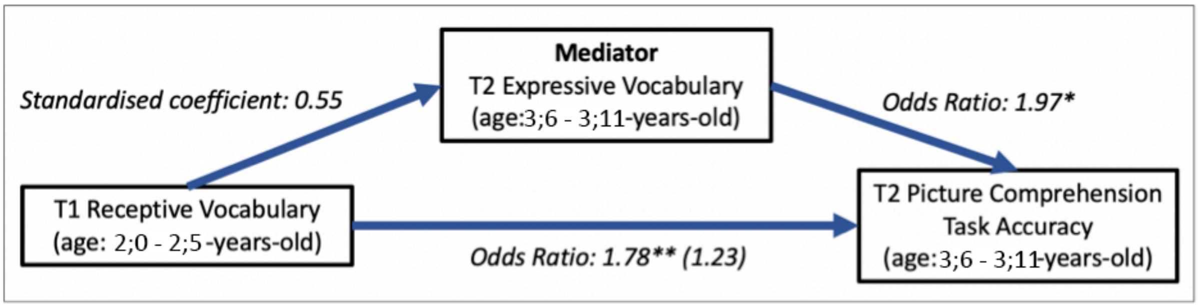
[a] *Intercept corresponds to no language scaffolding (0) and object unfamiliar (0)*

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;6-years-old, and later expressive vocabulary at ~3;6-years-old predicted concurrent task accuracy at ~3;6-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;6-years-old (total effect: 0.10) due to earlier receptive vocabulary at 2;0 – 2;5-years-old, 0.07 was estimated to be mediated through later expressive vocabulary at 3;6 – 3;11-years-old, and 0.03 was estimated to be from earlier receptive vocabulary at 2;0 – 2;5-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.


Mediator
T2 Expressive Vocabulary
(age:3;6 - 3;11-years-old)
Standardised coefficient: 0.55
Odds Ratio: 1.97*
T1 Receptive Vocabulary
(age: 2;0 - 2;5-years-old)
Odds Ratio: 1.78** (1.23)
T2 Picture Comprehension
Task Accuracy
(age:3;6 - 3;11-years-old)

**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;0 – 2;5-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. General linear mixed effect model results predicting 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.

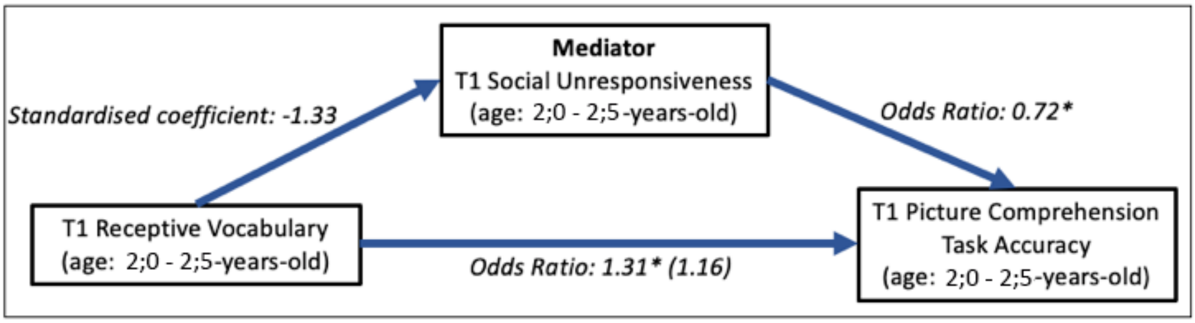
T1: age 2;0 – 2;5-years-old				
Fixed effect	*estimate*	*SE*	*z-value*	*p-value*
(*intercept*)[a]	0.54	0.42	1.28	.200
Familiar	-0.75	0.21	-3.54	<.001
Standard	0.06	0.21	0.32	.790
Familiar*Standard	0.98	0.30	3.14	.001
T1 receptive vocabulary (CDI)	0.09	0.09	1.11	.265
T1 social ability (SRS-2)[b]	-1.45	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 T1 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.


Mediator
T1 Social Unresponsiveness
(age: 2;0 - 2;5-years-old)
Standardised coefficient: -1.33
Odds Ratio: 0.72*
T1 Receptive Vocabulary
(age: 2;0 - 2;5-years-old)
Odds Ratio: 1.31* (1.16)
T1 Picture Comprehension
Task Accuracy
(age: 2;0 - 2;5-years-old)

**p* <.05

6.6 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;0 – 2;5 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;0 – 2;5 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;5 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 clinically significant levels of impairment using the SRS-2. Rather, the results reflected individual differences in social proficiency. Future studies that examine significant 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;6-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;0-years-old, and expressive vocabulary predicting picture comprehension at 3;6-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.

7 Chapter 7: General discussion

The contribution of this thesis to the literature is discussed in two parts, focusing on typical development, and then atypical development. The gaps in the literature and the results of the relevant papers are summarised first in each section. The implications of these findings are then discussed with limitations and future directions for the relevant research field.

7.1 What does this thesis add to the typically developing literature?

The first part of this thesis aimed to identify how and when gesture cues, as part of a multiple cue model, interact with variability in the environment to affect word learning in typical development.

Gaps in the typical development literature

In typical development, how gesture cues interact with the environment to affect word learning remains uncertain. Whilst we know that caregivers adapt their labels and gestures to accommodate their child's perspective (e.g. Masur, 1997; Vigliocco et al., 2019), whether or not they adapt their gestures to the degree of referential ambiguity in the environment, and the impact this may have on children's word learning accuracy, have not yet been tested. Furthermore, despite calls to identify the contextual cues within- and across-trial learning as the process unfolds (L. B. Smith et al., 2010), and naturalistic studies demonstrating that labels and gestures are tightly woven together in time (Trueswell et al., 2016), few studies of cues and cross-situational word learning directly test the temporal unfolding of how visual and auditory cues interact with each other. Investigating these topics is vital to understanding how cues can be integrated into multiple cue models of word learning, but also allows us to observe how cues interact with the variability that naturally accompanies child language acquisition.

The gaps in the typically developing literature were addressed by Chapter 2 and 3 (first and second papers). Chapter 2 utilised a computational model and a behavioural study that examined the effect of referential ambiguity on caregiver gesture use and child word learning. Chapter 3 reported a series of behavioural experiments in adults, with and without eye-tracking, to test the effects of manipulating the timing of a pointing gesture cue in relation to label utterance on cross-situational word learning.

Summary of results: Chapter 2 and 3

In Chapter 2, a computational model (Multi-Modal Integration Model, MIM; Monaghan et al., 2017) was adapted to test the influence of a pointing gesture cue during a cross-situational word learning task across three referential ambiguity conditions. These were: 1) one referent only (no ambiguity), 2) two referents (some ambiguity), and 3) six referents (high ambiguity). The variability of the gesture cue within each referential ambiguity condition was also manipulated across four conditions, occurring: 1) 0% of the time, 2) 33% of the time, 3) 67% of the time, and 4) 100% of the time. The model learnt most robustly when cues were present 33% or 67% of the time, and when there was more than one referent. This model was tested in a behavioural study of 18 – 24-month-olds and their caregivers. Children learnt three words under the same referential ambiguity conditions as the model, and caregivers' gesture and speech use when teaching their children words was video-coded. Children were then tested on their word learning accuracy. The results showed that caregivers used more deictic gestures in the two- and six-referent conditions as compared to the one-referent condition, but there was no difference in deictic gesture use between the two- and six-referent conditions, despite the higher referential ambiguity in the six-referent condition. Children also learnt most accurately in the two- and six-referent conditions, rather than in the one-referent condition. These results were consistent with the computational model.

In Chapter 3, the role of deictic gesture cues during cross-situational word learning was examined in a study of adult learners across three experiments. The first experiment tested the presence and absence of a pointing gesture cue across two conditions, one with two referents (referential ambiguity), and one with only one referent (no referential ambiguity). Learners performed equivalently in conditions with no referential ambiguity (one-referent, with and without gesture cue) and two referents with a gesture cue. This suggested that the benefit of gesture cues was due to reduction of referential ambiguity to the same level as no ambiguity. The second experiment fixed the presence of referential ambiguity to two referents, and tested the effect of altering the timing of a gesture cue, shifting the gesture cue one second before (early condition) and one second after (later) the auditory label. This showed a learning advantage for the early condition. The third experiment tested the same conditions but with an eye-tracker to examine the time course of learning. This revealed that participants' looking behaviour just after label utterance was the critical time period during training that predicted their learning accuracy, and they were more likely to fixate on the target object in the early, rather than late, gesture cue condition from even the first exposures to novel words.

Implications: how do deictic gestures help noisy language learning?

The role of gesture in vocabulary development is relatively well established; we know that infant gesture is tied to vocabulary development, and parent gesture is positively correlated with infant gesture (Rowe et al., 2008). However, Chapters 2 and 3 advance understanding of *how* gesture can support word-referent mappings within caregiver-child interactions.

In Chapter 2, we showed that computational modelling and behavioural experiments can be combined to provide evidence for how gesture cues can interact with referential ambiguity to affect learning. Crucially, by showing that the modulation of gesture cues by caregivers is dependent on the number of referents in the environment, and that infants are able to learn more accurately in conditions with more than one referent, we highlighted how language outcomes can be robust even in the face of referential ambiguity. This advances the literature around variability in language learning, showing how some variability may actually benefit learning. This is important, as the availability of cues provided by caregivers and information provided by the environment is not always consistent, and yet children still manage to learn words despite this variation. Gesture cues are thus useful tools employed by caregivers in the face of referential ambiguity during word learning exchanges, and their use is contingent on the environmental context.

In Chapter 3, we examined how gesture cues influence the accuracy of word-referent mappings, asking whether their benefit was in the reduction of referential ambiguity, or whether having two objects during training might enable a comparison and contrast strategy. Experiment 1 of this study showed that the benefit of gesture cues was in reducing the amount of referential ambiguity, as the addition of a gesture cue to a two-referent condition yielded word learning accuracy on par with a one-referent condition where there was no referential ambiguity. Gesture cues thus support cross-situational word learning through reducing referential ambiguity.

Together, Chapters 2 and 3 demonstrated that deictic gestures not only serve an important function in managing referential ambiguity during word learning exchanges between caregivers and infants, but that pointing gestures also provide a clear cue to meaning that can be manipulated during learning itself. The beauty of deictic gestures more generally is that they *can* potentially serve the dual process of referring to a desired object from a socio-pragmatic sense, and also as an attentional cue, without lessening the contribution towards a desired goal of identifying which referent a label refers to. In line with multiple cue models of word learning, an integrative approach between socio-pragmatic and attentional cues may also help further understanding of how children use these cues to learn words. For example, socio-pragmatic principles may underscore the later use of attentional cues. Wu, N. Kirkham, and colleagues (Wu et al., 2014; Wu & Kirkham, 2010) have shown that infants as young as 8-months-old are able to associate ostensive cues such as faces addressing infants with novel attention cues such as flashing lights. They suggest that ostensive signals such as eye gaze are relied on earlier in life, but infants may learn to use other cues such as pointing and arrows later on when they are paired first with communicative cues. Deictic gesture cues are thus prime candidates for further examination during language acquisition as they occur naturally during speech, unlike other potential endogenous cues such as arrows, and can be studied during caregiver-infant interactions. As such, they are useful and valuable signals that can help relate multiple cue models of word learning to naturalistic settings.

Implications: when are deictic gesture cues most useful?

Chapters 2 and 3 also provide evidence concerning *when* gesture cues are most useful. Contrary to our predictions, but consistent with the computational model, Chapter 2 showed that caregivers did not gesture significantly more when there were six-referents compared to two; the presence, rather than degree, of referential ambiguity was most influential on caregivers. Thus, deictic gesture cues were most useful in the *presence* of referential ambiguity irrespective of the amount. This raises the possibility that caregivers may utilise deictic gesture cues to reduce cognitive load, for example, by reducing the choice between potential referents to one, or more than one. In other words, conditions with more than one referent may be equivalent to one another; here, the role of the gesture may have been to highlight the intended referent from an array of non-target competitors, effectively communicating to the infant “look at this, not that/those.” This also suggests that caregivers may constrain the input for infants through gesture during real-world word learning, which might affect infant attention to objects, and subsequent word-referent mappings (Pereira et al., 2014).

The way in which the temporal process of word learning unfolds is also highly important to understanding how predictions about word-referent mappings are actively made in real time. In Chapter 3, we highlighted how cues preceding an auditory label for a visual object conferred a learning advantage to cues that came after the label, theorising that this was due to the higher predictive value of early cues, and potentially weaker associations between the target label and non-referent distractors. As naturalistic data shows, caregiver gestures in natural language learning exchanges with infants tend to occur before, rather than after, labels for objects (Frank et al., 2013). Early and late gestures may therefore also serve different purposes within language learning. Whereas an early gesture points a learner to an intended referent prior to label exposure, enabling effective learning through better prediction of novel word-referent mappings, or reducing spurious associations, late gestures may be less effective due to an increased period of referential ambiguity prior to word exposure. However, where late gestures do occur, our results indicated this is better than no gesture at all, as learners in Experiments 2 and 3 in the late gesture condition (two referents, gesture after the word) learnt better than those in Experiment 1 (two referents, no gesture). Late gestures thus might serve a feedback role for learners; where cross-situational statistics are utilised to converge upon word-referent pairs, late gestures may confirm specific candidates, and in some cases, correct inaccurate ones.

7.2 Limitations and future directions for typical development

The use of adult populations for insights into child language acquisition

Chapter 3 utilised an adult population to assess in-the-moment learning within a constrained setting. Although this enabled us to precisely manipulate and measure the effects of timing, the use of an adult population does limit our capacity to generalise findings to a developmental population. During word learning, children appear to reason differently to adults, valuing informative context over deductive logic (Ramscar et al., 2013), and have different visual experiences compared to adults, with one object centred in view at a time, rather than several potential referents at once (Pereira et al., 2014; Yurovsky, Smith, et al., 2013). Children also have different memory and attention capacities that change over development (Gathercole et al., 1994). Whilst there is considerable value to understanding temporal processes through adult learners, and cross-situational word learning studies frequently use adults participants to identify developmental learning mechanisms (e.g. Fitneva & Christiansen, 2011; MacDonald et al., 2017; Monaghan et al., 2017; Yu et al., 2012; Yurovsky, Yu, et al., 2013; Yurovsky & Frank, 2015), findings from adult studies must not be generalised to developmental populations without further testing in younger age groups. Future directions thus include adapting the task for preschool aged children. Additionally, as Chapter 3 utilised highly controlled conditions, any adaptations to the task would need to take into account a more naturalistic word learning context.

Beyond deictic gesture cues: what's next for multiple cue models?

In Chapter 2, we evaluated different types of caregiver gesture but focused on deictic gestures in particular as they were the most frequent, and in Chapter 3, we focused on pointing gestures as a type of deictic gesture. However, other types of gestures can also provide supplementary information not contained in speech (Goldin-Meadow (2000; Goldin-Meadow & Wagner, 2005). Different gestures can play different roles during multimodal word learning; for example, Vigliocco et al. (2019) showed that caregivers use representational gestures to support their children's understanding of novel objects and labels when objects are absent.

More broadly, one avenue for advancing multiple cue models of word learning lies in disambiguating the contribution of other individual cues within an integrative framework in typical language learning. Other cues that have been examined include prosody (Monaghan et al., 2017), eye gaze (MacDonald et al., 2017), iconic and representational gestures (Vigliocco et al., 2019), and grammar (Monaghan et al., 2015). The interpretation of social and non-social cues in cross-situational word learning by populations with autism spectrum disorder has also proved to be insightful (Hartley et al., 2020). In addition, the use of head mounted cameras has offered a unique insight into the perspective of children as learners, and may provide another way of integrating naturalistic studies with cross-situational word learning tasks (Pereira et al., 2014; Yurovsky, Smith, et al., 2013), opening up possibilities for future research to integrate multiple cues under more dynamic and ecologically valid conditions.

Summary of implications for typical development

Combined, Chapters 2 and 3 highlight the interactive nature of how cues can be used by the caregiver, and how they interact with variability to support word learning. They also highlight the use of gesture cues by the learner themselves. Chapter 2 demonstrated that infants themselves appear to be adaptive to variability, learning more robustly when there were two- or six- referents as opposed to one. Chapter 3 revealed that learners make use of early gestures to accurately identify associations between novel word-referent pairs during learning. A key direction for future research thus concerns *why* infants track co-occurrences within statistical learning. Saffran (2020) suggests that this is partly motivated by the need and desire to communicate with caregivers, but argues that the primary incentive is to generate predictions about the environment, as '*learning itself is motivating, and infants are driven to attempt to reduce uncertainty*' (p.4). An alternative explanation is curiosity-driven learning, which proposes that infants are driven to pursue the most novel stimuli for themselves, utilising past experience to shift between novelty-seeking and novelty-reducing strategies accordingly (Twomey & Westermann, 2018).

Future research must therefore uncover not only how multiple cues interact, but also how they are integrated by both caregivers as teachers and by infants as learners, over different timescales. Chapters 2 and 3 documented the dynamic processes that take place during word learning itself. However, subsequent research must also investigate how such processes develop over a much longer timeframe – over days, weeks, and months.

Cross-situational word learning offers a chance to identify how cues, such as deictic gestures, interact with cognitive processes such as hypothesis testing and associative learning. It also provides a framework that allows for statistical regularities in the environment to influence the subsequent mapping of words and referents. However, relating statistical learning studies that utilise artificial languages in controlled and uncluttered laboratory conditions to languages in real world contexts that are messy and contain a great deal of noise and signal, has been a significant challenge for the field (Romberg & Saffran, 2010; L. B. Smith et al., 2014). Although quantifying and understanding the amount of noise – for example, determining the level of uncertainty – can help to identify how statistical learning relates to real world contexts (e.g. Trueswell et al., 2016), L. B. Smith et al. (2014) argue that the separation of input into 'signal' and 'noise' itself is counterproductive, as they may be one and the same. Rather, they advocate for identifying how various sources of apparent environmental noise integrate with sensory, attentional, and memory processes within the learner to map words to referents and build semantic networks.

Multiple cues are of key importance as sources of that input. Models such as Hollich et al. (2000) and Yu and Ballard (2007) provide frameworks for how multiple cues can contribute to word learning. However, identifying how cues interact with each other, with the variability of the environment, and with the robustness of learning itself, is an important barrier to scale for identifying how multiple cues and cross-situational word learning can relate to real world learning. Chapters 2 and 3 offer vital evidence for gesture cues in word learning that help to bridge this gap.

7.3 What does this thesis add to the atypical development literature?

The second part of this thesis investigated whether late talking (LT) children learn words differently to typically developing children, and whether their early expressive delay impacts on their symbolic understanding of pictures.

Gaps in the literature

LT children are at risk of Developmental Language Disorder. Large cohort studies from the UK (Boyd et al., 2013; Clegg et al., 2015; Dale et al., 2003; Hartas, 2011), the Netherlands (Henrichs et al., 2011), Finland (Lyytinen et al., 2005), Australia (Reilly et al., 2007, 2010, 2018), and the USA (Armstrong et al., 2017; Hammer et al., 2017; Horwitz et al.,

2003) have identified a number of consistent demographic predictors of later language delay (e.g. male gender, low socioeconomic status, maternal health, family history), but these explain only a small amount of variance in outcomes and have limited predictive use (Reilly et al., 2010).

Studies examining mechanisms underlying LT that utilise mixed effects models, rather than measures of central tendency, offer a chance to identify whether or not LT children are qualitatively different to TD children in how they learn words. Deficits in grammar (Moyle et al., 2007), speech processing (Fernald & Marchman, 2012), fast mapping (Weismer et al., 2013), and nonword repetition (MacRoy-Higgins & Dalton, 2015) have been found in late talking children, but these studies are relatively few in number compared to those examining DLD. Additionally, using statistical analyses that allow for individual variation is not yet the norm (Perry & Kucker, 2019). Furthermore, although language facilitates and interacts with development (Callaghan & Rankin, 2002; J. Kirkham et al., 2013), the interaction between early expressive delay *and* other developmental domains such as social ability and symbolic understanding has not been investigated previously.

The gaps in the late talking literature were addressed by Chapters 5 and 6 (third and fourth papers) concerning a longitudinal study of LT and TD children. These Chapters examined late talking children's proficiency across different stages of word learning, and the effects of receptive and expressive vocabulary and social ability on picture comprehension, making use of mixed effects models to do so.

Summary of results: Chapter 5 and 6

In Chapter 5, a longitudinal study was presented where a cohort of LT and TD children (identified at 2;0 – 2;5-years –old) were administered tasks that assessed different stages of word learning at 3;0 – 3;5-years-old. Children were tested on two mechanisms of word learning: phonological ability and fast mapping.

The phonological ability task involved immediately repeating real known words and unknown nonwords. Despite all but two LT children reaching typical expressive vocabulary, the LT children showed a significantly impaired ability to repeat both real words and nonwords at 3-years-old as compared to TD children. Concurrent expressive vocabulary also related to task performance.

The cohort were also tested on their ability to accurately select referents for novel words following single exposures to them and retain the novel word-referent mappings after 5 minutes during a fast mapping task. At a group level, LT children did not differ from TD children on either measure. Concurrent expressive vocabulary across the sample related to task performance. Pre-COVID-19, half of the cohort were also tested on a cross-situational word learning task, where children had to accurately select referents for target words by tracking co-occurrences of words and referents across trials. They were then tested on their retention of these words after a 5-minute delay. LT children showed no differences to TD children on referent selection, but did show less accurate performance on retention trials. Across the sample, nonword repetition and fast mapping retention predicted LT status at the first timepoint, and also predicted expressive vocabulary at the last time point. CSWL did not relate to early or later expressive vocabulary.

Finally, in Chapter 6, a longitudinal study was presented involving the same cohort as in Chapter 5. LT and TD children were tested on their picture comprehension at 2;0 – 2;5-years and again at 3;6 – 3;11-years-old. Children were asked to match simple line drawings of an object to one of two 3D referents across four conditions that manipulated the availability of verbal labels and the familiarity of the objects. These conditions included: 1) familiar objects with different labels (e.g. a cat and a rabbit), 2) familiar objects with the same label (e.g. a tabby cat and a calico cat), 3) unfamiliar objects with different labels (e.g. a narwhal and a manatee), and 4) unfamiliar objects with the same label (e.g. elkhorn coral and encrusted coral). This meant that children could use linguistic scaffolding only where labels were different and known. All children were able to use verbal labelling to match pictures of objects to their referents at both age 2;0 – 2;5-years and 3;6 – 3;11-years-old, but LT children showed delayed performance overall when both timepoints were combined. Furthermore, receptive and expressive ability differentially predicted task performance over time; at ~ age 2-years, receptive vocabulary predicted accuracy, whereas at ~age 3;6-years, expressive vocabulary predicted accuracy. The effects of receptive vocabulary on task performance were also mediated by social ability measured at 2;0 – 2;5-years-old.

Implications: late talking: one factor of many in word learning

A somewhat implicit characterisation of LT children is that they represent a separate category of children to those who are typically developing; even where studies highlight heterogeneity, statistical analyses often test for categorical between-group differences. This highlights difficulties in linking a clearly observable factor (i.e., how much children say) with the more nuanced processes that make up language acquisition. As models of typical development have demonstrated, word learning is a complex, multi-stage process (Hollich et al., 2000; McMurray et al., 2012), and as such, multiple things may go wrong with this system.

Our analysis of word learning (Chapter 5) indicated that LT children are impaired in some, but not all, word learning mechanisms. Chapter 5 indicated that LT children show intact receptive word learning mechanisms. However, even once reaching typically developing range, they struggled with phonological and articulatory processes, showing impaired production of familiar and novel words. Our results also suggested that LT children may exhibit broader problems in extracting statistical information from the environment for later retention. Future studies of language delay thus could recruit from across different percentiles and test expressive vocabulary as a continuous predictor from the outset. For the whole sample, the predictive effect of concurrent expressive vocabulary for nonword repetition and fast mapping, and the overall predictive effect of nonword repetition and fast mapping retention on later vocabulary, is also consistent with theories that conceptualise phonological and lexical development as dynamic and interrelated throughout development (e.g. Edwards et al., 2004; Stokes, 2010; 2014) – and suggest that weaker phonological representation may impede novel word learning.

The necessity of using expressive vocabulary to identify children at risk of further problems must be tempered with, not only the recognition that late talking remains a symptom of language delay rather than a diagnosis in itself (Leonard, 2009), but also the appreciation that methods which allow for individual variation and trajectories are likely to provide a better characterisation of this population. For example, our analysis of picture comprehension (Chapter 6) showed that whilst there were some between group differences between LT and TD children, these differences were somewhat mitigated by allowing for random effects of participant when assessing main effects, and that LT children were able to make use of verbal labels to scaffold their understanding of pictures even despite their concurrent expressive language delay. The study also demonstrated how and why expressive vocabulary may interact with social ability and picture comprehension more broadly, indicating that language and symbolic understanding are subject to variation by different social abilities in children. The results also suggested that children may rely on more social scaffolding in the initial stages of symbolic development, then shift to utilising the capacity to talk about pictures as their symbolic development advances over time. Furthermore, as LT children showed a delay in picture comprehension across both timepoints, Chapter 6 indicated that cross-domain relationships can be affected by expressive language delay over time, even if LT children are able to use verbal labels to scaffold picture comprehension. This might indicate that although LT children were not understanding pictures in a qualitatively different way to TD children and were still able to use language, they might have a more general symbolic delay in non-linguistic domains as a result of their delay in linguistic abilities. The results of Chapter 6 therefore demonstrate why it is important to identify other processes that may be affected as a result of language delays, rather than fixating only on vocabulary, and call for more research into cross-domain links between language, socio-cognitive skills, and symbolic understanding in atypical development

Implications: interventions for late talking children

Interventions for LT children can involve enriching the general environment around children, addressing specific problems such as learning particular words, and structured teaching with models and prompts; additionally, it may be clinician or parent based (DeVeney et al., 2017).

Chapter 5 indicated that phonological representation was particularly difficult for LT children even after reaching TD ranges of vocabulary. This may provide a good target for intervention. For example, Buschmann et al. (2015) tested the effectiveness of the Heidelberg Parent-Based Language Intervention on LT children (expressive vocabulary < 1 SD below the age-related mean on SETK-2; Grimm, 2000) recruited at 2-years-old. At 4-years-old, there was no difference between the intervention (*n* = 23) and control group (*n* = 20) on expressive vocabulary, but the intervention group outperformed the control group on phonological memory (nonword repetition, word span, number recall, word order). Over the whole sample, LT children who reached typical ranges for vocabulary, scored significantly better on the phonological memory measures.

In Chapter 5, data from the CSWL task also suggested that LT children may struggle with retention of word-referent pairs acquired through statistical learning mechanisms – although due to the much smaller sample, this must be interpreted very cautiously. This might potentially reduce the effectiveness of interventions based on statistical learning; however, equally, increasing exposure in a supported fashion might give LT children more opportunity to identify multiple sources of input. The broader concept of variable input in language development may also play a protective role in language delay. Collisson et al.

(2016) identified protective factors for LT in an epidemiological cohort study that included book reading, the provision of informal play opportunities, and attendance at childcare centres, theorising that the benefit of these came from the wide range of different contexts and communicative partners. Future intervention studies could potentially make use of variability within randomised controlled trials to potentially promote gains in vocabulary.

In Chapter 6, the mediating factor of social ability on pictorial understanding also indicates a fruitful area for intervention. For example, Robertson and Weismer (1999) tested the effect of a 12-week clinician-implemented programme for LT children at 24 months that used social scaffolding. They found significant gains across socio-communicative and language skills in the intervention (*n* = 11) versus the control group (*n* = 10) after the intervention, as well as a reduction in parental stress. Combined with our results, these results suggest that benefits from a social scaffolding intervention might also extend beyond the linguistic symbolic domain to the non-linguistic. However, robust studies that examine the efficacy of targeting social skills in LT children are few.

In sum, our longitudinal study adds to the evidence base for intervention in LT children. However, given our limited sample, alongside the lack of studies that examine word learning and the lack of studies that examine the interaction of socio-cognitive skills and symbolic development in LT children, further research is necessary. Particularly, broader issues around interventions must be taken into account. In a systematic literature review of interventions for LT children, DeVeney et al. (2017) noted that parent-implemented interventions may be more effective than clinician-implemented interventions, but also identified a lack of studies with robust data reporting and rigorous research design. Only eight studies were included in the review, some of which did not report important baseline information, such as receptive language ability and SES. Similar limitations were noted by Cable and Domsch (2010) in a separate systematic review. Methodologically rigorous studies that examine low intensity parent-implemented interventions report good feasibility and acceptability, but little evidence of improvement in language both immediately and at 3-years-old (Wake et al., 2011). A more recent retrospective study (Kwok et al., 2020) identified gains in vocabulary and general communication following a parent-implemented intervention, but did not have a control group. Subsequently, future research directions include, (1) examining the effectiveness of focusing on potential areas of weakness related to word learning in LT children, such as phonological memory, (2) exploring the possibility of promoting protective factors, such as socio-communicative skills and variability, in language-learning environments, and (3) utilising rigorous analyses and methods, such as randomised controlled trials, where possible.

7.4 Limitations and future directions for atypical development

Several limitations were highlighted within Chapters 5 and 6. However, three key limitations are expanded upon here: (1) the interruption of testing, (2) relevant abilities, such as attention and memory, not being tested in the longitudinal study, and (3) the homogeneity of the LT and TD sample itself limiting generalisation of findings.

The interruption of testing due to the COVID-19 pandemic

A major limitation of the longitudinal study was caused by the COVID-19 pandemic. This occurred during the final timepoint of data collection, and resulted in a reduced sample size, as well as the inability to identify whether all LT children who participated in the first time point recovered by the last. There was a good uptake of participation at the second timepoint that was a full 12 months after the first visit (~ 91%). All participants at this timepoint had agreed to participate again 6 months later, including all LT children, indicating good motivation to continue. Although 75% of our sample was able to continue with at least part of the study remotely, this was largely due to their goodwill and personal circumstances that afforded their participation in the study during a global pandemic. This means our research is limited by our smaller sample size, albeit beyond our immediate control.

The pandemic has forced the field to identify innovative ideas for collecting developmental data. From as early as March 2020, several developmental research labs rapidly began to test the effect of the pandemic on parent-reported CDI data (Kartushina et al., 2021). Others have advocated for increasing online data collection in developmental science (e.g. Lourenco & Tasimi, 2020; Sheskin et al., 2020). The broader implications of the pandemic, however, highlight the subsequent need to focus on speech and language in the early years, due to increased pressure on parents working from home or made redundant, and reduced access to childcare and schooling. For example, Bowyer-Crane et al. (2021) reported an increase in 4 – 5-year-olds starting school in the UK who were perceived to need help with language skills. Of the 58 schools surveyed, 76% of schools reported that children needed more support than previous cohorts, with 96% highlighting communication and language concerns.

The effects of the ongoing pandemic on language development and school readiness will become apparent over time. From a more optimistic standpoint, some efforts to mitigate the effects of the pandemic are underway in the UK, and offer hope for recovery. For example, two-fifths of English primary schools will take part in the Nuffield Early Language Intervention in the upcoming 2021 academic year, which provides specialised individual and small-group language teaching for 4-year-olds starting school (Nuffield Foundation, 2021).

Other processes in the late talking sample

The inability to test the whole cohort on nonverbal IQ as a result of COVID-19 has been mentioned previously. However, we also did not test attention or memory due to time constraints within testing sessions. This means that differences within the sample that relate to these mechanisms were not captured. For example, the nonword repetition test has been considered largely a test of phonological working memory, and research in DLD has identified poorer working memory as being related to language ability (e.g. Jackson et al., 2016, 2019). However, Petruccelli et al. (2012) found that the working memory of 5-year-olds did not differ between recovered LT children and TD children, but children with DLD had lower scores in comparison to both groups. Similarly, D'odorico et al. (2007) found that although LT children had significantly poorer nonword repetition ability, they showed no difference to TD children in attention, impulsivity, or in short term memory. Thus, the mechanisms that underlie DLD may be different to those implicated in LT children. Nevertheless, future studies could potentially test parallel cognitive abilities that are nonverbal, such as the Attention and Memory Battery in the Leiter-3 (Roid et al., 2013), to assess whether concurrent attention and memory deficits in LT children could contribute to outcomes.

The homogeneity of the late talking sample

Our sample of LT and TD children was particularly homogenous in SES and parental education. This had the advantage of excluding the possibility of results being influenced by similar environmental factors which have been highlighted as contributory factors to LT outcomes by cohort studies (e.g. Reilly et al., 2010). However, this does mean that our findings cannot be readily extended to other populations. As evidence suggests that poorer language outcomes are related to lower SES (e.g. Fernald et al., 2013; Hirsh-Pasek et al., 2015; Locke et al., 2002; Tomalski et al., 2013), the reasons for our homogenous sample and potential solutions for future recruitment require consideration.

From a practical perspective, the limitations of conducting a complete longitudinal study within the time, space, and funds allocated to a doctoral project make the barriers to testing a more diverse sample of participants particularly challenging. Engaging participants with more diverse backgrounds proved difficult when recruiting for the longitudinal study in particular. Recruitment took place through flyers and actively promoting the study across local nurseries and pre-schools in Lancashire (Manchester, Lancaster, Preston), the local health visiting service for West Lancashire, and through the Lancaster University Babylab. Despite equal distribution across geographical areas, those who got in contact were consistently from similar, mid-SES backgrounds, had an interest in child development, and were often involved in the National Health Service (NHS) itself. A collaboration with a specific local opportunity area in Derby was attempted to recruit a more diverse sample, but had no uptake by would-be participants.

Recruiting through a national public health service that often serves as a backdrop for fluctuations in society and politics (British Medical Association, 2021) also affected our recruitment. For example, the public NHS health visiting service utilised for recruitment was undergoing a takeover by a private healthcare company following a financial bidding competition with the local NHS Trust (Matthews-King, 2017). This brought significant restructuring and uncertainty over how much capacity health visitors had for additional work, such as recruiting for studies. The service itself was overstretched, with a small team allocated to large parts of the Lancashire county. The national Early Years Foundation Stage progress check is typically recommended at age 2 years, with a designated range of 24 – 36-months-old (Department of Health and NCB, 2015). However, as a result of heavily strained public services, this check often occurred at the upper limit of this range, meaning many households were not eligible to take part according to our inclusion criteria – which was already expanded prior to the study commencing (from 22 – 26-months-old to 24 – 29-months-old), following consultation with the health visiting service. The frequently delayed 2-year-check in the local service used in this study may, in part, also point to practical barriers for potential language screening and intervention before school-age (*The Bercow Report: 10 Years On*, 2018).

Potential solutions to widening participation in word learning studies include incentivising participation, studies of broad, national cohorts, and studies in specific, under-represented populations. These are briefly examined in turn.

Incentivising participation in child studies is somewhat fraught with ethical complications. Smaller forms of reimbursement, such as stickers, books, and reimbursement for travel costs (as in our study), may not be strong enough motivators for all populations to take part, particularly when a long-term study may take up time and resources. More significant financial reimbursement is seen as appropriate by some, whereas others fear exploitation or coercion (for a review, see M. Rice & Broome, 2004; Zutlevics, 2016). More broadly, some report financial incentives in child health initiatives do not show strong evidence for encouraging better health outcomes (Bassani et al., 2013), raising further ethical issues.

The large cohort studies that have examined early language delay described previously (e.g. Clegg et al., 2015; Hartas, 2011) are broader representations of LT children within the UK. However, the practicality of studying word learning mechanisms in large numbers of children is of course, limited. An alternative is to examine specific populations with poorer language outcomes using a wide range of research methods, and borrow from public health initiatives that allow for heterogeneity. Law et al. (2013) argue that language and communication services deal with aberrant outcomes rather than preventative factors, and that reconceptualising speech and language needs as a public health issue, rather than a clinical one, would benefit both prevention and intervention. Using public health principles prior to implementation of programmes, such as examining community readiness and providing integrated community support (Dickerson et al., 2019; Islam et al., 2019), may also help improve diversity of word learning studies. To provide a solid evidence base for populations most in need of an evidence base for expressive language delay, ambitious future projects could potentially embed short word learning tasks within larger scale projects that target speech and language initiatives in diverse populations such as birth cohorts or epidemiological studies. This would ideally require the qualifier that systemic issues that prevent research participation are examined and accounted for prior to study commencement.

For example, LT children's impairments in nonword repetition may be a good candidate for testing in larger cohort studies. One of the problems with traditional nonword repetition measures is the limited ability of very young children to tolerate the task (Roy & Chiat, 2004). Measures of nonword repetition were originally administered to children aged 4-years and above, and require the child to sit and dutifully repeat auditory stimuli from a recording (Gathercole et al., 1994). However, the PSRep Test (Chiat & Roy, 2007) assesses the phonological representation of real words and nonwords using live presentation, has good compliance in 2-year-olds, most children in our study managed to complete the task at ~3-years-old without difficulty, and it takes only a matter of minutes to administer. Administration at 2;6-years-old elsewhere also predicted language outcomes 18 months later in 163 LT children (≤ 15th percentile, CDI) when combined with receptive and expressive vocabulary, morphosyntactic ability, and socio-cognitive skills (Chiat & Roy, 2008). This means that nonword repetition could be a viable option for including alongside epidemiological and demographic factors in large cohort studies – although for children whose expressive output is severely limited or younger than 2, this is less practical.

Fast mapping tasks similarly take a short amount of time to administer, and again, compliance in our sample was good. Weismer (2007) found that a model combining non-verbal IQ, expressive vocabulary, and novel word comprehension at 2;6 years in 40 LT children (2-years-old, ≤10th percentile CDI) provided good sensitivity for predicting TD language skills 12 months later, correctly identifying 90% of recovered LT children, and correctly rejecting 91% of those who continued to show significant impairments. However, only fast mapping retention predicted expressive vocabulary at T3 in our study, indicating that further studies that examine retention and memory abilities, as well as referent selection, are necessary.

Summary of implications for atypical development

This thesis offers evidence not only for testing word learning mechanisms in children with expressive language delay, but also for the reconceptualisation of LT as a consequence of individual variation and a spectrum of language ability. Chapter 5 demonstrates the potential value of identifying candidate word learning mechanisms that provide a better characterisation of LT children, highlighting deficits in phonological representation, and potentially around the retention of statistical associative information. Chapter 6 highlights how language can interact with multiple other non-linguistic domains during child development, and the necessity of characterising these effects in atypical populations.

Of interest is that all of the LT children in our sample, except for two, were above the 30th percentile for expressive language by 3 – 3.4-years-old (T2). Both of the children who showed enduring deficits had additional factors that likely impacted their language skills. One child had intermittent otitis media (repeated ear infections that can lead to conductive hearing impairment) diagnosed between study visits at T2 and T3, which self-resolved without intervention. This child reached above the 30th percentile at the last timepoint (T3). The other child, who remained around the 10th percentile at T2 and at T3, developed an overjet across the latter part of the study (a malocclusion where the upper teeth protrude over the lower, typically not corrected under the age of 12-years-old; National Health Service, 2020). Dental malocclusion can impede speech (e.g. Inukai et al., 2006; Laine, 1992) and poor oral motor skills are related to language difficulties (Alcock, 2006). Thus, this particular child's reduced expressive vocabulary was likely a result of longer-term articulation difficulties, and potentially less practice during the course of language acquisition.

Overall, a multiple hit hypothesis may provide the most reasonable fit to our data. In other words, children in our study did not show prolonged severe expressive delay at time of follow-up unless another aspect of their language was also affected beyond their expressive vocabulary. Bishop (2006) writes that research has not identified distinctive subgroups of

DLD, but rather, that children with a single deficit are less likely to have DLD than those with multiple deficits, thus recommending a dimensional approach to conceptualising DLD. LT children are similarly heterogenous, and cohort studies have shown that variability in data cannot be well accounted for by epidemiological factors alone (Reilly et al., 2010). Our longitudinal study indicates that a dimensional approach is equally sensible for LT as Rescorla (2011) describes, and also highlights how evaluating mechanisms at each stage of word learning, using methods that allow for individual variation, and examining deficits in other domains may better characterise the LT population and their outcomes.

7.5 Conclusions

This thesis has argued that word learning research must be accommodated into a theory that makes use of multiple cues, as well as multiple mechanisms, for language learning. The first part of this thesis identified how and when gesture cues support word learning within TD populations as part of a multiple cue model. Chapter 2 demonstrated that caregiver use gestures based on the presence of referential ambiguity to help teach their infants new words, and that infants were able to learn more accurately in conditions with referential ambiguity. Chapter 3 demonstrated that deictic gestures not only direct attention as word learning unfolds in real-time, but that the temporal dynamics of how gesture and speech co-occur directly affect the accuracy of word learning. The results showed more accurate word-referent mappings with early, rather than late gestures, and more accurate mappings with late, rather than no gestures. Together, these findings demonstrate how two sources of information – gesture cues and environmental variability – can affect language learning, but also show how they may interact with other processes during word learning, such as attentional cueing.

The second part of this thesis identified how word learning mechanisms might be affected in atypical development, and also identified how subsequent effects of expressive vocabulary delay might affect non-linguistic domains. Chapter 5 highlighted how different impaired mechanisms of word learning can impact vocabulary acquisition in LT children and vice versa. LT children showed impairments across expressive, but not receptive, domains – but also showed potential impairments in retention of statistical information. This showed that as word learning comprises multiple processes, impairment in one does not necessarily mean impairment in another. Chapter 6 identified that LT children show delayed, but not functionally different, picture comprehension as compared to TD children, and also that receptive and expressive vocabulary differentially support picture comprehension over time. This demonstrated how a broader perspective of symbolic development is necessary to enrich and effectively position theories of language acquisition related to how children come to understand the world around them.

It is possible that the wider sphere of research has historically encouraged a divisive approach to understanding scientific concepts (Kaiser, 2012; T. Kuhn, 1970). However, given the wealth of information that different models of word learning have brought to the field of language acquisition, it seems unwise to pit one against the other. Each model – linguistic constraints (Markman, 1989), socio-pragmatic theory (Tomasello, 2003), and cross-situational word learning (Smith & Yu, 2008; Yu & Smith, 2007) – has contributed meaningfully and substantially to advancing language acquisition theory. This was recognised twenty years ago by integrative models, such as the Emergent Coalition Model (combining socio-pragmatic cues and lexical constraints; Hollich et al., 2000), then again some years later in Yu and Ballard's (2007) unified model (combining socio-pragmatic cues and statistical learning), and more recently, by the MIM (combining linguistic input, socio-pragmatic cues and statistical learning; Monaghan, 2017). Models that examine cross-situational word learning (McMurray et al., 2012; Yurovsky & Frank, 2015) also provide a framework for different stages of word learning during development that can integrate lexical constraints with statistical learning.

Now may be a particularly fitting time to continue the advancement of integrative models of word learning due to two relatively recent developments: firstly, the advances in computing technology that allow us to not only model multiple inputs on an outcome simultaneously, but also allow for a more dynamic interaction of these via the use of machine learning (L. B. Smith & Slone, 2017); and secondly, a degree of openness around data sharing and multi-lab collaborations afforded by the mass connectivity of the Internet and by Open Science initiatives (Munafò et al., 2017). Both of these developments recognise the contribution of, and provide the capacity for, multi-disciplinary approaches to research that make use of multiple theories of word learning.

Part of understanding integrative theories of word learning comprises the need to identify what happens in atypical development. Although LT children offer a unique opportunity to study the effects of expressive language delay, whether or not they are qualitatively different to TD children requires further investigation through the use of word learning mechanisms and synergistic developmental domains, such as symbolic understanding and socio-cognitive skills. Overall, expressive language delay may be succinctly summarised by Bishop's (2006) description of more general language delay, *'… there may be multiple routes to effective language acquisition, and if one route is blocked, another can usually be found. However, if two or more routes are blocked, then language learning will be compromised'* (p. 220).

More broadly, research can no longer exist in a vacuum. With the advent of newer and more technologically advanced methods of communication, and a greater awareness of social responsibility than ever before, research must appeal to greater numbers and must represent all populations. This is particularly important when studying language delay and identifying children at risk of further problems. However, this is only attainable through good communication outside of the academic community, and broadening testing beyond specific populations. For example, a consistent feature in the longitudinal study was that no clear standard for LT children was known to those outside of the academic community, and that referrals to speech and language services were not always made consistently. Addressing language delay in a public health context, as Law et al. (2013) describe, may not only improve prevention, detection, and intervention for early language delay, but may also promote better communication between researchers, practitioners, and the communities that they serve. Raising awareness more generally in the wider community may also be viable supplement to providing intensive programmes whilst services are already stretched.

Consolidated Bibliography

Ahufinger, N., Guerra, E., Ferinu, L., Andreu, L., & Sanz-Torrent, M. (2021). Cross-situational statistical learning in children with developmental language disorder. *Language, Cognition and Neuroscience*, *0*(0), 1–21. https://doi.org/10.1080/23273798.2021.1922723

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, *67*(2), 635–645. https://doi.org/10.2307/1131837

Akhtar, N., Dunham, F., & Dunham, P. J. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of Child Language*, *18*(1), 41–49. https://doi.org/10.1017/S0305000900013283

Alcock, K. J., Meints, K., & Rowland, C. F. (2017). *UK-CDI Words and Gestures: Preliminary Norms and Manual*. http://lucid.ac.uk/ukcdi

Alcock. (2006). The development of oral motor control and language. *Down Syndrome Research and Practice*, *11*(1), 1–8. https://doi.org/10.3104/reports.310

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

Ankowski, A. A., Vlach, H. A., & Sandhofer, C. M. (2013). Comparison versus contrast: Task specifics affect category acquisition. *Infant and Child Development*, *22*(1), 1–23. https://doi.org/10.1002/icd.1764

Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*, *35*(6), 1105–1138. https://doi.org/10.1111/j.1551-6709.2011.01181.x

Armstrong, R., Scott, J. G., Whitehouse, A. J. O., Copland, D. A., Mcmahon, K. L., & Arnott, W. (2017). Late talkers and later language outcomes: Predicting the different language trajectories. *International Journal of Speech-Language Pathology*, *19*(3), 237–250. https://doi.org/10.1080/17549507.2017.1296191

Axelsson, E. L., Churchley, K., & Horst, J. S. (2012). The right thing at the right time: Why ostensive naming facilitates word learning. *Frontiers in Psychology*, *3*, 1–8. https://doi.org/10.3389/fpsyg.2012.00088

Bahtiyar, S., & Küntay, A. C. (2009). Integration of communicative partner's visual perspective in patterns of referential requests. *Journal of Child Language*, *36*(3), 529–555. https://doi.org/10.1017/S0305000908009094

Baldwin, D. A., & Tomasello, M. (1998). Word learning: A window on early pragmatic understanding. In *The Proceedings of the 29th Annual Child Language Research Forum* (pp. 3–23).

Bannard, C., Rosner, M., & Matthews, D. (2017). What's worth talking about? Information Theory reveals how children balance informativeness and ease of production. *Psychological Science*, *28*(7), 954–966. https://doi.org/10.1177/0956797617699848

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). https://doi.org/10.1016/j.jml.2012.11.001

Barrett, M. D. (1978). Lexical development and overextension in child language*. *Journal of Child Language*, *5*(2), 205–219. https://doi.org/10.1017/S030500090000742X

Bassani, D. G., Arora, P., Wazny, K., Gaffey, M. F., Lenters, L., & Bhutta, Z. A. (2013). Financial incentives and coverage of child health interventions: A systematic review and meta-analysis. *BMC Public Health*, *13*(3), S30. https://doi.org/10.1186/1471-2458-13-S3-S30

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

Berger, A., Henik, A., & Rafal, R. (2005). Competition between endogenous and exogenous orienting of visual attention. *Journal of Experimental Psychology: General, 134*(2), 207 – 221. https://doi.org/10.1037/0096-3445.134.2.207

Bergmann, K., Aksu, V., & Kopp, S. (2011). The Relation of speech and gestures: Temporal synchrony follows semantic synchrony. *Proceedings of the 2nd Workshop on Gesture and Speech in Interaction (GeSpIn 2011)*. https://pub.uni-bielefeld.de/record/2392953

Bion, R. A. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, *126*(1), 39–53. https://doi.org/10.1016/j.cognition.2012.08.008

Bishop, D. V. M. (2006). What Causes Specific Language Impairment in Children? *Current Directions in Psychological Science*, *15*(5), 217–221. https://doi.org/10.1111/j.1467-8721.2006.00439.x

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

Bishop, D. V. M., Price, T. S., Dale, P. S., & Plomin, R. (2003). Outcomes of Early Language Delay: II. Etiology of Transient and Persistent Language Difficulties. *Journal of Speech, Language & Hearing Research*, *46*(3), 561–565.

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Bowyer-Crane, C., Bonetti, S., Compton, S., Nielsen, D., D'Apice, K., & Tracey, L. (2021). *The impact of Covid-19 on School Starters: Interim briefing 1Parent and school concerns about children starting school*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Impact_of_Covid19_on_School_Starters_-_Interim_Briefing_1_-_April_2021_-_Final.pdf

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S., & Davey Smith, G. (2013). Cohort Profile: The 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*, *42*(1), 111–127. https://doi.org/10.1093/ije/dys064

Brignani, D., Guzzon, D., Marzi, C. A., & Miniussi, C. (2009). Attentional orienting induced by arrows and eye-gaze compared with an endogenous cue. *Neuropsychologia*, *47*(2), 370–381. https://doi.org/10.1016/j.neuropsychologia.2008.09.011

British Medical Association. (2021). *Written evidence submitted by the BMA (HSC0873)*. British Medical Association. https://committees.parliament.uk/writtenevidence/24974/pdf/

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(1), 207–220. https://doi.org/10.1017/s030500090700829x

Brown-Schmidt, S., & Duff, M. C. (2016). Memory and common ground processes in language use. *Topics in Cognitive Science*, *8*(4), 722–736. https://doi.org/10.1111/tops.12224

Bunce, J. P., & Scott, R. M. (2017). Finding meaning in a noisy world: Exploring the effects of referential ambiguity and competition on 2·5-year-olds' cross-situational word learning. *Journal of Child Language*, *44*(3), 650–676. https://doi.org/10.1017/S0305000916000180

Buschmann, A., Multhauf, B., Hasselhorn, M., & Pietz, J. (2015). Long-Term Effects of a Parent-Based Language Intervention on Language Outcomes and Working Memory for Late-Talking Toddlers. *Journal of Early Intervention*, *37*(3), 175–189. https://doi.org/10.1177/1053815115609384

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 L. S. Liben, U. Müller, & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Cognitive processes* (pp. 250–295). Hoboken, NJ: John Wiley & Sons, Inc.

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.

Callanan, M. A., & Sabbagh, M. A. (2004). Multiple labels for objects in conversations with young children: Parents’ language and children’s developing expectations about word meanings. *Developmental Psychology*, *40*(5), 746–763. https://doi.org/10.1037/0012-1649.40.5.746

Capone, N. C., & McGregor, K. K. (2004). Gesture development: A review for clinical and research practices. *Journal of Speech, Language, and Hearing Research*, 173186.

Carey, S. (1988). Conceptual differences between children and adults. *Mind & Language*, *3*(3), 167–181. https://doi.org/10.1111/j.1468-0017.1988.tb00141.x

Carey, S., & Bartlett, E. (1978). Acquiring a Single New Word. *Papers and Reports on Child Language Development*, *15*, 17–29.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4), i–174. https://doi.org/10.2307/1166214

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Casby, M. W. (1997) Symbolic play of children with language impairment: a critical review. *Journal of Speech, Language, and Hearing Research*, 40(3), 468-479

Chiat, S., & Roy, P. (2007). The Preschool Repetition Test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research*, *50*(2), 429–443. https://doi.org/10.1044/1092-4388(2007/030)

Chiat, S., & Roy, P. (2008). Early phonological and sociocognitive skills as predictors of later language and social communication outcomes. *Journal of Child Psychology and Psychiatry*, *49*(6), 635–645. https://doi.org/10.1111/j.1469-7610.2008.01881.x

Clark, E. V. (1983). Convention and Contrast in Acquiring the Lexicon. In T. B. Seiler & W. Wannenmacher (Eds.), *Concept Development and the Development of Word Meaning* (pp. 67–89). New York, NY: Springer. https://doi.org/10.1007/978-3-642-69000-6_5

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In *Mechanisms of language aquisition* (pp. 1–33). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cleave, P. L., & Bird, E. K.-R. (2006). Effects of familiarity on mothers' talk about nouns and verbs. *Journal of Child Language*, *33*(3), 661–676. https://doi.org/10.1017/S0305000906007549

Clegg, J., Law, J., Rush, R., Peters, T. J., & Roulstone, S. (2015). The contribution of early language development to children's emotional and behavioural functioning at 6 years: An analysis of data from the Children in Focus sample from the ALSPAC birth cohort. *Journal of Child Psychology and Psychiatry*, *56*(1), 67–75. https://doi.org/10.1111/jcpp.12281

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders*, *43*(1), 1–40. https://doi.org/10.1080/13682820601116485

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)*. Los Angeles, CA: 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

Conyers, L. M., Reynolds, A. J, & Ou, S. R. (2003) The effect of early childhood intervention and subsequent special education services: Findings from the Chicago child-parent center. *Educational Evaluation and Policy Analysis*, 25(1), 75–95

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

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

D'odorico, L., Assanelli, A., Franco, F., & Jacob, V. (2007). A follow-up study on Italian late talkers: Development of language, short-term memory, phonological awareness, impulsiveness, and attention. *Applied Psycholinguistics*, *28*(1), 157–169. https://doi.org/10.1017/S0142716406070081

Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research*, *46*(3), 544–560.

de Diego-Balaguer, R., Martinez, A., & Pons, F. (2016) Temporal attention as a scaffold for language development. *Frontiers in Psychology, 7(44),* 1–15, https://doi.org/ 10.3389/fpsyg.2016.00044

Deák, G. O. (2000). Hunting the fox of word learning: Why “constraints” fail to capture it. *Developmental Review*, *20*(1), 29–80. https://doi.org/10.1006/drev.1999.0494

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

Department of Health and NCB. (2015). *The Integrated Review: Bringing together health and early education reviews at age two to two-and-a-half*. National Childen’s Bureau Early Childhood Unit. https://www.foundationyears.org.uk/files/2015/03/IR_Supporting_Material.pdf

Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *International Journal of Language & Communication Disorders*, *43*(4), 361–389. https://doi.org/10.1080/13682820701546854

DeVeney, S. L., Hagaman, J. L., & Bjornsen, A. L. (2017). Parent-Implemented Versus Clinician-Directed Interventions for Late-Talking Toddlers: A Systematic Review of the

Literature. *Communication Disorders Quarterly*, *39*(1), 293–302. https://doi.org/10.1177/1525740117705116

Dickerson, J., Bird, P. K., Bryant, M., Dharni, N., Bridges, S., Willan, K., Ahern, S., Dunn, A., Nielsen, D., Uphoff, E. P., Bywater, T., Bowyer-Crane, C., Sahota, P., Small, N., Howell, M., Thornton, G., Pickett, K. E., McEachan, R. R. C., Wright, J., … the Better Start Bradford Innovation Hub. (2019). Integrating research and system-wide practice in public health: Lessons learnt from Better Start Bradford. *BMC Public Health*, *19*(1), 260. https://doi.org/10.1186/s12889-019-6554-2

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

Edwards, J., Beckman, M. E., & Munson, B. (2004). The Interaction Between Vocabulary Size and Phonotactic Probability Effects on Children's Production Accuracy and Fluency in Nonword Repetition. *Journal of Speech, Language, and Hearing Research*, *47*(2), 421–436. https://doi.org/10.1044/1092-4388(2004/034)

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

Ellis, E. M., Borovsky, A., Elman, J. L., & Evans, J. L. (2015). Novel word learning: An eye-tracking study. Are 18-month-old late talkers really different from their typical peers? *Journal of Communication Disorders*, *58*, 143–157. https://doi.org/10.1016/j.jcomdis.2015.06.011

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

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), 1–185.

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

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically-developing and late-talking toddlers. *Child Development*, *83*(1), 203–222. https://doi.org/10.1111/j.1467-8624.2011.01692.x

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*(2), 209–221. https://doi.org/10.1037/0012-1649.27.2.209

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248. https://doi.org/10.1111/desc.12019

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

Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science*, *35*(2), 367–380. https://doi.org/10.1111/j.1551-6709.2010.01156.x

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development*, *9*(1), 1–24. https://doi.org/10.1080/15475441.2012.707101

Friedrich, M., & Friederici, A. D. (2011). Word learning in 6-month-olds: Fast encoding-weak retention. *Journal of Cognitive Neuroscience*, *23*(11), 3228–3240. https://doi.org/10.1162/jocn_a_00002

Fries, C. C. (1952). *The Structure of English*. Harlow, UK: Longmans.

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

Gathercole, S. E., Willis, C. S., Baddeley, A., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory*, *2*(2), 103–127. https://doi.org/10.1080/09658219408258940

Gillette, J., Gleitman, H. Gleitman, L., & Lederer, A. (1999) Human simulations of vocabulary learning. *Cognition, 73(2),* 135 – 176, https://doi.org/ 10.1016/S0010-0277(99)00036-0

Goldin-Meadow, S. (2000). Beyond words: The importance of gesture to researchers and learners. *Child Development*, *71*(1), 231–239. https://doi.org/10.1111/1467-8624.00138

Goldin-Meadow, S. (2007). Pointing sets the stage for learning language-and creating language. *Child Development*, *78*(3), 741–745. https://doi.org/10.1111/j.1467-8624.2007.01029.x

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Sciences*, *9*(5). https://doi.org/10.1016/j.tics.2005.03.006

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, 99–108. https://doi.org/10.1037/0012-1649.28.1.99

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Development*, *90*(3), 985–992. https://doi.org/10.1111/cdev.13128

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, *21*(1), 125–155. https://doi.org/10.1017/S0305000900008692

Goodwyn, S. W., Acredolo, L. P., & Brown, C. A. (2000). Impact of symbolic gesturing on early language development. *Journal of Nonverbal Behavior*, *24*(2), 81–103. https://doi.org/10.1023/A:1006653828895

Gorman, K. S., Gegg-Harrison, W., Marsh, C. R., & Tanenhaus, M. K. (2013). What's learned together stays together: Speakers' choice of referring expression reflects shared experience. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(3). https://doi.org/10.1037/a0029467

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. https://doi.org/10.1037/11867-003

Gray, S. (2004). Word learning by preschoolers with Specific Language Impairment: Predictors and poor learners. *Journal of Speech, Language & Hearing Research*, *47*(5), 1117–1132. https://doi.org/10.1044/1092-4388(2004/083)

Gray, S. (2006). The relationship between phonological memory, veceptive Vocabulary, and fast mapping in young children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *49*(5), 955–969. https://doi.org/10.1044/1092-4388(2006/069)

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279. https://doi.org/10.1111/1467-9280.00255

Grimm, H. (2000). *SETK-2—Sprachentwicklungstest für zweijährige Kinder (2;0–2;11 Jahre)*. Hogrefe. https://www.testzentrale.de/shop/sprachentwicklungstest-fuer-zweijaehrige-kinder-2-0-2-11-jahre.html

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*(1), B23-34. https://doi.org/10.1016/s0010-0277(02)00186-5

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, *53*(4), 310–344. https://doi.org/10.1016/j.cogpsych.2006.04.003

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

Hammer, C. S., Morgan, P., Farkas, G., Hillemeier, M., Bitetti, D., & Maczuga, S. (2017). Late Talkers: A Population-Based Study of Risk Factors and School Readiness Consequences. *Journal of Speech, Language, and Hearing Research*, *60*(3), 607–626. https://doi.org/10.1044/2016_JSLHR-L-15-0417

Hart, B. (1991). Input frequency and children's first words. *First Language*, *11*(32), 289–300. https://doi.org/10.1177/014272379101103205

Hartas, D. (2011). Families' social backgrounds matter: Socio-economic factors, home learning and young children's language, literacy and social outcomes. *British Educational Research Journal*, *37*(6), 893–914. https://doi.org/10.1080/01411926.2010.506945

Hartley, C., & Allen, M. (2014) Brief report: generalisation of word-picture relations in children with autism and typically developing children. *Journal Of Autism And Developmental Disorders*, 44(8), 2064-2071

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., Bird, L.-A., & Monaghan, P. (2019). Investigating the relationship between fast mapping, retention, and generalisation of words in children with autism spectrum disorder and typical development. *Cognition*, *187*, 126–138. https://doi.org/10.1016/j.cognition.2019.03.001

Hartley, C., Bird, L.-A., & Monaghan, P. (2020). Comparing cross-situational word learning, retention, and generalisation in children with autism and typical development. *Cognition*, *200*, 104265. https://doi.org/10.1016/j.cognition.2020.104265

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

Hauer, B. J. A., & Macloed, C. M. (2006) Endogenous versus exogenous attentional cuing effects on memory. *Acta Psychologica, 122(3),* 305–320, https://doi.org/ 10.1016/j.actpsy.2005.12.008

Hebb, H. O. (1949). *The organization of behavior: A neuropsychological theory*. Wiley.

Heisler, L., & Goffman, L. (2016). The influence of phonotactic probability and neighborhood density on children's production of newly learned words. *Language Learning and Development : The Official Journal of the Society for Language Development*, *12*(3), 338–356. https://doi.org/10.1080/15475441.2015.1117977

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)

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust, P. K. S., & Suma, K. (2015). The Contribution of Early Communication Quality to Low-Income Children's Language Success. *Psychological Science*, *26*(7), 1071–1083. https://doi.org/10.1177/0956797615581493

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*(1), 1–27. https://doi.org/10.1017/S0305000910000759

Holler, J., & Levinson, S. (2019) Multimodal language processing in human communication. *Trends in Cognitive Sciences, 23*(8), 639–652.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., Hennon, E., Rocroi, C., & Bloom, L. (2000). Breaking the language barrier: An Emergentist Coalition Model for the origins of word learning. *Monographs of the Society for Research in Child Development*, *65*(3), i–135..

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

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157. https://doi.org/10.1080/15250000701795598

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

Inukai, S., Hideshima, M., Sato, M., Nishiyama, A., Ando, T., Ohyama, T., & Matsuura, H. (2006). Analysis of the Relationship between the Incisal Overjet in a Maxillary Denture and Phonetic Function Using a Speech Recognition System. *Prosthodontic Research & Practice*, *5*(3), 171–177. https://doi.org/10.2186/prp.5.171

Irwin, J. R., Carter, A. S. Briggs-Gowan, M. (2002) The social-emotional development of "late-talking" toddlers. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41(11), 1324 – 1332.

Islam, S., Small, N., Bryant, M., Bridges, S., Hancock, N., & Dickerson, J. (2019). Assessing community readiness for early intervention programmes to promote social and emotional health in children. *Health Expectations*, *22*(3), 575–584. https://doi.org/10.1111/hex.12887

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother-child interactions. *Cognitive Development*, *14*(1), 57–75. https://doi.org/10.1016/S0885-2014(99)80018-5

Jackson, E., Leitao, S., & Claessen, M. (2016). The relationship between phonological short-term memory, receptive vocabulary, and fast mapping in children with specific language impairment. *International Journal of Language & Communication Disorders*, *51*(1), 61–73. https://doi.org/10.1111/1460-6984.12185

Jackson, E., Leitao, S., Claessen, M., & Boyes, M. (2019). Fast mapping short and long words: Examining the influence of phonological short-term memory and receptive vocabulary in children with developmental language disorder. *Journal of Communication Disorders*, *79*, 11–23. https://doi.org/10.1016/j.jcomdis.2019.02.001

Johnston, J. C., Durieux-Smith, A., & Bloom, K. (2005). Teaching gestural signs to infants to advance child development: A review of the evidence. *First Language*, *25*(2), 235–251. https://doi.org/10.1177/0142723705050340

Jonides, J. (1981) Towards a model of the mind's eyes' movement. *Canadian Journal of Psychology, 34(2)*, 103 – 112, https://doi.org/ 10.1037/h0081031

Kachergis, G., Yu, C., & Shiffrin, R. M. (2014). Cross-situational word learning is both implicit and strategic. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00588

Kaiser, D. (2012). In retrospect: The Structure of Scientific Revolutions. *Nature*, *484*(7393), 164–165. https://doi.org/10.1038/484164a

Kan, P. F., & Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *53*(3), 739–756. https://doi.org/10.1044/1092-4388(2009/08-0248)

Kartushina, N., Mani, N., Aktan-Erciyes, A., Alaslani, K., Aldrich, N. J., Almohammadi, A., Alroqi, H., Anderson, L., Andonova, E., Aussems, S., Babineau, M., Barokova, M., Bergmann, C., Cashon, C. H., Custode, S., Carvalho, A. de, Dimitrova, N., Dynak, A., Farah, R., … Mayor, J. (2021). *COVID-19 first lockdown as a unique window into* *language acquisition: What you do (with your child) matters.* PsyArXiv. https://doi.org/10.31234/osf.io/5ejwu

Khoe, Y. H., Perfors, A., & Hendrickson, A. (2019). Modeling individual performance in cross-situational word learning. *Proceedings of the 41st Annual Conference of the Cognitive Science Society.*, 560–566. https://doi.org/10.31234/osf.io/4rtw9

Kidd, E., & Holler, J. (2009). Children's use of gesture to resolve lexical ambiguity. *Developmental Science*, *12*(6), 903–913. https://doi.org/10.1111/j.1467-7687.2009.00830.x

Kirk, E., Howlett, N., Pine, K. J., & Fletcher, B. C. (2013). To sign or not to sign? : The impact of encouraging infants to gesture on infant language and maternal mind-mindedness. *Child Development*, 574–590. https://doi.10.1111/j.1467-8624.2012.01874.x

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

Kirkham, N., Rea, M., Osborne, T., White, H., & Mareschal, D. (2019). Do cues from multiple modalities support quicker learning in primary schoolchildren? *Developmental Psychology, 55*(10), 2048 – 2059. https://doi.org/10.1037/dev0000778

Kuhn, G., & Kingstone, A. (2009). Look away! Eyes and arrows engage oculomotor responses automatically. *Attention, Perception, & Psychophysics*, *71*(2), 314–327. https://doi.org/10.3758/APP.71.2.314

Kuhn, L. J., Willoughby, M. T., Wilbourn, M. P., Vernon-Feagans, L., & Blair, C. B. (2014). Early communicative gestures prospectively predict language development and executive function in early childhood. *Child Development*, *85*(5), 1898–1914. https://doi.org/10.1111/cdev.12249

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. University of Chicago Press.

Kwok, E. Y. L., Cunningham, B. J., & Cardy, J. O. (2020). Effectiveness of a parent-implemented language intervention for late-to-talk children: A real-world retrospective clinical chart review. *International Journal of Speech-Language Pathology*, *22*(1), 48–58. https://doi.org/10.1080/17549507.2019.1584643

Laine, T. (1992). Malocclusion traits and articulatory components of speech. *European Journal of Orthodontics*, *14*(4), 302–309. https://doi.org/10.1093/ejo/14.4.302

Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the Communicative Development Inventories. *Child and Adolescent Mental Health*, *13*(4), 198–206. https://doi.org/10.1111/j.1475-3588.2008.00503.x

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: Findings from a systematic review of the literature. *International Journal of Language & Communication Disorders*, *35*(2), 165–188. https://doi.org/10.1080/136828200247133

Law, J., Reilly, S., & Snow, P. C. (2013). Child speech, language and communication need re-examined in a public health context: A new direction for the speech and language therapy profession. *International Journal of Language & Communication Disorders*, *48*(5), 486–496. https://doi.org/10.1111/1460-6984.12027

Law, J., Robert Rush, Schoon, I., & Parsons, S. (2009). Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes. *Journal of Speech, Language, and Hearing Research*, *52*(6), 1401–1416. https://doi.org/10.1044/1092-4388(2009/08-0142)

Le Corre, M., & Carey, S. (2007). One, Two, Three, Four, Nothing More: An Investigation of the Conceptual Sources of the Verbal Counting Principles. *Cognition*, *105*(2). https://doi.org/10.1016/j.cognition.2006.10.005

LeBarton, E. S., Goldin-Meadow, S., & Raudenbush, S. (2015). Experimentally-induced increases in early gesture lead to increases in spoken vocabulary. *Journal of Cognition and Development*, *16*(2), 199–220. https://doi.org/10.1080/15248372.2013.858041

Leonard, L. B. (2009). Is expressive language disorder an accurate diagnostic category? *American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association*, *18*(2), 115–123. https://doi.org/10.1044/1058-0360(2008/08-0064)

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). Cambridge, MA: MIT Press.

Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science*, *12*(2), 264–271. https://doi.org/10.1111/j.1467-7687.2008.00758.x

Liszkowski, U., & Tomasello, M. (2011). Individual differences in social, cognitive, and morphological aspects of infant pointing. *Cognitive Development*, *26*(1), 16–29. https://doi.org/10.1016/j.cogdev.2010.10.001

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & Vos, C. de. (2012). A prelinguistic gestural universal of human communication. *Cognitive Science*, *36*(4), 698–713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Locke, A., Ginsborg, J., & Peers, I. (2002). Development and disadvantage: Implications for the early years and beyond. *International Journal of Language & Communication Disorders*, *37*(1), 3–15. https://doi.org/10.1080/13682820110089911

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

Lourenco, S. F., & Tasimi, A. (2020). No Participant Left Behind: Conducting Science During COVID-19. *Trends in Cognitive Sciences*, *24*(8), 583–584. https://doi.org/10.1016/j.tics.2020.05.003

Luce, M. R., & Callanan, M. A. (2010). Parents' object labeling: Possible links to conventionality of word meaning? *First Language*, *30*(3–4), 270–286. https://doi.org/10.1177/0142723710370543

Luyster, R., & Lord, C. (2009). Word learning in children with autism spectrum disorders. *Developmental Psychology*, *45*(6), 1774–1786. https://doi.org/10.1037/a0016223

Lyytinen, P., Eklund, K., & Lyytinen, H. (2005). Language development and literacy skills in late-talking toddlers with and without familial risk for dyslexia. *Annals of Dyslexia*, *55*(2), 166–192. https://doi.org/10.1007/s11881-005-0010-y

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)

MacDonald, K., Yurovsky, D., & Frank, M. C. (2017). Social cues modulate the representations underlying cross-situational learning. *Cognitive Psychology*, *94*, 67–84. https://doi.org/10.1016/j.cogpsych.2017.02.003

MacRoy-Higgins, M., & Dalton, K. P. (2015). The influence of phonotactic probability on nonword repetition and fast mapping in 3-year-olds with a history of expressive language delay. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1773–1779. https://doi.org/10.1044/2015_JSLHR-L-15-0079

MacRoy-Higgins, M., Schwartz, R. G., Shafer, V. L., & Marton, K. (2013). Influence of phonotactic probability/neighbourhood density on lexical learning in late talkers. *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, *48*(2), 188–199. https://doi.org/10.1111/j.1460-6984.2012.00198.x

MacWhinney, B. (2005). Extending the competition model. *International Journal of Bilingualism*, *9*(1), 69–84. https://doi.org/10.1177/13670069050090010501

Marini, A., Ruffino, M., Sali, M. E., & Molteni, M. (2017). The role of phonological working memory and environmental factors in lexical development in Italian-speaking late talkers: A one-year follow-up study. *Journal of Speech, Language and Hearing Research*, *60*(12), 3462–3473. https://doi.org/10.1044/2017_JSLHR-L-15-0415

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

Markman, E. M., & Wachtel, G. F. (1988). Children’s use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology*, *20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241–275.

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.

Masur, E. F. (1997). Maternal labelling of novel and familiar objects: Implications for children’s development of lexical constraints. *Journal of Child Language*, *24*(2), 427–439. https://doi.org/10.1017/S0305000997003115

Matthews-King, A. (2017, December 13). Virgin’s £100m children’s health services contract signals ‘galloping privatisation’ of NHS, warn MPs. *The Independent*. https://www.independent.co.uk/news/health/virgin-children-health-service-contract-nhs-privatisation-mps-warn-drugs-healthcare-lancashire-richard-branson-a8105606.html

McGregor, K. K., Rohlfing, K. J., Bean, A., & Marschner, E. (2009). Gesture as a support for word learning: The case of under. *Journal of Child Language*, *36*(4), 807–828. https://doi.org/10.1017/S0305000908009173

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

McNeil, N. M., Alibali, M. W., & Evans, J. L. (2000). The role of gesture in children's comprehension of spoken language: Now they need it, now they don't. *Journal of Nonverbal Behavior*, *24*(2), 131–150. https://doi.org/10.1023/A:1006657929803

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, *108*(22), 9014–9019. https://doi.org/10.1073/pnas.1105040108

Merriman, W. E., Bowman, L. L., & MacWhinney, B. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, *54*(3/4), i–129. https://doi.org/10.2307/1166130

Mervis, C. B. (1987). Child-basic object categories and early lexical development. In *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 201–233). Cambridge, UK: Cambridge University Press.

Mervis, C. B., & Bertrand, J. (1994). Acquisition of the Novel Name-Nameless Category (N3C) Principle. *Child Development*, *65*(6), 1646–1662. https://doi.org/10.2307/1131285

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*(1), 91–117. https://doi.org/0.1016/s0010-0277(03)00140-9

Mirak, J., & Rescorla, L. (1998). Phonetic skills and vocabulary size in late talkers: Concurrent and predictive relationships. *Applied Psycholinguistics*, *19*(1), 1–17. https://doi.org/10.1017/S0142716400010559

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton, FL: CRC Press: Taylor & Francis Group, LLC.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006

Monaghan, P. (2017). Canalization of language structure from environmental constraints: A computational model of word learning from multiple cues. *Topics in Cognitive Science*, *9*(1), 21–34. https://doi.org/10.1111/tops.12239

Monaghan, P., & Mattock, K. (2012). Integrating constraints for learning word–referent mappings. *Cognition*, *123*(1), 133–143. https://doi.org/10.1016/j.cognition.2011.12.010

Monaghan, P., Brand, J., Frost, R. L. A., & Taylor, G. (2017). Multiple variable cues in the environment promote accurate and robust word learning. *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, 817–822.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, *55*(4), 259–305. https://doi.org/10.1016/j.cogpsych.2006.12.001

Monaghan, P., Mattock, K., Davies, R. A. I., & Smith, A. C. (2015). Gavagai is as gavagai does: Learning nouns and verbs from cross-situational statistics. *Cognitive Science*, *39*(5), 1099–1112. https://doi.org/10.1111/cogs.12186

Monaghan, P., Ruiz, S., & Rebuschat, P. (2021). The role of feedback and instruction on the cross-situational learning of vocabulary and morphosyntax: Mixed effects models reveal local and global effects on acquisition. *Second Language Research*, 37(2), 261 - 289. https://doi.org/10.1177/0267658320927741

Moyle, M. J., Weismer, S. E., Evans, J. L., & Lindstrom, M. J. (2007). Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *Journal of Speech, Language, and Hearing Research*, *50*(2), 508–528. https://doi.org/10.1044/1092-4388(2007/035)

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 1–9. https://doi.org/10.1038/s41562-016-0021

Munson, B., Edwards, J., & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research*, *48*(1), 61–78. https://doi.org/10.1044/1092-4388(2005/006)

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336. https://doi.org/10.1111/j.0956-7976.2002.00460.x

National Health Service. (2020, January 27). *Overview: Orthodontics*. https://www.nhs.uk/conditions/orthodontics/

Neam, S. Y., Baker, E., Hodges, R., & Munro, N. (2020). Speech production abilities of 4- to 5-year-old children with and without a history of late talking: The tricky tyrannosaurus. *International Journal of Speech-Language Pathology*, *22*(2), 184–195. https://doi.org/10.1080/17549507.2019.1638968

Nelson, H. D., Nygren, P., Walker, M., & Panoscha, R. (2006). Screening for speech and language delay in preschool children: Systematic evidence review for the US Preventive Services Task Force. *Pediatrics*, *117*(2), e298-319. https://doi.org/10.1542/peds.2005-1467

Nelson, K. (2007). *Young Minds in Social Worlds: Experience, Meaning, and Memory : Experience, Meaning, and Memory*. Cambridge, MA: Harvard University Press. http://ebookcentral.proquest.com/lib/lancaster/detail.action?docID=3300052

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, *58*(2), 220–249. https://doi.org/10.1016/j.cogpsych.2008.07.002

Nuffield Foundation. (2021, April 27). 62,000 reception-age children in England to take part in Nuffield Early Language Intervention. *Nuffield Foundation*. https://www.nuffieldfoundation.org/news/62000-children-ake-part-in-nuffield-early-language-intervention

O'Neill, D. K. (1996). Two-Year-Old Children's Sensitivity to a Parent's Knowledge State When Making Requests. *Child Development*, *67*(2), 659–677. https://doi.org/10.1111/j.1467-8624.1996.tb01758.x

Özçalişkan, Ş., & Goldin-Meadow, S. (2005). Do parents lead their children by the hand? *Journal of Child Language*, *32*(3), 481–505. https://doi.org/10.1017/S0305000905007002

Özçalışkan, S., & Dimitrova, N. (2013). How gesture input provides a helping hand to language development. *Seminars in Speech and Language*, *34*(4), 227–236. https://doi.org/10.1055/s-0033-1353447

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, *76*(4), 763–782. https:// 10.1111/j.1467-8624.2005.00876.x

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

Pearlmutter. (1989). Learning state space trajectories in recurrent neural networks. *International 1989 Joint Conference on Neural Networks*, 365–372 vol.2. https://doi.org/10.1109/IJCNN.1989.118724

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, *21*(1), 178–185. https://doi.org/10.3758/s13423-013-0466-4

Perry, L. K., & Kucker, S. C. (2019). The Heterogeneity of Word Learning Biases in Late-Talking Children. *Journal of Speech, Language and Hearing Research*, *62*(3), 554–563. https://doi.org/10.1044/2019_JSLHR-L-ASTM-18-0234

Petruccelli, N., Bavin, E. L., & Bretherton, L. (2012). Children with Specific Language Impairment and resolved late talkers: Working memory profiles at 5 years. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1690–1703. https://doi.org/10.1044/1092-4388(2012/11-0288)

Pierce, J. W., & MacAskill, M. R. (2018). *Building Experiments in PsychoPy.* Thousand Oaks, CA: Sage.

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

Posner, M.I., (1980) Orienting of attention. *The Quarterly Journal of Experimental Psychology, 32*(1), 3–25. https://doi-org/10.1080/00335558008248231

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

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

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

Rader, N. de V., & Zukow-Goldring, P. (2012). Caregivers' gestures direct infant attention during early word learning: The importance of dynamic synchrony. *Language Sciences*, *34*(5), 559–568. https://doi.org/10.1016/j.langsci.2012.03.011

Rakoczy, H., Tomasello, M., & Striano, T. (2005). How children turn sbjects into Symbols: A cultural learning account. In L. Namy (Ed) *Symbolic Use and Symbolic Representation: Developmental and Comparative Perspectives,* (pp.69 – 97). Hove, UK: Psychology Press

Ramscar, M., Dye, M., & Klein, J. (2013). Children value informativity over logic in word learning. *Psychological Science*, *24*(6), 1017–1023. https://doi.org/10.1177/0956797612460691

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*(6), 909–957. https://doi.org/10.1111/j.1551-6709.2009.01092.x

Reilly, S., Cook, F., Bavin, E. L., Bretherton, L., Cahir, P., Eadie, P., Gold, L., Mensah, F., Papadopoullos, S., & Wake, M. (2018). Cohort profile: the Early Language in Victoria Study (ELVS). *International Journal of Epidemiology*, *47*(1), 11–20. https://doi.org/10.1093/ije/dyx079

Reilly, S., Wake, M., Bavin, E. L., Prior, M., Williams, J., Bretherton, L., Eadie, P., Barrett, Y., & Ukoumunne, O. C. (2007). Predicting language at 2 years of age: A prospective community study. *Pediatrics*, *120*(6), e1441–e1449. https://doi.org/10.1542/peds.2007-0045

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

Rescorla, L. (1989). The Language Development Survey: a screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders*, *54*(4), 587–599. https://doi.org/10.1044/jshd.5404.587

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. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language, and Hearing Research*, *52*(1), 16–30. https://doi.org/10.1044/1092-4388(2008/07-0171)

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., 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

Rescorla, L., Hadicke-Wiley, M., & Escarce, E. (1993). Epidemiological investigation of expressive language delay at age two: *First Language*, *13*(37), 5–22. https://doi.org/10.1177/014272379301303702

Rescorla, L., Roberts, J., & Dahlsgaard, K. (1997). Late Talkers at 2: Outcome at age 3. *Journal of Speech, Language, and Hearing Research*, *40*(3), 556–566. https://doi.org/10.1044/jslhr.4003.556

Rescorla, L., Ross, G.S., & McClure, S. (2007) Language delay and behavioral/emotional problems in toddlers: findings from two developmental clinics. *Journal of Speech, Language, and Hearing Research, 50(4),* 1063-1078*,* https://doi.org/10.1044/1092-4388(2007/074)

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. https://doi.org/10.1044/jshr.3701.106

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)

Rice, M., & Broome, M. E. (2004). Incentives for Children in Research. *Journal of Nursing Scholarship*, *36*(2), 167–172. https://doi.org/10.1111/j.1547-5069.2004.04030.x

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). Hove, UK: Psychology Press

Roembke, T., & McMurray, B. (2016). Observational word learning: Beyond Propose-But-Verify and associative bean counting. *Journal of Memory and Language*, *87*, 105–127. https://doi.org/10.1016/j.jml.2015.09.005

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

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *WIREs Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

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

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, *323*(5916), 951–953. https://doi.org/10.1126/science.1167025

Rowe, M. L., Özçalişkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language*, *28*(2), 182–199. https://doi.org/10.1177/0142723707088310

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research*, *47*(1), 223–234. https://doi.org/10.1044/1092-4388(2004/019)

Rujas, I., Casla, M., Mariscal, S., Lázaro López-Villaseñor, M., & Murillo Sanz, E. (2019). Effects of grammatical category and morphology on fast mapping in typically developing and late talking toddlers. *First Language*, 0142723719828258. https://doi.org/10.1177/0142723719828258

Rujas, I., Mariscal, S., Casla, M., Lázaro, M., & Murillo, E. (2017). Word and nonword repetition abilities in Spanish language: Longitudinal evidence from typically developing and late talking children. *The Spanish Journal of Psychology*, *20*, E72. https://doi.org/10.1017/sjp.2017.69

Saffran, J. R. (2020). Statistical language learning in infancy. *Child Development Perspectives*, *14*(1), 49–54. https://doi.org/10.1111/cdep.12355

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Samuelson, L. K. (2021). Toward a precision science of word learning: Understanding individual vocabulary pathways. *Child Development Perspectives*, *15*(2), 117–124. https://doi.org/10.1111/cdep.12408

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Aarpenter, and Tomasello. *Child Development*, *69*(1), 94–104. https://doi.org/10.1111/j.1467-8624.1998.tb06136.x

Samuelson, L. K., Kucker, S. C., & Spencer, J. P. (2017). Moving word learning to a novel space: A dynamic systems view of referent selection and retention. *Cognitive Science*, *41*, 52–72. https://doi.org/10.1111/cogs.12369

Schjølberg, S., Eadie, P., Zachrisson, H. D., Oyen, A.-S., & Prior, M. (2011). Predicting language development at age 18 months: Data from the Norwegian Mother and Child Cohort Study. *Journal of Developmental and Behavioral Pediatrics*, *32*(5), 375–383. https://doi.org/10.1097/DBP.0b013e31821bd1dd

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

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *The Quarterly Journal of Experimental Psychology*, *37B*, 1–21. https://doi.org/10.1080/14640748508402082

Shepherd, M. & Müller, H. J. (1989) Movement versus focusing of visual attention. *Perception & Psychophysics,* 46(2), 146 – 154. https://doi.org/10.3758/BF03204974

Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Fei-Fei, L., Keil, F. C., Gweon, H., Tenenbaum, J. B., Jara-Ettinger, J., Adolph, K. E., Rhodes, M., Frank, M. C., Mehr, S. A., & Schulz, L. (2020). Online Developmental Science to Foster Innovation, Access, and Impact. *Trends in Cognitive Sciences*, *24*(9), 675–678. https://doi.org/10.1016/j.tics.2020.06.004

Singleton, N. C. (2018). Late talkers. *Pediatric Clinics of North America*, *65*(1), 13–29. https://doi.org/10.1016/j.pcl.2017.08.018

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1–2), 39–91.

Siu, A. L. (2015). Screening for speech and language delay and disorders in children aged 5 years or younger: US Preventive Services Task Force recommendation statement. *Pediatrics*, *136*(2), e474–e481. https://doi.org/10.1542/peds.2015-1711

Smith, A. C., Monaghan, P., & Huettig, F. (2017). The multimodal nature of spoken word processing in the visual world: Testing the predictions of alternative models of multimodal integration. *Journal of Memory and Language*, *93*, 276–303. https://doi.org/10.1016/j.jml.2016.08.005

Smith, K., Smith, A. D. M., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, *35*(3), 480–498. https://doi.org/10.1111/j.1551-6709.2010.01158.x

Smith, K., Smith, A. D. M., Blythe, R. A., & Blythe, R. A. (2009). Reconsidering human cross-situational learning capacities: A revision to Yu & Smith’s (2007) experimental paradigm. In *N. Taatgen & H. van Rijn (Eds.), Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 2711–2716.

Smith, L. B. (2000). How to learn words: An associative crane. In R. Golinkoff, & K. Hirsh-Pasek (Eds.), *Breaking the word learning barrier* (pp. 51-80). Oxford, UK: Oxford University Press.

Smith, L. B., & Slone, L. K. (2017). A Developmental Approach to Machine Learning? *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.02124

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: Contextually cued attention and early word learning. *Cognitive Science*, *34*(7), 1287–1314. https://doi.org/10.1111/j.1551-6709.2010.01130.x

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word–referent learning. *Trends in Cognitive Sciences*, *18*(5), 251–258. https://doi.org/10.1016/j.tics.2014.02.007

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother’s view: The dynamics of toddler visual experience: Dynamics of toddler visual experience. *Developmental Science*, *14*(1), 9–17. https://doi.org/10.1111/j.1467-7687.2009.00947.x

Southgate, V., Maanen, C. V., & Csibra, G. (2007). Infant pointing: Communication to cooperate or communication to learn? *Child Development*, *78*(3), 735–740. https://doi.org/10.1111/j.1467-8624.2007.01028.x

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958.

Stokes, S. F. (2010). Neighborhood density and word frequency predict vocabulary size in toddlers. *Journal of Speech, Language, and Hearing Research*, *53*(3), 670–683. https://doi.org/10.1044/1092-4388(2009/08-0254)

Stokes, S. F. (2014). The impact of phonological neighborhood density on typical and atypical emerging lexicons. *Journal of Child Language*, *41*(3), 634–657. https://doi.org/10.1017/S030500091300010X

Stokes, S. F., & Klee, T. (2009). The diagnostic accuracy of a new test of early nonword repetition for differentiating late talking and typically developing children. *Journal of Speech, Language, and Hearing Research*, *52*(4), 872–882. https://doi.org/10.1044/1092-4388(2009/08-0030)

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. https://doi.org/10.1017/S0305000911000031

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

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, *126*, 395–411. https://doi.org/10.1016/j.jecp.2014.06.003

Thal, D. J., Miller, S., Carlson, J., & Vega, M. M. (2005). Nonword repetition and language development in 4-year-old children with and without a history of early language delay. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1481–1495. https://doi.org/10.1044/1092-4388(2005/103)

The Bercow Report: 10 Years On. (2018). ICAN and Royal College of Speech and Language Therapists. http://www.bercow10yearson.com/wp-content/uploads/2018/03/337644-ICAN-Bercow-Report-WEB.pdf

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

Tomalski, P., Moore, D. G., Ribeiro, H., Axelsson, E. L., Murphy, E., Karmiloff-Smith, A., Johnson, M. H., & Kushnerenko, E. (2013). Socioeconomic status and functional brain development – associations in early infancy. *Developmental Science*, *16*(5), 676–687. https://doi.org/10.1111/desc.12079

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition* (pp. viii, 388). 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). Hoboken, NJ: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444325485.ch9

Tomasello, M., & Barton, M. E. (1994). Learning words in nonostensive contexts. *Developmental Psychology*, *30*(5), 639–650. https://doi.org/10.1037/0012-1649.30.5.639

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development, 57*(6), 1454–1463. https://doi.org/10.2307/1130423

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*(3), 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x

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

Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language*, *23*(1), 157–176. https://doi.org/10.1017/S0305000900010138

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

Trueswell, J. C., Lin, Y., Armstrong, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, *148*, 117–135. https://doi.org/10.1016/j.cognition.2015.11.002

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156. https://doi.org/10.1016/j.cogpsych.2012.10.001

Twomey, K. E., & Westermann, G. (2018). Curiosity-based learning in infants: A neurocomputational approach. *Developmental Science*, *21*(4), e12629. https://doi.org/10.1111/desc.12629

Vallotton, C. D. (2012). Infant signs as intervention? Promoting symbolic gestures for preverbal children in low-income families supports responsive parent–child relationships. *Early Childhood Research Quarterly*, *27*(3), 401–415. https://doi.org/10.1016/j.ecresq.2012.01.003

Veale, R., Schermerhorn, P., & Scheutz, M. (2011). Temporal, environmental, and social constraints of word-referent learning in young infants: A neurorobotic model of multimodal habituation. *IEEE Transactions on Autonomous Mental Development*, *3*(2), 129–145. https://doi.org/10.1109/TAMD.2010.2100043

Venker, C. E. (2019). Cross-situational and ostensive word learning in children with and without autism spectrum disorder. *Cognition*, *183*, 181–191. https://doi.org/10.1016/j.cognition.2018.10.025

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

Vigliocco, G., Motamedi, Y., Murgiano, M., Wonnacott, E., Marshall, C., Maillo, I. M., & Perniss, P. (2019). Onomatopoeia, gestures, actions and words: How do caregivers use multimodal cues in their communication to children? *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, 1171–1177.

Vlach, H. A., & DeBrock, C. A. (2019). Statistics learned are statistics forgotten: Children's retention and retrieval of cross-situational word learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *45*(4), 700–711. https://doi.org/10.1037/xlm0000611

Vlach, H. A., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support children’s retention of learned words. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00046

Vlach, H. A., & Sandhofer, C. M. (2014). Retrieval dynamics and retention in cross-situational statistical word learning. *Cognitive Science*, *38*(4), 757–774. https://doi.org/10.1111/cogs.12092

Vygotsky, L. S. (1980). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Wake, M., Tobin, S., Girolametto, L., Ukoumunne, O. C., Gold, L., Levickis, P., Sheehan, J., Goldfeld, S., & Reilly, S. (2011). Outcomes of population based language promotion for slow to talk toddlers at ages 2 and 3 years: Let’s Learn Language cluster randomised controlled trial. *British Medical Journal*, *343*, d4741. https://doi.org/10.1136/bmj.d4741

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

Weismer, S. (2007). Typical talkers, late talkers, and children with specific language impairment: A language endowment spectrum? In *The influence of developmental perspectives on research and practice in communication disorders: A Festschrift for Robin S. Chapman.* (pp. 83–102). Mahwah, NJ: Erlbaum.

Weismer, S. E., Venker, C. E., Evans, J. L., & Moyle, M. J. (2013). Fast mapping in late-talking toddlers. *Applied Psycholinguistics*, *34*(1), 69–89. https://doi.org/10.1017/S0142716411000610

Werker, J. F., & Yeung, H. H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences*, *9*(11), 519–527. https://doi.org/10.1016/j.tics.2005.09.003

Whitehouse, A., Robinson, M., & Zubrick, S. R. (2011) Late talking and the risk for Psychosocial problems during childhood and adolescence. *Pediatrics*, peds.2010-2782, https//doi.org/10.1542/peds.2010-2782

Winter, S. M. & Kelley, M. F. (2008) Forty years of school readiness research: What have we learned?, *Childhood Education*, 84:5, 260-266, https://doi.org/10.1080/00094056.2008.10523022

Wray, C., & Norbury, C. F. (2018). Parents modify gesture according to task demands and child language needs. *First Language*, *38*(4), 419–439. https://doi.org/10.1177/0142723718761729

Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*, *107*(2), 118–136. https://doi.org/10.1016/j.jecp.2010.04.014

Wu, R., Tummeltshammer, K. S., Gliga, T., & Kirkham, N. Z. (2014). Ostensive signals support learning from novel attention cues during infancy. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00251

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193. https://doi.org/10.1016/0010-0277(90)90003-3

Xu, F., & Tenenbaum, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, *10*(3), 288–297. https://doi.org/10.1111/j.1467-7687.2007.00590.x

Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences*, *13*(4), 167–174. https://doi.org/10.1016/j.tics.2009.01.008

Yoshida, H., & and Hanania, R. (2007) Attentional highlighting as a mechanism behind early word learning. In McNamara D. S. & Trafton J. G. (eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, 719–724.

Yoshida, H., & Burling, J. (2012) Highlighting: a mechanism relevant for word learning. *Frontiers in Psychology, 3*(262), 1-12. https://doi.org/ 10.3389/fpsyg.2012.00262

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*(13), 2149–2165. https://doi.org/10.1016/j.neucom.2006.01.034

Yu, C., & Smith, L. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*(2), 165–180.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review*, *119*(1), 21–39. https://doi.org/10.1037/a0026182

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, *22*(1), e12735. https://doi.org/10.1111/desc.12735

Yu, C., Zhong, Y., & Fricker, D. (2012). Selective attention in cross-situational statistical learning: Evidence from eye tracking. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00148

Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition*, *145*, 53–62. https://doi.org/10.1016/j.cognition.2015.07.013

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, n/a-n/a. https://doi.org/10.1111/desc.12036

Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science*, *37*(5), 891–921. https://doi.org/10.1111/cogs.12035

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)

Zutlevics, T. (2016). Could providing financial incentives to research participants be ultimately self-defeating? *Research Ethics*, *12*(3), 137–148. https://doi.org/10.1177/1747016115626756

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*(1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

Appendix A: Chapter 2 Supporting information
(supplementary analyses from computational model and behavioural study)

The following information complements the paper Cheung et al. (in press), '*Caregivers use gesture contingently during word learning'*, Developmental Science, e13098, https://doi.org/10.1111/desc.13098. Data and code repository: https://osf.io/6frcw/?view_only=72344789a6294aa19d63a8bd93a628f3

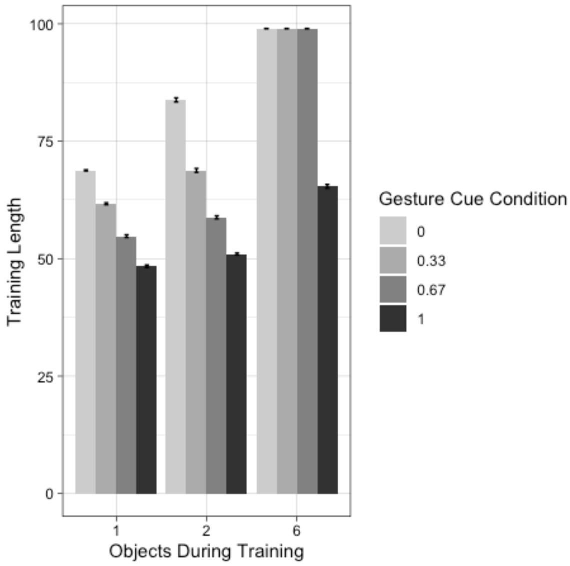
Computational model: additional figure

For the main analysis (main manuscript, Table 1 and Figure 2A & 2B), we collapsed across availability of gesture cue conditions to allow for easier comparison between the model and the behavioural study. Figure S1 shows the breakdown of each gesture availability condition (0%, 33%, 67%, 100%) for training length and for performance during testing.

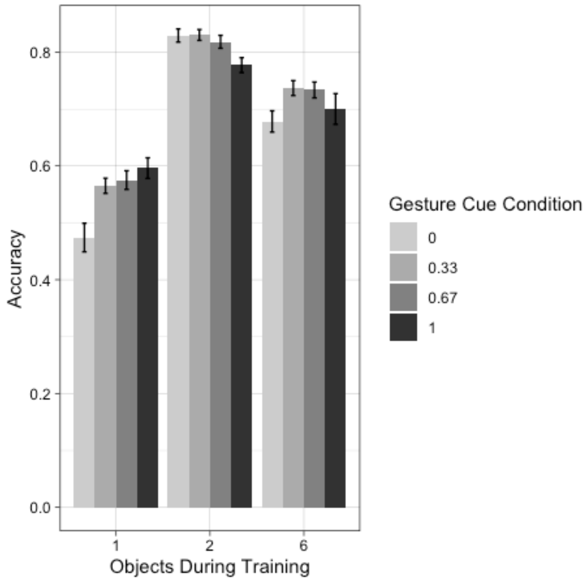
Figure S1.

Multimodal integration model (MIM) results, where model is run to 95% criterion/100 epochs, showing breakdown by availability of gesture cue: effect of number of objects present during training and gesture cue reliability on (A) length of training required; (B) testing accuracy.

(A)



(B)



Computational model: supplementary analyses

Controlling for quantity of exposure

As the original model (main manuscript, Computational Model) learned to criterion more quickly in the one- compared to two- and six-referent conditions, there was also a potential confound in the model's performance affected by referential ambiguity and quantity of exposure. To control for quantity of exposure, we tested also this model on its accuracy when it had been trained to 100 epochs. The results were largely similar to those reported in the main paper, with the exception that the effect of gesture reliability was now not significantly different between the two and six-referent condition. The results are shown in Figure S2 and Table S1.

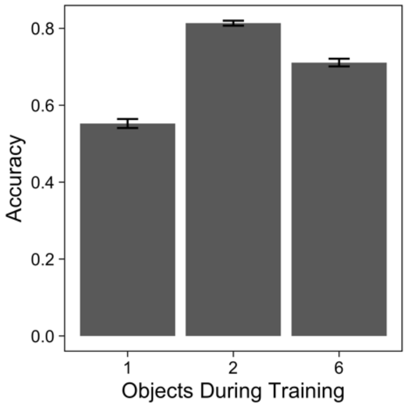
Table S1.

Linear mixed effects model results of the MIM computational model's performance after training to 100 epochs, to control for exposure time across the one-, two-, and six-referent conditions.

Dependent variable	Independent variables	*Estimate*	*SE*	*z*	*p*
Testing accuracy after training to criterion	(intercept – one object)	-0.16	0.16	-1.00	.317
	One v. two objects	3.32	0.20	16.90	< .001
	One v. six objects	1.36	0.20	6.80	< .001
	Two vs. six objects	-1.93	0.15	-13.10	< .001
	Gesture cue	2.49	0.33	7.56	< .001
	One v. two objects x Gesture cue	-1.87	0.40	-4.72	< .001
	One v. six objects x Gesture cue	-2.08	0.37	-5.56	< .001
	Two v. six objects x Gesture cue	-0.19	0.27	-0.69	.493

Figure S2.

Mean and standard error bars for results of the multimodal integration model (MIM) for testing accuracy proportion correct after training to 100 epochs, by number of objects present during training (calculated across gesture cue condition).



Additional simulations

The original model with six objects during training failed in many cases to learn to criterion. We thus conducted a supplementary analysis where we repeated the simulations but increased the number of units in the integrative layer from 100 to 200, reduced the learning criterion from 95% to 90% of the patterns correct on four consecutive blocks of training, and increased the maximum number of training iterations from 100,000 to 200,000.

For training time, the results of the model are shown in Figure S3A. When the gesture cue was low in availability, the model with 6 objects present during training failed to learn to criterion in all cases, but did reach criterion when cue reliability was higher. In the linear effects model, there was a significant effect of number of objects during training on model fit, ($\chi^2(2) = 99.59$, $p < .001$), with one object during training resulting in quicker learning than six objects ($t(116) = 11.017$, $p < .001$), and quicker learning for two than six objects ($t(116) = 10.340$, $p < .001$), but no significant difference between one and two objects ($t(116) = .682$, $p = .497$). There was a significant effect of cue availability ($\chi^2(1) = 71.75$, $p < .001$), with quicker learning when cues were available more often. The interaction also significantly improved fit ($\chi^2(2) = 205.79$, $p < .001$), with no significantly different effect of cue reliability for one than two objects (t(113) = -1.53, p - .127), but a larger effect of gesture reliability for six than one and two objects (t(113) = -20.617, *t*(113) = -19.096, both *p* < .001). The final model is shown in Table S2.

For performance during testing, results are shown in Figure S3B. Generalised linear mixed effects analyses demonstrated a significant effect of number of objects present during training, ($\chi^2(2) = 8.42$, $p = .015$), with one object resulting in lower accuracy than two and six objects (z = 23.50, z = 19.93, both $p < .001$), but accuracy from two and six objects was not significantly different (z = .075, $p = .399$). For cue reliability, the effect was also significant, ($\chi^2(1) = 7.81$, $p = .005$), with increasing availability increasing accuracy. The interaction was also significant, ($\chi^2(2) = 29.5$, $p < .001$), with a greater effect of cue availability on six objects than two objects (z = -4.50, $p < .001$), and a greater effect of gesture on two objects than one object (z = -8.89, $p < .001$). The final model is shown in Table S3.

Table S2.

Linear mixed effects model results of the MIM computational model's performance after training to 90% correct or 200 epochs of training. Results show model fit for the effects of number of objects during training and gesture cue condition on length of training time and accuracy.

Dependent variable	Independent variables	*Estimate*	*SE*	*df*	*t*	*p*
Length of training time	(intercept – one object)	47.35	3.23	113	14.67	< .001
	One v. two objects	11.53	4.56	113	2.53	.013
	One v. six objects	174.39	4.65	113	37.51	< .001
	Two vs. six objects	162.86	4.65	113	35.03	< .001
	Gesture cue	-15.99	5.17	113	-3.09	.003
	One v. Two object x Gesture cue	-11.25	7.31	113	-1.54	.127
	One v. Six object x Gesture cue	-152.60	7.40	113	-20.62	< .001
	Two v. Six object x Gesture cue	-141.34	7.40	113	-19.10	< .001
		Estimate	*SE*		z	*p*
Testing accuracy after training to criterion	(intercept – one object)	-0.97	0.13		-7.50	< .001
	One v. two objects	2.91	0.17		17.35	< .001
	One v. six objects	3.58	0.23		15.81	< .001
	Two vs. six objects	0.70	0.20		3.51	< .001
	Gesture cue	0.54	0.06		9.87	< .001
	One v. two objects x Gesture cue	-0.62	0.07		-9.08	< .001
	One v. six objects x Gesture cue	-.090	0.08		-11.18	< .001
	Two v. six objects x Gesture cue	-0.29	0.06		-4.50	< .001

Table S3.

Linear mixed effects model results of the behavioural study demonstrating the effects of number of objects during training and child vocabulary scores† on caregiver gesture and speech with gesture subtypes during training trials.

Dependent variable	Independent variables	*Estimate*	*SE*	*df*	*t*-value	*p*-value	χ^2 (*df*)	*p*-value
All gestures	(intercept – one object)	3.66	0.36	12.56	10.07	<.001	11.73(2)	.003
	One v. two objects	0.74	0.35	94.00	2.12	.037		
	One v. six objects	1.23	0.35	94.00	3.51	<.001		
	Two v. six objects	0.49	0.35	94.00	1.39	0.167		
Supp. speech with gesture (receptive vocab)	(intercept)	-1.90	2.24	37.21	-0.85	.400	8.09 (3)	.044
	Receptive vocab	0.02	0.01	36.20	2.51	.017		
	Symb. gesture vocab	0.08	0.06	36.56	1.38	.177		
	Receptive*Symb. vocab	-<0.01	<0.01	36.27	-2.33	.026		
Supp. speech with gesture (expressive vocab)	(intercept)	1.66	0.35	21.67	4.76	<.001	18.04(5)	.003
	Expressive vocab	-0.0005	0.001	116.40	-0.34	.73		
	One v. two objects	-0.13	0.34	94.17	-0.37	.71		
	One v. six objects	1.10	0.34	92.53	3.25	.002		
	Two v. six objects	1.22	0.34	94.24	3.62	<.001		
	Expressive*One obj.	-0.0008	0.002	94.25	-0.45	.657		
	Expressive*Two obj.	-0.0008	0.002	94.25	0.45	.657		
	Expressive*Six obj.	-0.005	0.002	93.59	-2.88	.005		

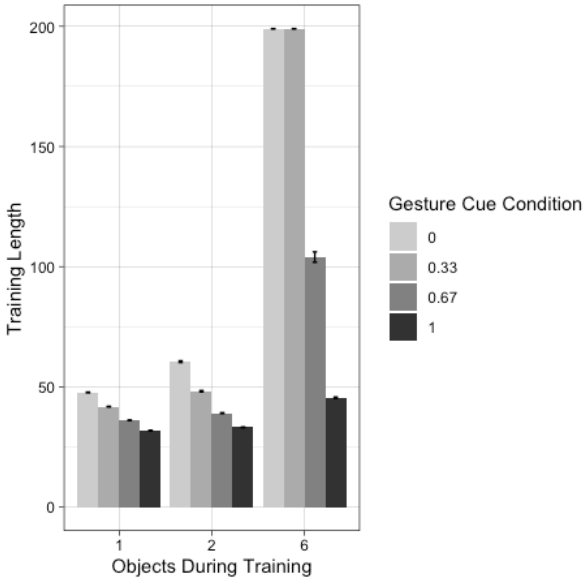
obj. = object; supp. = supplementary; symb. = symbolic; vocab = vocabulary

†Analyses where separate expressive and receptive vocabulary model fits gave different best-fitting model results for the dependent variable are been indicated in the table.

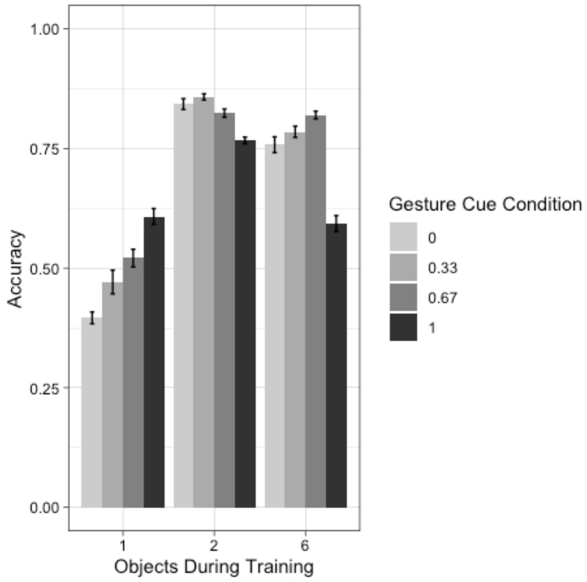
Figure S3.

Multimodal integration model (MIM) results where model is run to 90% criterion/200 epochs: effect of number of objects present during training and gesture cue reliability on (A) length of training required and (B) testing accuracy.

(A)



(B)



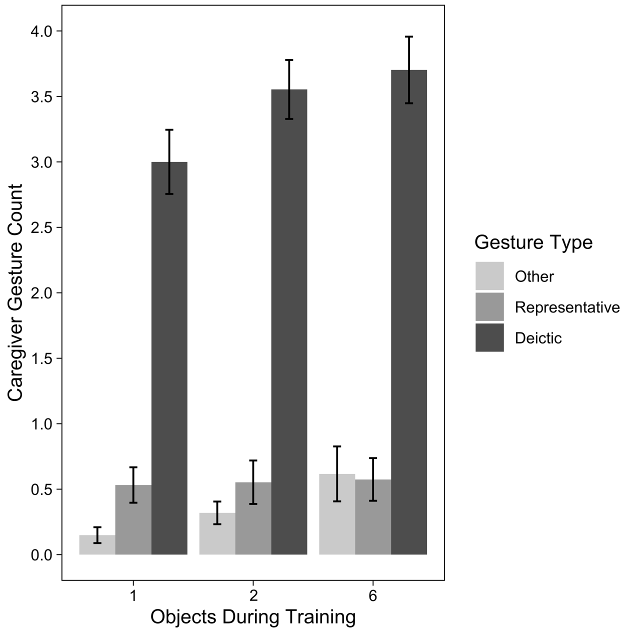
Behavioural study: additional figures

For the main analysis (main manuscript, Table 3 and Figure 2C & 2D), we focused on deictic gesture use. Figure S4 shows a breakdown of the effect of the number of objects present on training for caregiver subtypes of gesture, and also shows the effect of condition on other caregiver subtypes of speech and gesture.

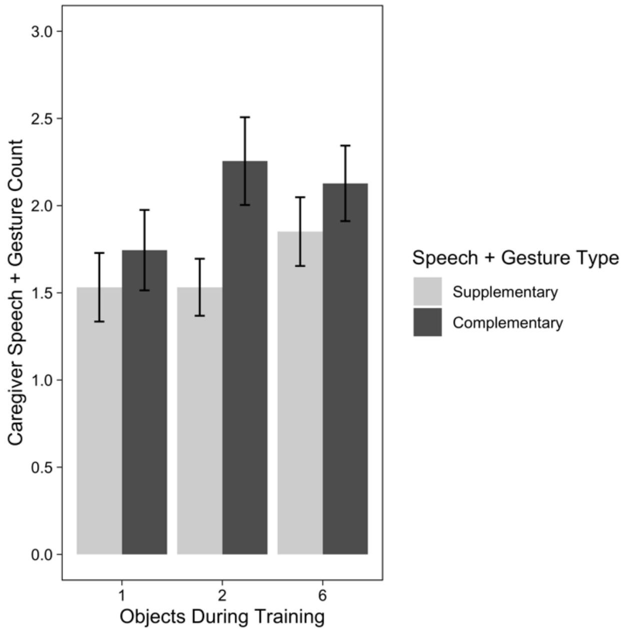
Figure S4.

Mean and standard error bars for results of the behavioural study demonstrating the effect of number of objects present during training on: (A) caregiver use of subtypes of gesture; (B) caregiver use of subtypes of speech and gesture; (C) caregiver use of the referent label.

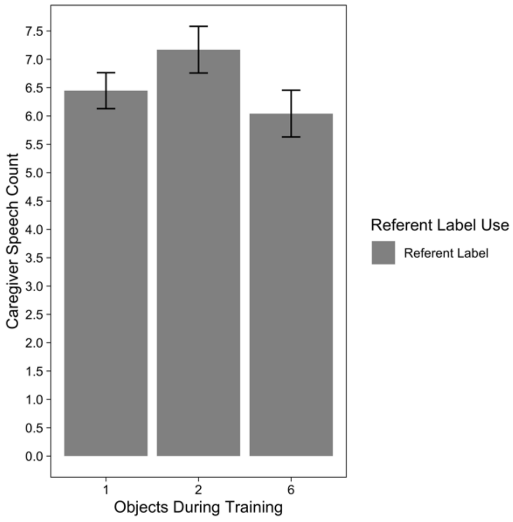
(A)



(B)


Caregiver Speech + Gesture Count
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1
2
6
Objects During Training
Speech + Gesture Type
Supplementary
Complementary

(C)


Caregiver Speech Count
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1
2
6
Objects During Training
Referent Label Use
Referent Label

Behavioural study: supplementary analyses

Part of the difficulty in determining the mechanisms through which gesture might aid language acquisition lies in distinguishing effects due to gesture alone, and those that are the result of gesture co-occurring with speech. Some studies have indicated that once verbal input is accounted for, parent gesture does not correlate with child vocabulary scores, suggesting that much of gesture's value may well be imbedded in the information it provides simultaneously with speech (Iverson et al., 1999; Pan et al., 2005; Rowe et al., 2008). For example, it may be that gestures provide visual information that is not present in speech, such as the hand action of a bird's wings flapping when saying the word 'eagle', or by reinforcing what is said during speech, such as pointing at an intended referent whilst naming it (Goldin-Meadow, 2000; Goldin-Meadow & Wagner, 2005). We thus conducted several further analyses for the behavioural data to examine the range of caregiver communication with their children in the different conditions, and to also investigate the effect of these on child accuracy at test. We also explored how child speech and gesture cues might be affected by referential uncertainty during training.

Cues during training

Separate linear mixed effects models were also constructed to predict caregiver and child gesture speech for each subtype described in Rowe et al.'s (2008) and Iverson and Goldin-Meadow's (2005) coding scheme. These were constructed in the same way as linear mixed effects models described in the main manuscript (Behavioural study). For full details of all models run, please see R code and data files on OSF: (https://osf.io/6frcw/?view_only=72344789a6294aa19d63a8bd93a628f3).

Caregiver supplementary speech and gesture cues during training

Due to high correlation between CDI expressive and receptive vocabulary subscales, separate linear mixed effects models were carried out, one with fixed effects of expressive, symbolic gesture, and communicative gesture vocabulary, and one with receptive, symbolic gesture, and communicative gesture vocabulary. The final results of these models can be found in Table S3.

In the expressive vocabulary linear mixed effect model, there was a main effect of condition and expressive vocabulary ($\chi^2(5) = 18.04$, *p* = .003). However, this was largely driven by effects of condition, with increased supplementary speech and gesture from one to six objects (*t*(92.53) = 3.25, *p* = .002), and from two to six objects (*t*(94.24) = 3.62, *p* < .001). The interaction between expressive vocabulary and the six-object condition was significant (*t*(93.59) = -2.88, *p* = .005), but this estimate was very small.

In the receptive vocabulary linear mixed effect model, there was also marginal interaction between receptive vocabulary and child gesture vocabulary without a fixed effect of condition ($\chi^2(3) = 8.09$, *p* = .044) that suggested caregivers of children with higher receptive and gesture vocabulary gave less supplementary speech with gesture overall, but these estimates were very low.

For overall gesture, only an effect of condition was found $\chi^2(2) = 11.73$, *p* = .003) without any additional effects of child vocabulary. This demonstrated similar results to those in the main manuscript (Behavioural study), with an increase in overall gesture from one to two objects (*t*(94) = 2.12, *p* = .037), and from one to six objects (*t*(94) = 3.51, *p* < .001), but no significant difference from two to six objects (*t*(94) = 1.39, *p* = .167).

Child use of novel label during training

Models of child data examining the use of referent label demonstrated no fixed effect of condition, but there was an effect of child receptive, symbolic gesture, and communicative gesture vocabulary ($\chi^2(3) = 8.86$, *p* = .036, Table S4) indicating that higher receptive vocabulary with lower gesture vocabulary predicted more frequent child use of the novel word overall – although these estimates were very small. Similarly, examining a fixed effect of expressive vocabulary demonstrated an effect without condition or other gesture subscales ($\chi^2(1) = 5.33$, *p* = .012, Table S4); again, this estimate was very small.

Child gesture and speech use during training

Models of child data revealed a significant effect of condition on overall gesture use ($\chi^2(2) = 9.09$, $p = .011$, Table S4), which differed from the way in which caregivers used gestures. Children gestured significantly more in the six-object condition ($t(90) = 3.09$, $p = .003$) compared to the one-object condition, but showed no significant difference between the one- and two-object conditions ($t(89) = 1.60$, $p = .113$), or between the two- and six-object conditions ($t(92) = 1.49$, $p = .139$). When examining gesture subtypes, no significant fixed effects were found in predicting deictic or representational gestures related to the referent. A significant effect of condition was found when predicting other gestures which appeared to drive the effect of condition on overall gesture use ($\chi^2(2) = 13.86$, $p < .001$). When examining the other gesture subtypes post-hoc, the vast majority of other gestures were deictic gestures aimed at non-referent items. Models of child data for co-occurrence of speech (supplementary and complementary) with gesture did not reveal any significant fixed effects or interactions.

Children gestured far less frequently during training than caregivers and spoke very little, which could explain the lack of significant differences in speech and gesture co-occurrence between conditions. However, they gestured more often in the six-object condition than the one-object condition, although this comprised primarily gestures towards non-target items. The obvious difference between knowledge states of caregiver and child may play a role here – caregivers knew what the target was, and so pointed at it more often. Children did not have the benefit of prior knowledge, and subsequently may have pointed more than their parents at the other novel objects on the tray in the six-object condition simply because there were more of them compared to the one-object condition.

Table S4.

Linear mixed effects model results of the behavioural study demonstrating the effects of number of objects during training and child vocabulary scores† on child gesture and speech with gesture subtypes during training trials.

Dependent variable	Independent variables	*Estimate*	*SE*	*df*	*t*-value	*p*-value	χ^2 (*df*)	*p*-value
All gestures	(intercept – one object)	1.27	0.32	35.95	4.00	<.001	9.09 (2)	.011
	One v. two objects	0.52	0.33	88.64	1.60	.113		
	One v. six objects	1.02	0.33	90.39	3.09	.003		
	Two v. six objects	0.49	0.33	92.31	1.49	.139		
Other gestures	(intercept – one object)	0.53	0.21	46.59	2.54	.014	13.86(2)	<.001
	One v. two objects	0.63	0.25	90.06	2.46	.012		
	One v. six objects	0.97	0.26	92.19	3.81	<.001		
	Two v. six objects	0.35	0.26	93.96	1.35	.0179		
Referent label use (receptive vocab)	(intercept)	1.00	0.73	39.99	1.37	.177	8.56 (3)	.036
	Receptive vocab	0.01	0.002	39.78	2.90	.006		
	Symb. gesture vocab	-0.01	0.02	39.89	-0.50	.622		
	Comm. gesture vocab	-0.06	0.04	39.98	-1.68	.101		
Referent label use (expressive vocab)	(intercept)	0.26	0.19	33.93	1.36	.181	6.33(1)	.012
	Expressive vocab	0.002	0.001	46.91	2.60	.012		

comm. = communicative; symb. = symbolic; vocab = vocabulary

†Analyses where separate expressive and receptive vocabulary model fits gave different best-fitting model results for the dependent variable are been indicated in the table.

Accuracy at test

We used Generalised Estimated Equations (GEE; *geeglm* package; *geepack* in R[v3.4.1, 2017]) to examine the effect of condition, caregiver behaviour and child vocabulary during training on test trial accuracy. Separate GEEs were constructed to examine child vocabulary variables, condition, and each training behaviour gesture subtype as independent variables. For full details of all models run, please see R code and data files on OSF: (https://osf.io/6frcw/?view_only=72344789a6294aa19d63a8bd93a628f3).

Caregiver behaviour on accuracy at test

When entered as independent variables alongside that of condition and other child vocabulary subscales (receptive, expressive, symbolic, and communicative subscales), caregiver deictic gesture use, novel label use, complementary speech and gesture, and supplementary speech and gesture were not a significant predictors of child accuracy at test (Tables S5-S8). These results are similar to that of the original manuscript that examined caregiver deictic gesture use and receptive child vocabulary (main manuscript, Table 3; see also General Discussion).

Table S5.

Generalised estimated equations results of the behavioural study predicting effect of caregiver deictic gesture cues during training and child vocabulary scores† (UK-Communicative Development Inventories) on child accuracy at test.

GEE model	Independent variables	*Estimate*	*SE*	*Wald*	*p*-value
1	(intercept – one object)	-1.33	0.53	6.24	.012
	One v. two objects	0.90	0.43	4.41	.036
	One v. six objects	0.85	0.46	3.36	.067
	Two v. six objects	-0.05	0.49	0.01	.923
	Expressive vocab	0.001	0.001	1.23	.267
	Caregiver deictic gesture	0.03	0.10	0.11	.745
2	(intercept – one object)	-1.72	1.19	2.08	.149
	One v. two objects	0.88	0.45	3.86	.049
	One v. six objects	0.71	0.71	2.24	.134
	Two v. six objects	-0.18	0.50	0.12	.726
	Symb. gesture vocab	0.02	0.03	0.59	.443
	Caregiver deictic gesture	-0.001	0.10	0.00	.989
3	(intercept – one object)	-2.53	1.15	4.83	.028
	One v. two objects	0.89	0.43	4.33	.038
	One v. six objects	0.84	0.48	3.15	.076
	Two v. six objects	-0.05	0.05	0.01	.920
	Comm. gesture vocab	0.07	0.05	1.85	.173
	Caregiver deictic gesture	0.05	0.10	0.22	.639

comm. = communicative; symb. = symbolic; vocab = vocabulary

† Receptive vocabulary was reported in the main manuscript

Table S6.

Generalised estimated equations results of behavioural study predicting effect of caregiver referent label use during training and child vocabulary scores (UK-Communicative Development Inventories) on child accuracy at test.

GEE model	Independent variables	*Estimate*	*SE*	*Wald*	*p*-value
1	(intercept – one object)	-0.92	0.72	1.64	.20
	One v. two objects	0.99	0.46	4.70	.03
	One v. six objects	0.80	0.50	2.52	.11
	Two v. six objects	-0.19	0.50	0.16	.69
	Receptive vocab	0.002	0.002	1.41	.24
	Caregiver referent label use	-0.11	0.08	2.19	.14
2	(intercept – one object)	-0.50	0.57	0.75	.39
	One v. two objects	0.99	0.46	4.69	.03
	One v. six objects	0.80	0.50	2.55	.11
	Two v. six objects	-0.19	0.50	0.14	.70
	Expressive vocab	0.0009	0.001	0.77	.38
	Caregiver referent label use	-0.11	0.08	1.84	.17
3	(intercept – one object)	-1.26	0.93	1.83	.176
	One v. two objects	0.98	0.49	3.03	.044
	One v. six objects	0.67	0.51	1.72	.19
	Two v. six objects	-0.31	0.51	0.36	.548
	Symb. gesture vocab	0.02	0.02	0.99	.319
	Caregiver referent label use	-0.10	1.08	1.43	.232
4	(intercept – one object)	-1.70	1.02	2.80	.094
	One v. two objects	1.01	0.46	4.74	.029
	One v. six objects	0.79	0.51	2.39	.122
	Two v. six objects	-0.22	0.51	0.18	.669
	Comm. gesture vocab	0.07	0.05	2.42	.12
	Caregiver referent label use	-0.13	0.07	2.87	.09

comm. = communicative; symb. = symbolic; vocab = vocabulary

Table S7.

Generalised estimated equations results of the behavioural study predicting effect of caregiver complementary speech with gesture cues during training and child vocabulary scores (UK-Communicative Development Inventories) on child accuracy at test.

GEE model	Independent variables	*Estimate*	*SE*	*Wald*	*p*-value
1	(intercept – one object)	-1.73	0.61	8.11	.004
	One v. two objects	0.88	0.42	4.32	.038
	One v. six objects	0.84	0.48	3.06	.08
	Two v. six objects	-0.04	0.50	0.01	.938
	Receptive vocab	0.002	0.002	1.15	.284
	Caregiver complementary speech + gesture	0.06	0.12	0.28	.6
2	(intercept – one object)	-1.35	0.44	9.20	.002
	One v. two objects	0.88	0.42	4.29	.038
	One v. six objects	0.84	0.48	3.02	.082
	Two v. six objects	-0.03	0.50	0.00	.945
	Expressive vocab	0.001	0.001	1.31	.252
	Caregiver complementary speech + gesture	0.07	0.12	0.38	.535
3	(intercept – one object)	-1.70	1.10	2.37	.123
	One v. two objects	0.85	0.44	3.68	.055
	One v. six objects	0.69	0.48	2.07	.151
	Two v. six objects	-0.16	0.50	0.10	.749
	Symb. gesture vocab	0.02	0.02	0.48	.489
	Caregiver complementary speech + gesture	0.05	0.12	0.16	.693
4	(intercept – one object)	-2.39	1.05	5.16	.023
	One v. two objects	0.89	0.43	4.39	.036
	One v. six objects	0.85	0.49	3.03	.082
	Two v. six objects	-0.04	0.50	0.01	.932
	Comm. gesture vocab	0.06	0.05	1.62	.203
	Caregiver complementary speech + gesture	0.05	0.11	0.22	.642

comm. = communicative; symb. = symbolic; vocab = vocabulary

Table S8.

Generalised estimated equations results of the behavioural study predicting effect of caregiver supplementary speech with gesture cues during training and child vocabulary scores (UK-Communicative Development Inventories) on child accuracy at test.

GEE model	Independent variables	*Estimate*	*SE*	*Wald*	*p*-value
1	(intercept – one object)	-1.91	0.65	8.63	.003
	One v. two objects	0.91	0.44	4.27	.039
	One v. six objects	0.84	0.49	2.91	.088
	Two v. six objects	-0.08	0.51	0.02	.876
	Receptive vocab	0.002	0.002	1.41	.236
	Caregiver supplementary speech + gesture	0.11	0.12	0.90	.343
2	(intercept – one object)	-1.44	0.43	11.18	<.001
	One v. two objects	0.92	0.44	4.29	.038
	One v. six objects	0.84	0.49	2.90	.089
	Two v. six objects	0.08	0.50	0.02	.878
	Expressive vocab	0.05	0.001	1.51	.22
	Caregiver supplementary speech + gesture	0.11	0.12	0.87	.35
3	(intercept – one object)	-2.08	1.20	2.98	.084
	One v. two objects	0.89	0.47	3.64	.056
	One v. six objects	0.69	0.50	1.91	.167
	Two v. six objects	-0.20	0.51	0.15	.697
	Symb. gesture vocab	0.02	0.03	0.78	.376
	Caregiver supplementary speech + gesture	0.12	0.12	1.00	.316
4	(intercept – one object)	-2.79	1.16	5.77	.016
	One v. two objects	0.92	0.44	4.26	.039
	One v. six objects	0.83	0.50	2.78	.095
	Two v. six objects	-0.09	0.51	0.03	.865
	Comm. gesture vocab	0.07	0.05	2.13	.144
	Caregiver supplementary speech + gesture	0.14	0.12	1.48	.224

comm. = communicative; symb. = symbolic; vocab = vocabulary

Behavioural study: training of coders

Pilot data was used as training data. Four parent-dyads were run through the training procedure only to examine whether children could tolerate the paradigm, but their results were not analysed. Different video clips of each subtype of gesture and speech with gesture were isolated as training examples by an experienced coder, e.g. for deictic gestures, showing gestures such as extending the arm and presenting the palmar surface of the hand, or an index point. These were then used to train an independent coder (who had previous experience in video coding) on this specific coding system. Full videos of complete pilot training sessions were coded by the experienced coder. The independent coder then coded the same videos separately, and received feedback on the quality of their coding until percent agreement for categorisation into subtypes was above 80% (Hallgren, 2012).

Behavioural study: excluded participants

Table S9 shows the demographics and CDI subscores of participants who were excluded from both training and testing. Of the six children who were excluded, *n* = 1 was unwell (not disclosed until training began); *n* = 1 child needed the bathroom during a training trial; *n* = 4 were excluded due to 'typical' child fussiness, e.g. irritability or excessive fidgeting.

Table S9:

Behavioural study: demographics and child vocabulary scores (UK-Communicative Development Inventories) of excluded versus included training sample.

	Excluded from training (n = 6)	Completed training (n = 47)
	mean (sd)	*mean (sd)*
Sex (m:f ratio)	3:3	27:20
Age (months)	20.9 (1.7)	20.5 (1.7)
Receptive vocab	303.8 (70.5)	276 (91.5)
Expressive vocab	83.8(66.9)	146 (114)
Comm. gesture vocab	15.8 (6.4)	19.9 (3.79)
Symb. gesture vocab	31.7 (13.7)	41.1 (6.9)

References

Goldin-Meadow, S. (2000). Beyond words: The importance of gesture to researchers and learners. *Child Development*, *71*(1), 231–239. https://doi.org/10.1111/1467-8624.00138

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Sciences*, *9*(5). https://doi.org/10.1016/j.tics.2005.03.006

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother-child interactions. *Cognitive Development*, *14*(1), 57–75. https://doi.org/10.1016/S0885-2014(99)80018-5

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, *76*(4), 763–782. https://doi.org/10.1111/j.1467-8624.2005.00876.x

Rowe, M. L., Özçalişkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language*, *28*(2), 182–199. https://doi.org/10.1177/0142723707088310

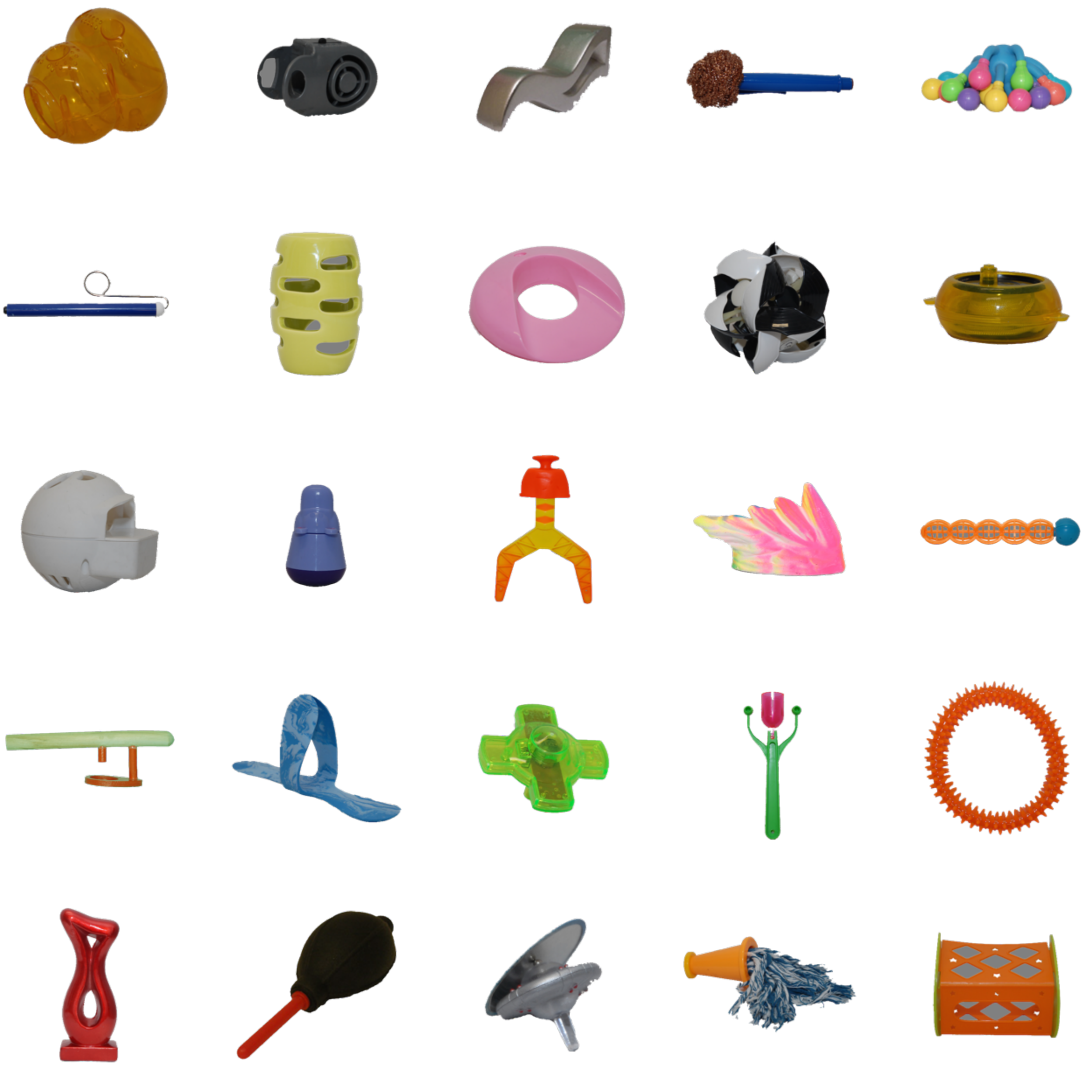
Appendix B: Chapter 3 Supporting information

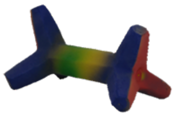
(list of stimuli used in cross-situational word learning task)

Stimuli lists

All items from: Horst, J. S. & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: a collection of novel images for use in experimental research. *Behavior Research Methods 48 (4), 1393-1409*. These items and further information around their properties can be viewed at the original NOUN Database source: http://www.sussex.ac.uk/wordlab/noun.

Novel objects





Novel words: sound files created using ‘Serena’ voice in Mac OS X

agen
akar
blicket
boskot
chatten
colat
coodle
eder
eget
fiffin
gasser
isot
jefa
kaki
kita
koba
manu
modi
osip
pentants
pizer
reda
regli
tannin
tanzer
teebu
tever
toma
tulver
upos
virdex
wiso

Appendix C: Chapter 5 Supporting information

(supplementary analyses of nonword repetition task and stimuli for fast mapping and CSWL tasks)

A number of additional analyses were undertaken testing the effect of receptive vocabulary across all tasks, the effect of expressive vocabulary on syllable loss in the nonword repetition task, and cross-correlations with expressive and receptive vocabulary over time. These are listed here. For all linear mixed effects (LME) and general linear mixed effects models (GLME) reported, the same procedure was utilised to build models as detailed in the main manuscript.

Preschool Repetition (PSRep) Test (Chiat & Roy, 2007)

Table S1 shows a breakdown of accuracy and syllable loss across groups, word type, and word length.

Table S1. PSRep Test: accuracy and syllable loss mean and standard error

Word length	Non words: mean (*SE*)			
	Late talking (n = 19)		*Typically developing (n = 31)*	
	Accuracy	*Syllable loss*	*Accuracy*	*Syllable loss*
One-syllable words	0.61 (0.05)	*n.a.*	0.91 (0.02)	*n.a.*
Two-syllable words	0.41 (0.05)	0.10 (0.03)	0.70 (0.03)	0.06 (0.02)
Three-syllable words	0.31 (0.04)	0.32 (0.05)	0.67 (0.03)	0.09 (0.02)
	Real words: mean (*SE*)			
	Late talking (n = 19)		*Typically developing (n = 31)*	
	Accuracy	*Syllable loss*	*Accuracy*	*Syllable loss*
One-syllable words	0.63 (0.05)	*n.a.*	0.92 (0.02)	*n.a.*
Two-syllable words	0.49 (0.05)	0.18 (0.05)	0.85 (0.03)	0.04 (0.01)
Three-syllable words	0.41 (0.05)	0.40 (0.05)	0.76 (0.03)	0.14 (0.03)

n.a. = not applicable

Expressive vocabulary and PSRep Test syllable loss: For the PSRep Test, expressive vocabulary was also tested with regard to syllable loss (only accuracy was reported in the main manuscript).

We predicted syllable loss using two LME analyses, with fixed effects of 1) T1 Population (0 = TD, 1 = LT) and 2) T2 expressive vocabulary (EOWPVT-4 score). Each model also had a fixed effect of word type (real word = 0, nonword = 1) and random effects of participant and target item.

There was an effect of T1 Population on syllable loss ($\chi^2(1) = 16.56$, *p* <.001; Table S2), indicating that LT children lost more syllables than TD children (*p* <.001), with no effect of word type.

There was also an effect of T2 expressive vocabulary on syllable loss. The best-fitting model to the data contained fixed effects of T2 expressive vocabulary, trial type, and an interaction between expressive vocabulary and trial type, with random effects of participant and target ($\chi^2(2) = 9.43$, *p* = .009; Table S2). This model indicated that those with higher vocabularies lost less syllables (*p* <.001), that all children lost fewer syllables on non-word items in comparison to word items (*p = .002).* The interaction term indicated that children who had higher expressive vocabularies lost more syllables in non-word trials as compared to word trials (*p* = .003).

Children identified as LT at T1 thus lost more syllables, despite all but one having recovered using expressive percentile criteria (Table S1). Those with higher concurrent (T2) expressive vocabularies also lost fewer syllables. Interestingly, those with higher expressive vocabularies lost more syllables in non-word trials, perhaps indicating some reliance on existing expressive vocabulary to perform well on word trials.

Table S2.

Nonword repetition task: linear mixed effects results predicting syllable loss by fixed effects of T1 and T2 expressive vocabulary.

Relation with T1 expressive vocabulary at 2;0 – 2;5-years-old					
Fixed effect	*estimate*	*SE*	*t-value*	*df*	*p-value*
(intercept – typically developing)	0.05	0.02	2.28	65.39	.023
T1 population (CDI; late talking, 1)	0.11	0.03	4.41	52.13	<.001
Relation with T2 expressive vocabulary at 3;0 – 3;5-years-old					
Fixed effect	*estimate*	*SE*	*t-value*	*df*	*p-value*
(intercept)	0.86	0.13	6.46	90.05	<.001
T2 expressive vocabulary (EOWPVT-4)[a]	-0.64	0.11	-5.79	83.53	<.001
Trial type (non-word, 1)	-0.39	0.13	-3.06	1371.00	.002
T2 expressive[a] * trial type (non-word, 1)	0.32	0.10	3.01	1698.79	.003

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

<u>Receptive vocabulary and PSRep Test accuracy:</u> We predicted accuracy (item correct) using two GLME analyses, with fixed effects of 1) T1 receptive vocabulary (CDI), and 2) T2 receptive vocabulary (ROWPVT-4 score). Each model also had fixed effects of word length (number of syllables) and word type (word = 0, nonword = 1), and random effects of participant and target item. Random slopes of participant per target were attempted but caused non-convergence, so were omitted from the model.

There was a significant effect of T1 receptive vocabulary on accuracy. The best-fitting model contained an effect of T1 receptive vocabulary ($\chi^2(2) = 12.71$, $p = .002$; Table S3), indicating that the higher children's receptive vocabularies were, the more accurately they scored ($p < .001$), and a fixed effect of word length, indicating that children scored less accurately with when words were longer (2-syllable, $p = .008$; 3-syllable:, $p <.001$). There was no effect of word type.

There was also an effect of T2 receptive vocabulary on accuracy. The best-fitting model to the data contained fixed effects of receptive vocabulary, word length, word type, and one interaction between receptive vocabulary and word type, and another interaction between receptive vocabulary and word length ($\chi^2(2) = 6.50$, $p = .039$; Table S4). This model indicated that those with higher receptive vocabularies scored more accurately ($p = .006$). The interaction term between receptive vocabulary and word type indicated that children with higher receptive vocabularies scored less accurately on nonwords ($p = .034$), and the interaction between receptive vocabulary and word length indicated that children with higher receptive vocabularies scored less accurately on 2-syllable words, although this was at chance ($p = .050$).

Table S3.

Nonword repetition task: linear mixed effects results predicting syllable loss by fixed effects of T1 and T2 receptive vocabulary.

Relation with T1 receptive vocabulary at 2;0 – 2;5-years-old				
Fixed effect	*estimate*	*SE*	*z-value*	*p-value*
(intercept)	-1.43	0.92	-1.55	.120
T1 receptive vocabulary (CDI)[a]	1.04	0.25	4.18	<.001
2-syllable words	-1.22	0.46	-2.67	.008
2-syllable words	-1.73	0.46	-3.78	<.001
Relation with T2 receptive vocabulary at 3;0 – 3;5-years-old				
Fixed effect	*estimate*	*SE*	*z-value*	*p-value*
(intercept)	-7.71	2.60	-2.97	.003
T2 receptive vocabulary (ROWPVT-4)[a]	8.97	2.29	3.92	<.001
2-syllable words	2.33	1.85	1.26	.209
3-syllable words	-2.14	1.95	-1.10	.272
Word type (nonword, 1)	2.52	1.49	1.69	.091
T2 receptive * 2-syllables	-3.17	1.61	-1.96	.050
T2 receptive * 3-syllables	0.33	1.69	0.19	.847
T2 receptive * word type (nonword, 1)	-2.73	1.29	-2.12	.034

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

<u>Receptive vocabulary and PSRep Test syllable loss:</u> We predicted syllable loss using GLME analyses with fixed effects of T1 receptive vocabulary (CDI), and T2 receptive vocabulary (ROWPVT-4 score), and word type (word or non-word), and random effects of participant and target item. Random slopes of participant per target were attempted but caused non-convergence, so were omitted from the model.

There was a significant predictive effect of T1 receptive vocabulary on nonword repetition accuracy. The best-fitting model to the data contained fixed effects of receptive vocabulary, with random effects of participant and target word ($\chi^2(1) = 5.05$, $p = .025$; Table S4). Children with higher receptive vocabularies scored more accurately on the task ($p = .025$). There was no interaction between receptive vocabulary and word length.

There was also a predictive effect of T2 receptive vocabulary on accuracy. The best-fitting model to the data contained fixed effects of T2 receptive vocabulary and word length, with an interaction between receptive vocabulary and word length, and random effects of participant and target word ($\chi^2(1) = 9.06$, $p = .003$; Table S4). Nonword repetition accuracy increased with higher receptive vocabulary (*p* .003).

Table S4.

Nonword repetition task: general linear mixed effects model results predicting accuracy by fixed effects of T1 and T2 receptive vocabulary.

Relation of accuracy with T1 receptive vocabulary at 2;0 – 2;5-years-old					
Fixed effect	*estimate*	*SE*	*t-value*	*df*	*p-value*
(intercept)	0.23	0.06	3.67	60.53	<.001
T1 receptive vocabulary (CDI)[a]	-0.04	0.02	-2.31	51.39	.025
Relation of accuracy with T2 receptive vocabulary at 3;0 – 3.5-years-old					
Fixed effect	*estimate*	*SE*	*t-value*	*df*	*p-value*
(intercept)	0.53	0.14	3.79	53.19	<.001
T2 receptive vocabulary (ROWPVT-4) [a]	-0.39	0.12	-3.15	52.48	.003

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

Receptive vocabulary and fast mapping and retention task (Hartley et al., 2019)

Referent selection: We predicted referent selection accuracy using two GLME analyses with fixed effects of 1) T1 receptive vocabulary (CDI score), and 1) T2 receptive vocabulary (ROWPVT-4). These models also had random effects of participant and target item. Random slopes of participant per target did not converge, and so were omitted.

There was no effect of T1 receptive vocabulary on referent selection accuracy. There was, however, an effect of T2 receptive vocabulary. A model with fixed effects of T2 receptive vocabulary provided the best fit to the data ($\chi^2(1) = 17.67(1)$, *p*-value <.001; Table S5). This showed that the higher participants' receptive vocabulary, the more accurately they scored on referent selection trials (*p* <.001).

Retention: We predicted retention accuracy using two GLME analyses with fixed effects of 1) T1 receptive vocabulary (CDI score), and 1) T2 receptive vocabulary (ROWPVT-4). These models also had a fixed effect of previous referent selection accuracy for the same word (incorrect = 0, correct = 1), and random effects of participant and target item. Random slopes of participant per target did not converge, and so were omitted. There was no effect of T1 or T2 receptive vocabulary, yielded no significant improvements in fit over the null model.

In sum, although receptive vocabulary at T1 did not predict fast mapping proficiency, having concurrent higher receptive vocabulary at T2 predicted accuracy on referent selection trials. When data from referent selection and retention trials were combined, higher receptive vocabulary also predicted performance across the task and within retention trials. The effect of concurrent receptive vocabulary on retention trial accuracy was smaller than concurrent expressive vocabulary.

Table S5.

Fast mapping task: general linear mixed effects model results predicting accuracy in referent selection trials by fixed effects of T2 receptive vocabulary.

Relation with T2 receptive vocabulary at 3;0 – 3;5-years-old				
Fixed effect	*estimate*	*SE*	*z-value*	*p-value*
(intercept)	-6.81	1.94	-3.51	<.001
T2 receptive vocabulary (ROWPVT-4)[a]	7.67	1.78	4.31	<.001

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

Receptive vocabulary and cross-situational word learning task (CSWL; Hartley et al., 2020)

We predicted training trial accuracy and retention accuracy using two GLME analyses with fixed effects of 1) T1 receptive vocabulary (CDI score), and 1) T2 receptive vocabulary (ROWPVT-4). These models also had random effects of participant and target item. Random slopes of participant per target did not converge, and so were omitted. There was no effect of T1 or T2 receptive vocabulary on accuracy.

Correlations between task performance and vocabulary

We conducted Kendall's rank correlation *tau-b* values (one-tailed) to assess the relationships between task performance and receptive and expressive vocabulary over time (Table S6).

At T1 (2;0 – 2;5-years-old), receptive and expressive vocabulary significantly correlated with PSRep Test accuracy, with expressive vocabulary ($\tau = 0.49$) yielding higher correlations than receptive ($\tau = 0.36$).

At T2 (3;0 – 3;5-years-old), expressive vocabulary had higher significant correlations with the tasks than receptive on the PSRep Test (expressive: $\tau = 0.45$; receptive: $\tau = 0.31$) and fast mapping referent selection (expressive*:* $\tau = 0.31$; receptive: $\tau = 0.35$). Expressive vocabulary also predicted fast mapping retention performance ($\tau = 0.20$), whereas receptive vocabulary did not.

At T3 (3;6 – 3;11-years-old), expressive vocabulary yielded higher correlations with the tasks than receptive on the PSRep Test (expressive: $\tau = 0.45$; receptive: $\tau = 0.31$) and fast mapping retention (expressive: $\tau = 0.31$; receptive: $\tau = 0.27$). Expressive vocabulary also predicted fast mapping referent selection performance ($\tau = 0.31$), whereas receptive vocabulary did not.

Table S6.

Kendall's rank tau correlations between tasks and vocabulary over time.

Vocabulary	PSRep Test	Fast mapping		CSWL	
	Accuracy	*Referent selection*	*Retention*	*Referent selection*	*Retention*
		T1: 2;0 – 2;5-years-old			
Expressive (Oxford-CDI)	**τ = 0.49, *p* < .001**	τ = 0.06, *p = n.s.*	τ = 0.11, *p = n.s*	τ = -0.10, *p = n.s*	τ = 0.08, *p = n.s*
Receptive (Oxford-CDI)	**τ = 0.36, *p* <.001**	τ = 0.134, *p = n.s*	τ = 0.05, *p = n.s*	τ = -0.02, *p = n.s*	τ = 0.05, *p = n.s*
		T2: 3;0 – 3;5-years-old			
Expressive (EOWPVT-4)	**τ = 0.33, *p* < .001**	**τ = 0.40, *p* < .001**	**τ = 0.20, *p* = .050**	τ = 0.08, *p = n.s*	τ = 0.09, *p = n.s*
Receptive (ROWPVT-4)	**τ = 0.29, *p* = .003**	**τ = 0.35, *p* = .001**	τ = 0.10, *p = n.s*	τ = 0.19, *p = n.s*	τ = 0.18, *p = n.s*
		T3: 3;6 – 3;11-years-old			
Expressive (EOWPVT-4)	**τ = 0.45, *p* < .001**	**τ = 0.31, *p* = .023**	**τ = 0.31, *p* = .014**	τ = 0.13, *p = n.s*	τ = 0.29, *p = n.s*
Receptive (ROWPVT-4)	**τ = 0.31, *p* = .010**	τ = 0.20, *p = n.s*	**τ = 0.27, *p* = .035**	τ = 0.12, *p = n.s*	τ = 0.20, *p = n.s*

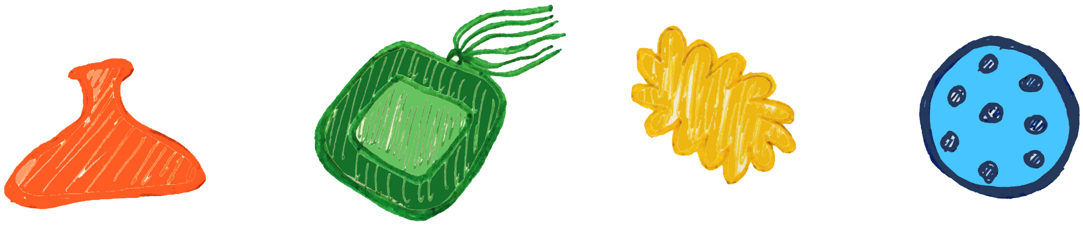
CDI = Communicative Development Inventories; CSWL = Cross-situational word learning; EOWPVT-4 = Expressive One Word Picture Vocabulary Test 4th Edition; T = timepoint

Stimuli used for fast mapping task

Familiar objects (all toys, none live)

1. Frog
2. Grapes
3. Tomato
4. Cup
5. Car
6. Spoon
7. Knife
8. Fish

Unfamiliar objects

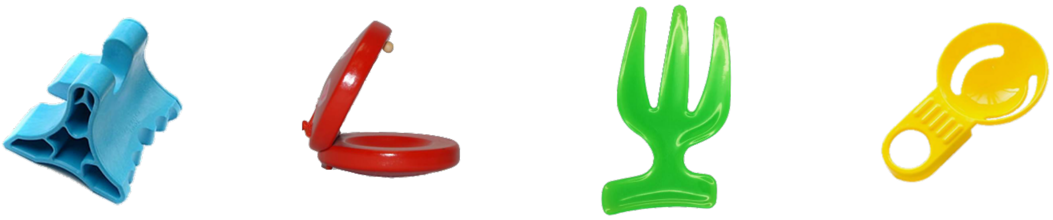


Labels (Horst & Hout, 2016)

1. dax
2. wug
3. yok
4. lep

Stimuli used for cross-situational word learning task

Stimuli



Labels (Horst & Hout, 2016)

1. blicket
2. teebu
3. fiffin
4. virdex

References

Chiat, S., & Roy, P. (2007) The Preschool Repetition Test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research,* 50(2): 429-443

Hartley, C., Bird, L. A., & Monaghan, P. (2019) Investigating the relationship between fast mapping, retention, and generalisation of words in children with autism spectrum disorder and typical development. *Cognition*, 187: 126-138

Hartley, C., Bird, L. A., & Monaghan, P. (2020) Comparing cross-situational word learning, retention, and generalisation in children with autism and typical development. *Cognition*, 200: 104265

Horst, J., & Hout, M. (2016) The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, 48(4): 1393-1409

Appendix D: Chapter 6 Supporting information

(list of stimuli used in verbal scaffolding task)

Category	Control Familiar	Control Unfamiliar	Standard Familiar	Standard Unfamiliar
Animal	German shepherd (sitting) + Burnese 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 + silver metal Western spoon	Metal garlic press + wooden garlic press	Black hairbrush + wooden comb	Binoculars + safety science glasses