Do multiple exemplars promote preschool children’s retention and generalisation of words learned from pictures?

Word count: 5,781 (excluding abstract, tables, figures & references)

**Abstract**

We investigated whether preschool children’s extension of labels from memory representations of pictures is enhanced by exposure to multiple exemplars during teaching. Neurotypical 2-year-olds (N = 23) and 3-year-olds (N = 19) mapped novel word-picture associations in a referent selection task. Their retention and generalisation of labels was then assessed after 5 minutes with depicted 3-D objects. During referent selection, children were presented with a single variant of each novel picture (single exemplar condition) or two differently coloured variants of each novel picture (multiple exemplar condition). Both age groups extended labels to similarly coloured objects with significantly greater accuracy when taught with multiple exemplars. Three-year-olds also generalised labels to differently coloured category members with significantly greater accuracy in the multiple exemplar condition, where they outperformed two-year-olds. We propose that comparing multiple pictures of to-be-learned referents strengthens encoding of category-defining shape, facilitating extension of labels to objects from memory.

*Keywords:*

Word learning; Pictures; Multiple Exemplars; Retention; Generalisation

**Do multiple exemplars promote preschool children’s retention and generalisation of words learned from pictures?**

Pictures are an invaluable resource for supporting language development – they enable children to learn names for objects without the need for direct experience. Many pictures represent categories, affording children the opportunity to learn labels for multiple objects in their environment. However, successful learning from pictures hinges on children’s understanding that information associated with symbolic representations, including verbal labels, actually refers to independently existing referents (i.e., pictures themselves are not the exclusive intended referent). In order to truly learn a word from a picture, children must also retain and appropriately extend labels to symbolised referents in the absence of the picture for comparison. Previous studies of word learning from objects have shown that associating a novel label with multiple exemplars of a referent can facilitate both retention (Twomey et al., 2014) and generalisation (Perry et al., 2010) by highlighting crucial category-relevant details (e.g., shape) and diverting attention away from category-irrelevant details (e.g., colour). Here, for the first time, we investigate whether two- and three-year-olds’ ability to extend labels from memory representations of pictures is facilitated by exposure to multiple exemplars during teaching.

As infants are immersed in a world of pictures, it is perhaps unsurprising that they are capable of extending labels from pictures to symbolised objects by their second year of life (Ganea et al., 2008; Preissler & Carey, 2004). In Preissler and Carey (2004), 18- and 24-month-olds were taught the name for a black-and-white line drawing of a novel object. Immediately after mapping the label, children were tasked with identifying the referent from an array including the picture and the previously unseen symbolised object. Both age groups consistently extended the label to the symbolised object, demonstrating awareness that the drawing was a representation of the intended referent. When learning from highly iconic colour photographs, even younger children are capable of immediately extending labels and other properties from 2D representations to 3D objects (Ganea et al., 2008; Geraghty et al., 2014; Keates et al., 2014; Khu et al, 2014). In Geraghty et al. (2014), 15- and 17-month-olds were taught a novel word in association with photographs of two differently coloured variants of the same unfamiliar object (e.g. a purple whisk and an orange whisk). At test, infants in both age groups extended the novel word to another differently coloured photograph of the unfamiliar object (e.g. a silver whisk) and the 3-D depicted object (an actual silver whisk), indicating their formation of a broad representational category that was not constrained by colour or dimensionality (also see Hartley & Allen, 2014, 2015b). Infants aged 13- to 18-months also attempt to activate hidden functions of 3-D objects after they have been revealed by colour photographs, demonstrating transfer of non-linguistic information from pictures to referents. Together, these findings suggest that children recognise word-picture-object relationships from very early in life.

Most studies investigating how children learn words from pictures have employed a similar approach: children are taught a novel word-picture association and then immediately tested on their ability to extend the label to a symbolised referent, usually with the picture visible for comparison. However, this task does not test true word learning, nor does it reflect how children usually learn from pictures under naturalistic conditions (Carter & Hartley, 2021). True word learning not only requires successful identification of meaning within the context of a naming event, but also encoding and retention of a word-referent representation in memory that can subsequently be retrieved and generalised (Hartley et al., 2019, 2020; McMurray et al., 2012). Critically, evidence suggests that identification, retention, and generalisation of meaning represent distinct steps in the process of learning a word. In Horst and Samuelson (2008), two-year-olds were presented with an unfamiliar object alongside two familiar objects and were asked to identify the referent for a novel word (a referent selection task). Under these circumstances, children can apply a range of heuristics and biases to correctly identify the unfamiliar object as the intended referent including the mutual exclusivity principle (the assumption that each referent only has a single label; Carey, 1978), the novel name-less category constraint (the assumption that a novel label refers to a novel object, rather than a familiar object with a known label; Golinkoff et al., 1994; Mervis & Bertrand, 1994), and general attentional preferences for novel objects (Mather & Plunkett, 2012). Despite accurately identifying the meanings of the novel words, Horst and Samuelson’s participants responded with chance-level accuracy when their retention of the word-object pairings was tested just 5 minutes later, demonstrating that accurate referent selection does not guarantee retention. Moreover, despite being capable of mapping word-referent associations from just 6 months of age (Friedrich & Friederici, 2011), children do not consistently generalise labels based on category-defining shape until approximately 24 months (Landau et al., 1988; Taxitari et al., 2020). Through building a vocabulary comprised of names for object categories that are distinguished by shape (rather than other perceptual features, such as colour or texture), children infer that shape determines what a thing is and, therefore, the label it should receive (they develop a ‘shape bias’; Gershkoff-Stowe & Smith, 2004; Samuelson & Smith, 1999). Thus, in order to truly assess children’s learning from pictures, it is vital to measure retention *and* generalisation of word-picture-object relationships.

To our knowledge, just a single study to date has investigated children’s extension of novel words from representations of pictures retained in memory. In Carter and Hartley (2021), children aged 30-53 months accurately mapped novel labels to colour photographs and black-and-white line drawings depicting unfamiliar objects in a referent selection task. After a 5-minute delay, children were presented with sets of three 3-D unfamiliar objects that were symbolised by pictures during training and asked to identify the referents of the newly-learned words. Testing children’s understanding of novel labels in the absence of associated pictures more accurately represents children’s naturalistic learning; a child is likely to hear the name for an unfamiliar picture on one occasion, and then be required to retrieve the name when they experience the symbolised referent on another occasion. Crucially, the results revealed that the children failed to extend labels to symbolised referents with above-chance accuracy and were not significantly influenced by iconicity.

Considering evidence that three-year-olds can retain labels paired with objects with above-chance accuracy after 5 minutes (Cheung et al., 2022), Carter and Hartley’s (2021) findings suggest that retaining words learned from pictures may be relatively more challenging for young children. They propose that young children’s difficulty extending labels from pictures to symbolised referents from memory may be explained by the scarcity of information provided by individual 2-D representations relative to 3-D objects. Although young children are capable of generating mental representations of symbolised referents from brief exposures to pictures (Hartley & Allen, 2015a), it may be that 2-D sensory input afforded by a single picture is insufficient to encode a robust word-picture-object representation that can be easily retrieved after a 5-minute delay.

Informed by research investigating word learning from objects, it may be that children’s retention and generalisation of word-picture-object relationships can be facilitated by teaching with multiple exemplars. In Twomey et al. (2014), two-year-olds accurately mapped three novel word-object pairings in a referent selection task. Each of the three novel objects was encountered on three training trials; in the single exemplar condition the same objects were presented repeatedly (i.e., children saw one variant of each referent), whereas in the multiple exemplar condition the objects differed in colour on each trial (i.e., children saw three variants of each referent). When retention of word-object pairings was tested after a 5-minute delay, only children in the multiple exemplar condition responded with above-chance accuracy (also see Aguilar et al., 2018). In Perry et al.’s (2010) longitudinal study, 18-month-olds generalised novel labels with significantly greater accuracy if they were trained with three distinctive category exemplars rather than three similar category exemplars (also see Geraghty et al. (2014) and Vukatana et al. (2015)). Teaching with multiple exemplars is proposed to benefit children’s learning because pairing an individual novel word with multiple variants of the referent with distinct category-irrelevant features (e.g., different colours) promotes a comparative process that increases attentional salience – and strengthens encoding – of category-defining shape (Gentner & Namy, 1999; Graham et al., 2010; Namy & Gentner, 2002).

The objective of this study was to discover whether young children extend labels from pictures to objects from memory more accurately when taught via multiple exemplars versus single exemplars. Samples of preschool children aged two and three years mapped a series of novel word-picture pairings in a referent selection task. In one condition, children were presented with a single variant of each novel picture during training. In another condition, children were presented with two differently coloured variants of each novel picture during training. Children’s retention and generalisation of word-picture-object relationships was then assessed after a 5-minute delay by asking children to identify 3-D object referents for each novel label. Based on previous evidence regarding object name learning (e.g., Perry et al., 2010; Twomey et al., 2014), we predicted that children would retain and generalise labels with significantly greater accuracy in the multiple exemplar condition. We also tentatively predicted that three-year-olds may extend labels to symbolised referents with significantly greater accuracy than two-year-olds due to their increased experience of learning from pictures (Callaghan et al., 2011) and more developed memory mechanisms subserving general word learning (Cheung et al., 2022). Importantly, our results will advance theoretical understanding of how children’s learning from pictures can be facilitated at early ages, potentially informing the design and development of educational resources (e.g., picture books).

**Method**

**Participants**

Based on the most relevant data that informed this research, our sample sizes were determined by power simulations performed using ‘Power ANalysis for GEneral Anova’ (PANGEA, Version 0.2; Westfall, 2015). We aimed to detect a key finding reported in Twomey et al. (2014): that neurotypical two-year-olds retain novel words with significantly greater accuracy when taught via multiple exemplars than single exemplars. Twomey et al. utilised a between-subjects design in which 12 children were assigned to the single exemplar condition and 12 were assigned to the multiple exemplar condition. Their data showed that children who encountered multiple exemplars retained novel words with significantly greater accuracy (*M* = 0.72, *SD* = 0.27) than children who encountered the same exemplars repeatedly (*M* = 0.47, *SD* = 0.32) – a large effect size (*d* = 0.84). Predicting that both 2- and 3-year-olds in our study would achieve higher word learning accuracy in the multiple exemplar condition, we required 16 participants per age group to detect this effect with 80% power in our within-subjects design. However, insufficient data exists to inform accurate power simulations for our interaction effect. Although previous studies indicate large differences between the word learning abilities of 2- and 3-year-olds (e.g. Vlach & Sandhofer, 2011), providing confidence in the power afforded by our sample sizes, we recommend that our results associated with age should be regarded as exploratory and interpreted with some degree of caution.

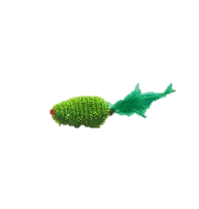
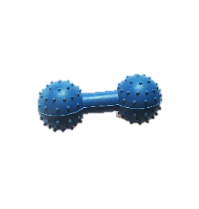
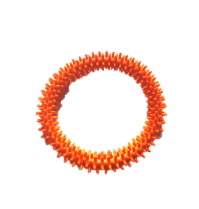
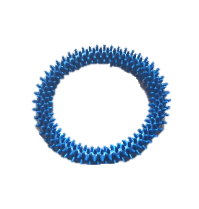
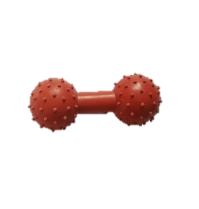
The final sample for this study included 42 neurotypical children, including 23 two-year-olds (*M* age = 2;7, *SD* = 0;3, 10 girls) and 19 three-year-olds (*M* age = 3;5, *SD* = 0;3, 11 girls). Children were recruited from four nurseries located around Warrington, UK. All children were from monolingual English homes, with no history of developmental or sensory disorders. An additional three two-year-olds participated in the study but were not included in the final sample due to consistently inaccurate performance on warm-up trials, indicating their failure to comprehend the task. Informed consent was obtained from parents prior to their children’s participation in the study. We tested all children for whom we received informed consent, ceasing recruitment after our planned sample sizes were reached in the final nursery. All procedures performed in this study were in accordance with the ethical standards of institutional and national research committees. The ethics approval reference number for this project is FST-2023-3928-2.

**Materials**

Stimuli included eight novel words (e.g., blicket) selected from the NOUN database (Horst & Hout, 2016), familiar objects, unfamiliar objects, and colour photographs of familiar and unfamiliar objects (sized 8 cm x 8 cm). Familiar objects were selected on the basis that most children understand their linguistic labels by 20 months (Fenson et al., 1994). Unfamiliar objects were selected on the basis that children would not know their linguistic labels. Six different familiar objects were used in warm-up trials (three per condition, counterbalanced; 1. Car, elephant, crayon, 2. Ball, monkey, plane).

Unfamiliar objects in the single exemplar condition were divided into four sets (see Figure 1). Each set included a to-be-named ‘target object’ and a ‘shape match’ (a differently-coloured version of the target object that was presented in generalisation trials). Each shape match was the same colour as a target object from a different set, and therefore served as a ‘colour match’ in generalisation trials. Colour photographs depicting the four target objects were printed and laminated.

**Figure 1.** Sets of unfamiliar objects used in single exemplar trials



**Named object**

**Shape match**

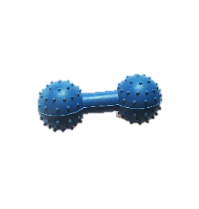
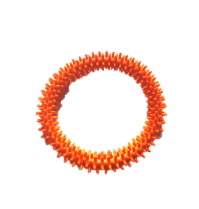
**Colour match**

**Set 1**

**Set 2**

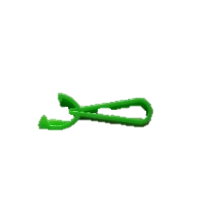
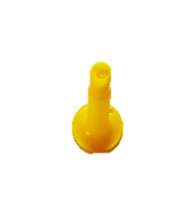
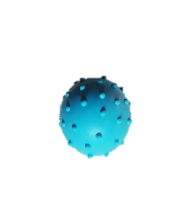
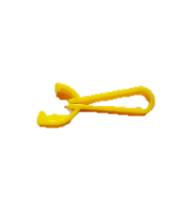
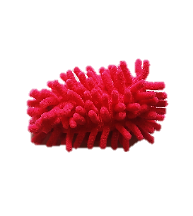
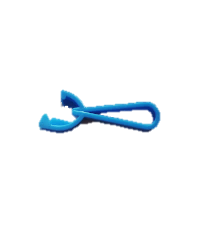
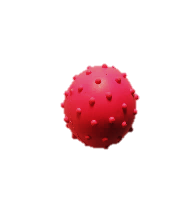
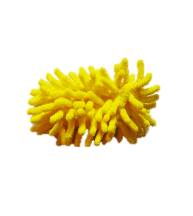
**Set 3**

**Set 4**



Unfamiliar objects in the multiple exemplar condition were also divided into four sets (see Figure 2). Each set included three differently coloured variants of a target object – two were labelled during referent selection trials via pictures, and one served as a shape match in generalisation trials. As for stimuli in the single exemplar condition, each shape match was the same colour as a target object from a different set, and therefore served as a ‘colour match’ in generalisation trials. Colour photographs depicting two of the colour variants for each target object were printed and laminated (the two presented in referent selection trials).

**Figure 2.** Sets of unfamiliar objects used in multiple exemplar trials



**Named object**

**version 1**

**Shape match**

**Colour match**

**Named object**

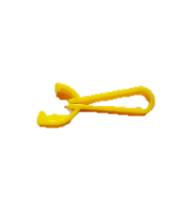
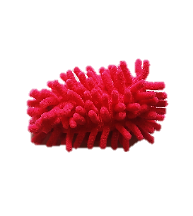
**version 2**

**Set 1**

**Set 2**

**Set 3**

**Set 4**



Colour photographs depicting 16 familiar objects were divided into pairs, with four pairs allocated to each condition (teddy bear, lion, sock, bird, clock, house, bike, apple, carrot, flower, tree, cat, balloon, hat, banana, key). These pairs of photographs depicting familiar objects were presented alongside the photographs of unfamiliar target objects in referent selection trials.

**Procedure**

Our experimental task was very similar to those reported by Hartley et al. (2019) and Carter and Hartley (2021). Participants were tested individually in their nurseries and were accompanied by a familiar member of nursery staff. Every child completed both the single exemplar condition and multiple exemplar condition on different days, approximately 2 days apart (order counterbalanced across participants). Children were verbally praised for attention and good behaviour. Corrective reinforcement was only provided in the warm-up trials. The task consisted of the following stages delivered in a fixed order: 1. warm-up trials, 2. referent selection trials, 3. shape/colour match familiarisation, 4. 5-minute delay, and 5. retention and generalisation trials.

**Warm-up trials.** The task began with three warm-up trials. On each trial, three familiar objects were placed in a row in front of the child. The experimenter waited silently for approximately 3 seconds before requesting one of the objects (e.g., “Which is the car? Show me the car.”). If the child was reluctant to respond, they were prompted up to two additional times by the experimenter and a further two times by the familiar nursery staff member. Following a correct response, the experimenter issued praise and reinforced the identity of the object (e.g., “Well done, that is the car!”). Following an incorrect response, the experimenter provided corrective feedback (e.g., “Actually, this is the car. Can you touch the car? Well done, you touched the car!”). Following the child’s response, the experimenter re-ordered the objects and requested a different object in a different row position until all three objects had been requested.

**Referent selection trials.** Sixteen referent selection trials were administered immediately after the warm-up trials and had an identical format, except that the stimuli were colour photographs rather than objects and children did not receive feedback related to their responses (the experimenter just said “thank you”). Children were presented with four sets of pictures, each including a photograph depicting one unfamiliar object and two photographs depicting familiar objects (see Figure 3). Each set was presented on four trials. On two trials per set, the experimenter requested the photograph of the unfamiliar object (‘novel trials’; e.g., “Which is the blicket? Show me the blicket.”) and on two trials they requested the photographs of familiar objects (‘familiar trials’; “Which is the car? Show me the car.”). Familiar trials were included to encourage children to examine every item in the array and prevent them from developing a strategy of repeatedly selecting the most novel object on every trial without considering referential meaning (see Horst et al., 2011). Novel trials were designed to promote active learning of new word-picture pairings. As children were likely to know names for the two familiar photographs, they should infer that the novel label refers to the photograph depicting the unfamiliar object (Carey, 1978).

**Figure 3.** Example stimuli presented at each stage of the word learning task. Images with black borders represent 2-D pictures and images without black borders represent 3-D objects. The target referent is positioned in the middle for referent selection trials, on the left for the retention trial, and on the right for the generalisation trial

**Referent selection**

**Trial 1**

**Retention trial**

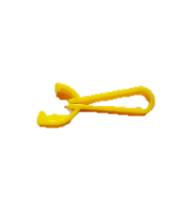
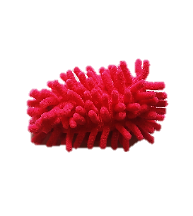
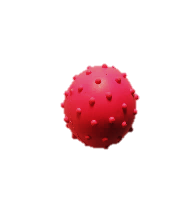
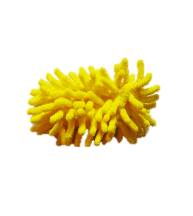
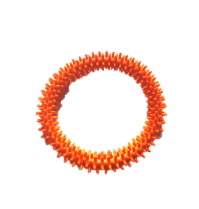
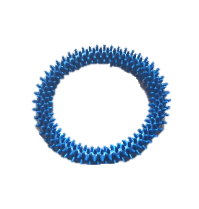
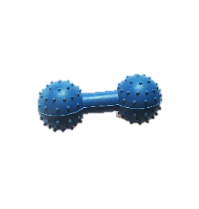
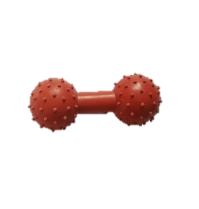
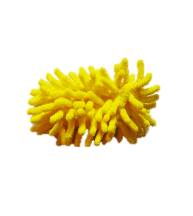
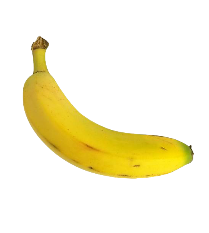
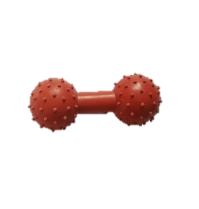
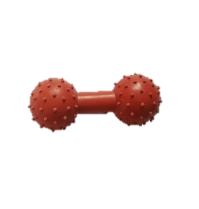
**Generalisation trial**

**Single Exemplar Condition**

**Referent selection**

**Trial 2**

**Multiple Exemplar Condition**



In the single exemplar condition, the photograph depicting the unfamiliar object was identical on all four trials per set (i.e., the same unfamiliar object was requested on both novel trials). In the multiple exemplar condition, the photographs depicting unfamiliar objects differed in colour (but not shape) within sets. For example, on one novel trial and one familiar trial a particular unfamiliar object was yellow, but on the other novel trial and familiar trial the same depicted unfamiliar object was purple. Thus, children had the opportunity to map the same novel label to two differently-coloured exemplars of each unfamiliar object.

The order of trials was pseudo-randomised with the constraint that the same set of objects was never presented on more than two consecutive trials and no more than two trials of the same type (familiar or novel) were experienced sequentially. Positioning of objects in the row (left, middle, right) was pseudo-randomised across trials with the constraint that the requested object did not appear in the same location more than twice consecutively. Novel word-object pairings were randomly generated for each participant.

**Shape/colour match familiarisation.** Immediately after the referent selection trials, participants were familiarised with photographs depicting the as-yet-unseen shape/colour matches, prior to the appearance of the corresponding objects in generalisation trials. The purpose of this stage was to minimise novelty and familiarity preferences, increasing the likelihood that children would select objects based on their memory of word-picture mappings. The experimenter presented a photograph of an unfamiliar shape/colour match object and simply instructed the child to “look”. Children were allowed to touch the photograph if they wished. After approximately 5s, the experimenter removed the photograph from view and presented the next photograph until images of all shape/colour match objects for the condition had been seen. The order in which photographs of shape/colour match objects were presented was randomised for each child.

**Delay.** After presenting the final shape/colour match photograph, all experimental stimuli were removed from the child’s view and a 5-minute delay was initiated. During this period, the experimenter made conversation and played an unrelated game with the child.

**Retention and generalisation trials.** To re-engage the participant’s attention, the experimenter administered one warm-up trial with familiar objects as described above. This was immediately followed by eight retention trials and eight generalisation trials (see Figure 3). Each novel word was tested on two retention trials and two generalisation trials, yielding 16 test trials in total per condition. For retention trials, three unfamiliar 3-D objects that were symbolised by photographs in the referent selection trials were presented in a row and the experimenter requested one of the objects (e.g., “Which is the blicket? Show me the blicket.”). In the single exemplar condition, the object associated with a particular novel word was the same colour on both retention trials. In the multiple exemplar condition, both colour variants that were named in referent selection trials were requested on one retention trial each. The purpose of retention trials was to assess whether children would extend the novel labels associated with photographs during referent selection to their corresponding 3-D objects. For generalisation trials, children were presented with a differently coloured unlabelled variant of the target object (shape match), an object that matched the target object on colour but not shape (colour match), plus a shape/colour match object from another set. Children were asked to identify which of these novel objects was a referent for the word paired with the photographs of the target object during referent selection (e.g., “Which is the blicket? Show me the blicket.”). The purpose of these trials was to assess whether children’s generalisation of labels from photographs to novel depicted referents is systematically influenced by shape. Importantly, all objects were of equivalent familiarity in both retention trials (all were objects that matched labelled photographs on shape and colour) and generalisation trials (all were objects symbolised by pictures introduced during the familiarisation stage and were differently coloured, unlabelled, variants of target objects).

To provide the necessary level of control when presenting stimuli, object groupings were fixed. The order of trials was pseudo-randomised with the constraint that the same set of objects was never presented on more than two consecutive trials and no more than two trials of the same type (retention, generalisation) were experienced sequentially. Positioning of objects in the row (left, middle, right) was pseudo-randomised across trials with the constraint that the requested object did not appear in the same location more than twice consecutively.

**Results**

**Referent selection**

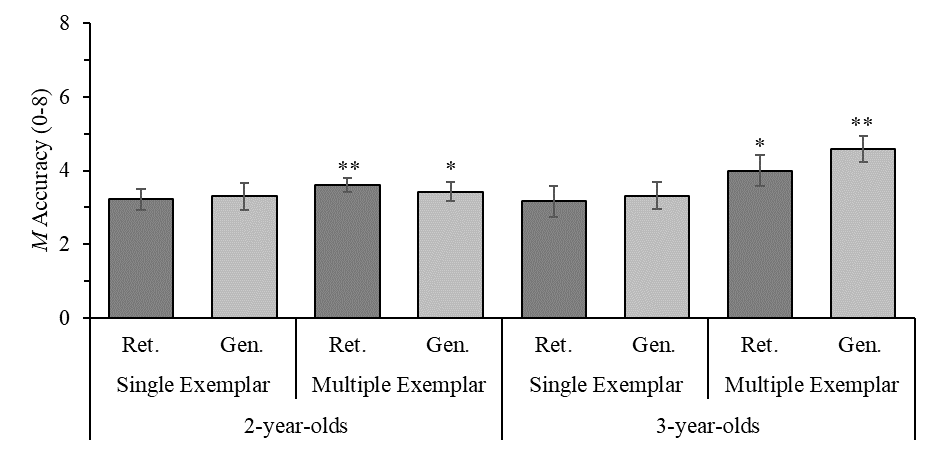
Participants were scored out of eight on novel and familiar referent selection trials (see Figure 4). One sample *t*-tests were used to compare children’s response accuracy against a chance level of 2.64 (0.33 chance rate per trial multiplied by eight). Both age groups responded with above-chance accuracy on familiar and novel trials in the single and multiple exemplar conditions (all *p* values < .001). Children’s referent selection accuracy data were entered into a 2(Age Group: 2 years, 3 years) x 2(Condition: single exemplar, multiple exemplars) x 2(Trial Type: familiar, novel) mixed ANOVA. The results revealed a significant main effect of Trial Type, *F*(1, 40) = 20.45, *p* < .001, ηp2 = .34, indicating that children across age groups and conditions responded significantly more accurately on familiar trials (*M* = 7.92) than novel trials (*M* = 7.48). No other main effects or interactions were significant (Age Group: *p* = .10; Condition: *p* = .45; Age Group x Condition: *p* = .32; Age Group x Trial Type: *p* = .60, Condition x Trial Type: *p* = .41; Age Group x Condition x Trial Type: *p* = .31).

**Figure 4.** Average referent selection accuracy for two-year-olds and three-year-olds in the single and multiple exemplar conditions. Error bars show ± 1 SE. Below-ceiling performance significantly exceeded chance level accuracy (2.64) with *p* < .001 for all measures

**Retention trials**

As retention and generalisation represent distinct word learning processes, we analysed children’s responses on these trial types separately. Participants were scored out of eight on retention (each word was tested twice; see Figure 5). Children responded correctly if they selected the object represented by the picture that was assigned the requested label during Referent Selection.

**Figure 5.** Average retention and generalisation accuracy for two-year-olds and three-year-olds in the single and multiple exemplar conditions. Error bars show ± 1 SE. Dashed line represents chance (2.64). Stars above columns indicate where response accuracy was significantly more accurate than expected by chance (\* *p* < .01; \*\* *p* < .001)



One sample *t*-tests were used to compare children’s response accuracy against a chance level of 2.64. Two-year-olds’ retention accuracy was above chance in the single exemplar condition (*p* = .052) and the multiple exemplar condition (*p* < .001). Three-year-olds’ retention accuracy did not differ from chance in the single exemplar condition (*p* = .22), but was above chance in the multiple exemplar condition (*p* = .004).

Children’s retention accuracy scores were entered into a 2(Age Group: 2 years, 3 years) x 2(Condition: single exemplar, multiple exemplars) mixed ANOVA, which revealed a significant main effect of Condition, *F(*1, 40) = 5.76, *p* = .021, ηp2 = .13. Across age groups, children responded significantly more accurately in the multiple exemplar condition (*M* = 3.80) than the single exemplar condition (*M* = 3.19). No other main effects or interactions were significant (Age Group: *p* = .67; Age Group x Condition: *p* = .39).

**Generalisation trials**

Participants were scored out of eight on generalisation trials (each word was tested twice; see Figure 5). Children responded correctly if they selected the shape match object corresponding to the picture that was assigned the requested label.

One sample *t*-tests were used to compare children’s generalisation accuracy against a chance level of 2.64. Two-year-olds’ generalisation accuracy did not differ from chance in the single exemplar condition (*p* = .081), but was above chance in the multiple exemplar condition (*p* = .007). Three-year-olds’ generalisation accuracy was above chance in the single exemplar condition (*p* = .076) and the multiple exemplar condition (*p* < .001).

Children’s generalisation accuracy scores were entered into a 2(Age Group: 2 years, 3 years) x 2(Condition: single exemplar, multiple exemplars) mixed ANOVA. The main effect of Age Group was not significant (*p* = .13). A significant main effect of Condition, *F(*1, 40) = 5.31, *p* = .027, ηp2 = .12, was qualified by a Condition x Age interaction that approached statistical significance, *F(*1, 40) = 3.51, *p* = .068, ηp2 = .08. We proceeded to deconstruct the relationship between Age and Condition, though note that these comparisons should be treated with caution as the interaction was marginally significant. Pairwise comparisons showed that three-year-olds generalised newly-learned words with significantly greater accuracy in the multiple exemplars condition (*M* = 4.58) than the single exemplar condition (*M* = 3.32), *t*(18) = 3.13, *p* = .006, *d* = 0.71, while the accuracy of two-year-olds did not significantly differ between the multiple exemplars (*M* = 3.43) and single exemplar conditions (*M* = 3.30; *t* = 0.30, *p* = .77). The three-year-olds responded significantly more accurately than two-year-olds in the multiple exemplars condition (*t* = 2.64, *p* = .012, *d* = 0.81), but not the single exemplar condition (*t* = 0.22, *p* = .98).

**Discussion**

This study investigated whether two- and three-year-olds’ ability to retain and generalise novel labels learnt from pictures is enhanced by teaching with multiple exemplars. During training, both age groups mapped novel word-picture pairings with comparable accuracy that significantly exceeded chance. After a 5-minute delay, in the single exemplar condition, neither age group’s response accuracy significantly exceeded chance on retention trials or generalisation trials (which required children to extend labels to similarly coloured or differently coloured 3-D referents, respectively). However, children responded significantly more accurately on retention trials in the multiple exemplar condition, with both groups significantly exceeding chance-level performance. While both age groups also responded with above-chance accuracy on generalisation trials in the multiple exemplar condition, three-year-olds significantly benefited from teaching with multiple exemplars relative to single exemplars and responded with significantly greater accuracy than two-year-olds. Together, these findings clearly demonstrate that associating words with multiple variants of pictures increases the likelihood that two- and three-year-olds will successfully extend labels to symbolised referents from memory.

The lack of age or condition effects at referent selection was unsurprising. In line with previous studies (e.g., Carter & Hartley, 2021), two- and three-year-olds accurately identified that pictures of unfamiliar objects were the intended referents for novel words. By 24-months, children reliably utilise a range of lexical heuristics and attentional biases to correctly identify novel objects as the intended referents of novel labels (Carey, 1978; Mather & Plunkett, 2012; Mervis & Bertrand, 1994). This is demonstrated by the relatively small difference between two- and three-year-olds’ average accuracy on novel referent selection trials (0.92 vs. 0.95). Also, our participants responded with significantly greater accuracy on trials requesting referents for familiar words than novel words. This trial type effect can be explained by children’s pre-existing representations of referents for familiar words, meaning that accurate responses on these trials simply required recognition of a known object rather than mapping a new word-referent relationship.

In the single exemplar condition, neither age group extended labels to depicted referents with above-chance accuracy when their retention of word-picture relationships was tested after 5 minutes. In light of evidence that children realise that labels associated with pictures refer to independently existing 3-D referents by 15 months (Ganea et al., 2008; Geraghty et al., 2014), it is unlikely that our participants’ inaccurate responding was due to insufficient symbolic understanding. Rather, it may be that word-picture representations encoded in the single exemplar condition were relatively fragile and vulnerable to decay over a short delay. Our participants’ performance in the single exemplar condition aligns with existing findings that preschool children can struggle to retain labels paired with individual objects (Horst & Samuelson, 2008) or pictures (Carter & Hartley, 2021) when tested after 5 minutes rather than immediately. By contrast, both age groups responded above-chance and significantly more accurately on retention trials in the multiple exemplar condition. This result is congruent with evidence reported in Twomey et al. (2014) that two-year-olds only retain novel words with above-chance accuracy after 5 minutes when taught with multiple differently-coloured referent variants. Notably, our sample of two-year-olds achieved above-chance retention for four words, rather than three as in this previous study. Our participants’ strong performance in the multiple exemplar condition cannot be explained by differences in statistical input, as both conditions presented the same number of exposures to word-picture pairings during referent selection. The key difference between conditions concerned the extent to which *variability* during teaching facilitated encoding of mental representations that enabled extension to 3-D referents from memory.

Informed by categorisation theories, we propose that teaching novel word-picture associations with multiple exemplars enhances retention through comparison of referent variants (e.g., Gentner & Namy, 1999; Graham et al., 2010; Namy & Gentner, 2002). Comparing differently coloured pictures that share the same label across trials simultaneously highlights common features (i.e., global shape) and reduces attention to unique attributes that vary between exemplars (e.g., colour). This is particularly valuable when learning from iconic pictures which relate to their independently existing referents by virtue of shape-based resemblance (Neander, 1987). Consequently, the child’s mental representation of the depicted object’s shape – its category-defining quality – may be strengthened, facilitating accurate label extension following a delay. Indeed, previous studies have demonstrated that encouraging children to focus on important, rather than irrelevant, stimuli can enhance their retention of novel word-object associations (e.g., Axelsson et al., 2012). By contrast, teaching children novel labels via individual pictures may encourage children to code more complex representations of depicted referents characterised by a specific shape, *plus* other category-irrelevant details (e.g., colour and texture), that are less easily retained (Twomey et al., 2014). Whereas children only required two fast mapping trials in the multiple exemplar condition to achieve above-chance retention and extension to symbolised referents, learning from single exemplars that do not cue children’s attention to shape appears to be less efficient and may require additional exposures to achieve similar performance.

We observed that both two- and three-year-olds generalised labels to differently coloured 3-D referents with above-chance accuracy after a 5-minute delay when taught with multiple exemplars, but not single exemplars. Thus, for both age groups, affording the opportunity to compare just two picture variants was sufficient to promote encoding of a category representation that could be generalised to novel 3-D objects. We also detected that exposure to multiple exemplars at referent selection significantly benefited three-year-olds’ generalisation accuracy, enabling them to respond more accurately than 2-year-olds. This effect may be explained by age-related differences in expressive vocabulary development. The shape bias emerges at approximately 24 months and coincides with children’s acquisition of approximately 50-150 count nouns (Landau et al., 1988; Samuelson & Smith, 1999). As many object categories are organised by shape, the process of learning object names selectively tunes children’s attention, training their ability to acquire shape-based lexical categories (Gershkoff-Stowe & Smith, 2004). As expressive vocabulary size increases rapidly between 2 and 3 years (Fenson et al., 1994), so too does children’s experience of mapping and generalising labels on the basis of shape. Thus, 3-year-olds’ larger vocabularies and greater depth of word learning experience may have enhanced their ability to capitalise on multiple exemplars cueing their attention to shape, which in turn facilitated their generalisation of novel labels from colour photographs.

The findings from this study have implications for the design of picture-based educational resources. From early in infancy, children spend considerable time engaged in picture book interactions with caregivers (DeBaryshe, 1993). In this context, adults may name pictures with the implicit assumption that children will generalise information to depicted objects when they are encountered. Our data show that, despite understanding that pictures represent 3-D referents, preschool children may struggle to extend information to symbolised objects from memory when taught with single exemplars. By contrast, the advantages displayed by both two- and three-year-olds in our multiple exemplars condition suggest that visual media designed to promote vocabulary acquisition should present differently coloured images of to-be-learned referents. Advancing beyond previous studies that report word learning benefits from comparing three or more exemplars (e.g., Perry et al., 2010; Twomey et al., 2014), here we demonstrate that presenting just two variants is sufficient to enhance pre-schoolers’ retention and generalisation of labels acquired from pictures. We recommend that future research investigates whether this strategy may also be advantageous when teaching communication skills to minimally verbal children via picture-based augmentative and alternative communication interventions, such as the Picture Exchange Communication System (Frost & Bondy, 2002).

This study is not without limitations. Firstly, if children’s retention accuracy was enhanced by comparing differently coloured target pictures across referent selection trials, then we may expect that individual differences in categorisation would predict their performance in the multiple exemplar condition. Including an independent categorisation task would have enabled us to test this hypothesis, potentially strengthening our theoretical explanation. Secondly, we propose that three-year-olds generalised labels with significantly greater accuracy than two-year-olds in the multiple exemplar condition due to their increased expressive vocabulary size, but the lack of a standardised vocabulary assessment prevents us from confirming this. Finally, while our data show that diversifying target referents can benefit young children’s learning from pictures, this represents only one source of variability. Increasing the variability of other aspects of children’s environments – such as increasing the number or diversity of competitors – has been shown to hinder word learning from objects (Horst et al., 2010; Lester et al., 2023). To optimise educational resources, additional research is required to investigate how other non-target sources of variability influence young children’s word learning from pictures.

Overall, this research has advanced theoretical understanding of how young children learn words from pictures by assessing their knowledge of word-picture-object relationships beyond immediate referent selection. When testing extension of labels from pictures to objects from memory 5 minutes after mapping – arguably a more valid reflection of children’s “real world” picture learning – two- and three-year-olds respond with significantly greater accuracy when taught with multiple exemplars. We propose that the process of comparing referent variants across trials strengthens encoding of category-defining shape, enhancing both retention and generalisation of word-picture associations. Three-year-olds’ also outperformed two-year-olds on generalisation trials in the multiple exemplar condition, perhaps indicating that children’s ability to capitalise on cues highlighting shape increases with receptive vocabulary development. On an applied level, these findings suggest that picture-based educational resources would benefit from diversifying representations of to-be-learned referents.

**Acknowledgements**

We wish to thank the children, parents, and staff at Puddleducks Day Nursery, Rainbow Day Nursery, Cheshire Day Nursery Thelwall, and Stretton Day Nursery.

**References**

Aguilar, J. M., Plante, E., & Sandoval, M. (2018). Exemplar variability facilitates retention of word learning by children with specific language impairment. *Language, speech, and hearing services in schools*, *49*(1), 72-84. http://dx.doi.org/10.1044/2017\_LSHSS-17-0031

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.

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., ... & Collins, W. A. (2011). Early social cognition in three cultural contexts. *Monographs of the society for research in child development*, i-142.

Carey, S. (1978). *The child as word learner*. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), Linguistic theory and psychological reality (pp. 264–293). Cambridge, MA: MIT Press.

Carter, C. K., & Hartley, C. (2021). Are children with autism more likely to retain object names when learning from colour photographs or black-and-white cartoons? *Journal of Autism and Developmental Disorders*, *51*(9), 3050-3062. http://dx.doi.org/10.1007/s10803-020-04771-2

Cheung, R. W., Hartley, C., & Monaghan, P. (2022). Multiple Mechanisms of Word Learning in Late-Talking Children: A Longitudinal Study. *Journal of Speech, Language, and Hearing Research*, *65*(8), 2978-2995. https://doi.org/10.1044/2022\_JSLHR-21-00610

DeBaryshe, B. D. (1993). Joint picture-book reading correlates of early oral language skill. *Journal of Child Language, 20*, 455–461. http://dx.doi.org/10.1017/S0305000900008370.

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*, 1–185. http://dx.doi.org/10.2307/1166093.

Friedrich, M., & Friederici, A. D. (2011). Word learning in 6-month-olds: fast encoding–weak retention. *Journal of Cognitive Neuroscience*, *23*(11), 3228-3240. http://dx.doi.org/10.1162/jocn\_a\_00002

Frost, L., & Bondy, A. (2002). *PECS training manual*. Newark, DE: Pyramid Educational Consultants

Ganea, P. A., Pickard, M. B., & DeLoache, J. S. (2008). Transfer between picture books and the real world by very young children. *Journal of cognition and development*, *9*(1), 46-66. http://dx.doi.org/10.1080/15248370701836592

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive development*, *14*(4), 487-513. http://dx.doi.org/10.1016/S0885-2014(99)00016-7

Geraghty, K., Waxman, S. R., & Gelman, S. A. (2014). Learning words from pictures: 15-and 17-month-old infants appreciate the referential and symbolic links among words, pictures, and objects. *Cognitive Development*, *32*, 1-11. https://doi.org/10.1016/j.cogdev.2014.04.003

Gershkoff‐Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child development*, *75*(4), 1098-1114. http://dx.doi.org/10.1111/j.1467-8624.2004.00728.x

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*, *28*(1), 99. http://dx.doi.org/10.1037/0012-1649.28.1.99

Graham, S. A., Namy, L. L., Gentner, D., & Meagher, K. (2010). The role of comparison in preschoolers’ novel object categorisation. *Journal of Experimental Child Psychology*, *107*(3), 280–290.

Hartley, C., & Allen, M. L. (2014). Generalisation of word-picture relations in children with autism and typically developing children. *Journal of Autism and Developmental Disorders*, *44*, 2064–2071. https://doi.org/10.1007/s10803-014-2074-1

Hartley, C., & Allen, M. L. (2015a). Iconicity influences how effectively minimally verbal children with autism and ability-matched typically developing children use pictures as symbols in a search task. *Autism*, *19*, 570–579. https://doi.org/10.1177/1362361314 536634

Hartley, C., & Allen, M. L. (2015b). Symbolic understanding of pictures in low-functioning children with autism: the efects of iconicity and naming. *Journal of Autism and Developmental Disorders*, *45*, 15–30. https://doi.org/10.1007/s10803-013-2007-4

Hartley, C., Bird, L., & 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.cogni tion.2019.03.001

Hartley, C., Bird, L., & Monaghan, P. (2020). Comparing cross-situational word learning, retention, and generalisation in children with autism and typical development. *Cognition*, *200*, 1–20. https://doi. org/10.1016/j.cognition.2020.104265

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

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

Horst, J. S., Samuelson, L. K., Kucker, S. C., & McMurray, B. (2011). What’s new? Children prefer novelty in referent selection. *Cognition*, *118*(2), 234-244. https://doi.org/10.1016/j.cognition.2010.10.015

Horst, J. S., Scott, E. J., & Pollard, J. A. (2010). The role of competition in word learning via referent selection. *Developmental Science*, *13*, 706–713. https://doi.org/10.1111/j.1467-7687.2009.00926.x

Keates, J., Graham, S. A., & Ganea, P. A. (2014). Infants transfer nonobvious properties from pictures to real-world objects. *Journal of Experimental Child Psychology*, *125*, 35-47. https://doi.org/10.1016/j.jecp.2014.02.003

Khu, M., Graham, S. A., & Ganea, P. A. (2014). Learning from picture books: infants’ use of naming information. *Frontiers in psychology*, *5*, 144. https://doi.org/10.3389/fpsyg.2014.00144

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*(3), 299–321. https://doi.org/10.1016/0885-2014(88)90014-7

Lester, N., Theakston, A., & Twomey, K. E. (2023). The role of the museum in promoting language word learning for young children. *Infant and Child Development*, e2400. https://doi.org/10.1002/icd.2400

Mather, E., & Plunkett, K. (2012). The role of novelty in early word learning. *Cognitive Science*, *36*(7), 1157-1177. https://doi.org/10.1111/j.1551-6709.2012.01239.x

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*, 831–877. https ://doi.org/10.1037/a0029872

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: young children's use of comparison in category learning. *Journal of experimental psychology: General*, *131*(1), 5. http://dx.doi.org/10.1037/0096-3445.131.1.5

Neander, K. (1987). Pictorial representation: A matter of resemblance. *The British Journal of Aesthetics, 27*, 213–226. doi:http://dx.doi.org/10.1093/bjaesthetics/27.3.213. http://dx.doi.org/10.1093/bjaesthetics/27.3.213

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological science*, *21*(12), 1894-1902. http://dx.doi.org/10.1177/0956797610389189

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*, 185–212

Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, *73*(1), 1–33. http://dx.doi.org/10.1016/S0010-0277(99)00034-7

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science, 13*(1), 13–19. http://dx.doi.org/10.1111/1467-9280.00403

Taxitari, L., Twomey, K. E., Westermann, G., & Mani, N. (2020). The limits of infants’ early word learning. *Language Learning and Development*, *16*(1), 1-21. http://dx.doi.org/10.1080/15475441.2019.1670184

Twomey, K. E., Ranson, S. L., & Horst, J. S. (2014). That's more like it: Multiple exemplars facilitate word learning. *Infant and Child Development*, *23*(2), 105-122. http://dx.doi.org/10.1002/icd.1824

Vlach, H. A., & Sandhofer, C. M. (2011). Developmental differences in children’s context-dependent word learning. *Journal of Experimental Child Psychology*, *108*(2), 394-401. https://doi.org/10.1016/j.jecp.2010.09.011

Vukatana, E., Graham, S. A., Curtin, S., & Zepeda, M. S. (2015). One is not enough: Multiple exemplars facilitate infants' generalizations of novel properties. *Infancy*, *20*(5), 548-575. http://dx.doi.org/10.1111/infa.12092

Westfall, J. (2015). *PANGEA: Power ANalysis for GEneral Anova designs*. http://jakewestfall.org/publications/pangea.pdf