British Journal of Developmental Psychology (2018)

The British Psychological Society

www.wileyonlinelibrary.com

How self-generated labelling shapes transfer of learning during early childhood: The role of individual differences

Alecia Moser[1,a], Sarah Olsen[1,a], Sylvia N. Rusnak[2], Rachel Barr[2] and Peter Gerhardstein[1,*]

[1]Department of Psychology, Binghamton University, New York, USA
[2]Department of Psychology, Georgetown University, Washington, District of Columbia, USA

Multiple factors influence imitation during toddlerhood, including task complexity, social contingency, and individual differences. We conducted a secondary data analysis of individual differences in self-generated labelling using data collected from a complex puzzle imitation task with 355 2- to 3-year-olds. This analysis indicated that toddlers' ability to label the completed puzzle (fish or boat) was associated with better imitation performance. Labelling occurs during social interactions; therefore, our second analysis tested how labelling differed as a function of the level of social scaffolding in each condition. This analysis revealed that self-generated labelling was lower when the social demonstrator was removed and the task was presented on a touchscreen. This study is one of the first to examine self-generated labelling during a complex imitation task in toddlers and increases our understanding of the complexity of memory processing needed for imitation learning.

Statement of contribution

What is already known on this subject?

- Toddlers exhibit a transfer of learning deficit from 2D media, including books, TV, and tablets.
- Self-generated labelling enhances children's learning, through attentional and cognitive mechanisms.
- Children are sensitive to reduced social cues in screen media contributing to the transfer deficit.

What does this study add?

- Self-generated labelling is associated with better goal imitation performance.
- Self-generated labelling occurs more frequently under social conditions.

Development of representations across toddlerhood and early childhood has been investigated by examining individual differences in self-generated labelling performance (e.g., Miller & Marcovitch, 2011). Self-generated labelling is a linguistic form of identification that may demonstrate a child's access to a conceptual representation of an object (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Miller & Marcovitch, 2011). This capacity for self-generated labelling suggests an ability to perceive the named object

*Correspondence should be addressed to Peter Gerhardstein, Department of Psychology, Old Johnson Hall, Binghamton University, Binghamton NY 13902-6000 (email: gerhard@binghamton.edu).
[a]These authors contributed equally.

DOI:10.1111/bjdp.12254

abstractly and symbolically. Specifically, the label may offer a scaffold to assist transfer of learning when contexts change; for example, a label might facilitate a child's ability to link an on-screen image with the real object that it represents.

Object representations and language are tightly linked, particularly by word learning (Osina, Saylor, & Ganea, 2013; Smith, 2013; Xu, 2002), such that representations influence language and language influences object representations (Clark, 2004; Fulkerson & Waxman, 2007; Mareschal & Quinn, 2001; Yee, Jones, & Smith, 2012). Learning object names effectively draws attention to similarities (based on shape, for instance; Goldstein *et al.*, 2010; Golinkoff *et al.*, 1994; Graham, Kilbreath, & Welder, 2004; Henderson & Graham, 2005; Samuelson & Smith, 2005). This learning, in turn, is argued to promote more abstract conceptual representations that are generalizable (Barr & Brito, 2014; Hayne, 2004, 2006). While labelling may only *reflect* this increasing ability to generalize and not bring it about, the capacity to label may provide a marker of a more generalized representation.

Linguistic cues, which establish a connection to the object representation (Clark, 2004), can enhance recognition and learning under challenging perceptual conditions and high cognitive load by offering a scaffold to link an object in different contexts (Gerson & Woodward, 2014; Hayne & Herbert, 2004; Herbert & Hayne, 2000; Miller & Marcovitch, 2011; Simcock & Hayne, 2002; Taylor, Liu, & Herbert, 2016; Troseth, 2010). Miller and Marcovitch (2011), for instance, examined how language cues describing the location of the hidden object influence 2-year-olds' object search. Results revealed that verbal labels facilitated performance; specifically, children performed best when they self-generated the label, and this was more beneficial than when the experimenter provided the label. These findings indicate that possessing a label helped children make the connection between learning and testing contexts in a search task.

Transfer of learning involves applying information originally learned in one context to a context that is distinctly different (Barnett & Ceci, 2002). Across a wide range of tasks, toddlers and young children show decrements in learning when there is a change in the physical context between the learning and testing environments (Barnett & Ceci, 2002; Hartshorn *et al.*, 1998; Herbert & Hayne, 2000; Strouse & Ganea, 2017; Strouse & Troseth, 2014). Specifically, toddlers demonstrate this *transfer deficit* (for reviews see Barr, 2010, 2013; Hipp *et al.*, 2017) when transferring from video (Dickerson, Gerhardstein, Zack, & Barr, 2013; Moser *et al.*, 2015; Nielsen, Simcock, & Jenkins, 2008) and touchscreens (Moser *et al.*, 2015; Zack, Gerhardstein, Meltzoff, & Barr, 2013; Zimmermann, Moser, Lee, Gerhardstein, & Barr, 2016) to real-world settings. This transfer deficit is bidirectional (Moser *et al.*, 2015; Zack *et al.*, 2013).

Explanations of the transfer deficit reference developmental changes in representational development, particularly children's limited representational flexibility (Barr, 2013) and symbolic understanding (Troseth, 2010), as major contributing factors. Representational flexibility allows information to be generalized, abstracted, and used across multiple contexts (Barnett & Ceci, 2002; Barr, 2010; Hayne, 2004, 2006). Children's lack of symbolic understanding, which requires symbolic objects to be mapped onto their real-world counterparts, can interfere with transfer of knowledge from the symbol to the object and vice versa (e.g., DeLoache, Simcock, & Marzolf, 2004; Troseth, 2010; Troseth & DeLoache, 1998). The ability to self-generate object labels (a toddler's capacity to name objects in their environment) is thought to reflect greater representational flexibility (Miller & Marcovitch, 2011) and may be an indication that children have successfully drawn a relation between a symbolic 2D object and its 3D referent (Golinkoff *et al.*, 1994); this second accomplishment is referred to as representational insight

(Suddendorf, 2003; Troseth, Bloom Pickard, & DeLoache, 2007). For instance, Strouse and Troseth (2014) found that toddlers transferred a novel label ('modi') of an object from 2D video learning context to a 3D test only when parents verbally told the children that the on-screen object was 'the same as' the one they were familiarized with. Considering the ubiquitous nature of the transfer deficit involving screen media, there has been surprisingly little investigation of how individual differences might impact such transfer during early childhood, such as individual differences in labelling. The present analysis examined whether self-generation of a label is associated with overall imitation performance and specifically with enhanced transfer of learning from media.

Aims of the current data analysis

Given the importance of self-generated labelling as a marker of a more flexible representation of the object, as well as evidence of the ability to generate a symbol to stand for an object, the present analysis first examined whether self-generated labelling was associated with imitation performance overall. Imitation tasks involve a demonstration, which is followed by a test of the child's ability to imitate the demonstrated actions. This type of test is frequently referred to as a social learning task, because there is a dynamic social interaction between the demonstrator and the imitating child (Hipp *et al.*, 2017). Imitation tasks therefore permit direct manipulation of the social context. Like imitation, labelling occurs within a social context. Parental object labelling is frequent and predictable during early development such that by 2 years of age, children reliably use social cues (eye-gaze) conveyed by a social agent to learn object-label pairings (Graham, Nilsen, Collins, & Olineck, 2010). These types of social learning environments also support toddlers' ability to use a provided label to transfer learning from pictures to real-world objects (Geraghty, Waxman, & Gelman, 2014). Therefore, imitation tasks, which are typically social in nature, offer an ideal opportunity to study the conditions that engender labelling and whether it is associated with imitation performance, particularly under conditions of transfer.

The puzzle imitation task was developed to examine factors affecting transfer of learning from screen media. This task involved construction of a 3-piece magnetic or virtual puzzle that formed either a 'fish' or a 'boat' (Barr *et al.*, 2016; Dickerson *et al.*, 2013; Moser *et al.*, 2015; Zimmermann *et al.*, 2015, 2016; for review see Hipp *et al.*, 2017). Across different studies, this puzzle task has been demonstrated for 2- to 3-year-olds via video, on a touchscreen with 2D virtual puzzle pieces, and on a magnet board using 3D magnet pieces. At test, children are given the opportunity to reproduce the goal of the task by constructing the fish or the boat as demonstrated; imitation performance is measured by this goal score. Transferring between dimensions is referred to as the *transfer* condition (i.e., 3D-2D, 2D-3D) and imitation performance following these demonstrations is compared to conditions in which the demonstration and test dimension are the same, referred to as the *no-transfer* condition (i.e., 3D-3D, 2D-2D). Despite the high degree of similarity between the 3D and 2D puzzle, children in transfer conditions imitate significantly less than children in no-transfer conditions (Moser *et al.*, 2015).

In addition to manipulating transfer between 2D and 3D contexts, this design allowed us to manipulate social cues. In one study, we reduced social scaffolds using a 'ghost' demonstration condition, where the 2D puzzle pieces moved by themselves across the screen, thus eliminating the human demonstration of the target actions. When social cues were removed and replaced with a ghost demonstration, children's imitation performance on the 2D puzzle test returned to baseline (Zimmermann *et al.*, 2016); in other words, they failed to construct the puzzle. Across studies, social scaffolds ranged from low social scaffolding (ghost demonstration) to high social scaffolding (live demonstration; see Figure 2).

The puzzle task provides an ideal opportunity to test the role of individual differences in labelling. Playing with puzzles is familiar to most children (Levine, Ratliff, Huttenlocher, & Cannon, 2012), and the puzzle task is challenging because the order of assembly is arbitrary (i.e., any order of assembly is possible). Furthermore, across multiple studies conducted using the puzzle task, the researchers observed that some children were able to name the completed puzzle (a stylized 'fish' or 'boat'), while others were not. As stated previously, labelling the puzzle is a linguistic form of identification that may signal the availability of a conceptual representation of an object. Therefore, we hypothesized that self-generated labelling would be associated with better imitation performance on the puzzle task, as indicated by higher goal scores. We also predicted that children who self-generate the object label would be more likely to successfully transfer across dimensions. That is, transfer conditions are expected to show the typical impairment in imitation performance, but self-generating the object label might correspond to a greater increase in imitation performance in the transfer conditions than the no-transfer conditions. We further examined the conditions under which labelling was likely to occur. In prior studies, we manipulated the amount of social scaffolding present during the demonstration. In this follow-up analysis, we hypothesized that more social scaffolding would be associated with more self-generated labelling.

Method

Sample

The current secondary data analyses included data drawn from five previously published studies using the puzzle imitation task (see Table 1; Barr *et al.*, 2016; Dickerson *et al.*, 2013; Moser *et al.*, 2015; Zimmermann *et al.*, 2015, 2016). A total of 355 (182 girls) typically developing 2-, 2.5-, and 3-year-olds (from two metropolitan areas) were included in the sample analysed in the secondary data analyses. All of the studies were procedurally identical except when explicitly noted. All studies were approved by the appropriate institutional review board, conformed to the ethical guidelines of the American Psychological Association, and included parental consent.

The present analyses were conducted using independent groups of children who were 2 years (n = 72; M age = 747 days, SD = 11 days), 2.5 years (n = 156; M age = 930 days, SD = 11 days), and 3 years (n = 127; M age = 1,113 days, SD = 13 days) of age. The sample was primarily Caucasian (79.72%) and from college-educated families (M years of education = 16.8, SD = 1.4, as reported by 96% of families). The remaining 20.28% of the sample included the following: mixed (14.37%), African American (1.13%), Asian (3.38%), and not reported (1.41%). Additionally, 5.35% of the sample was Latino. Socio-economic index rank (SEI; Nakao & Treas, 1992, 1994) was based on 169 families (47.7%; M = 80.7, SD = 13.1).

Apparatus and stimuli

The puzzle imitation task and apparatus were the same across the studies from which data were drawn. A rectangular enclosure contained a 17″ touchscreen for 2D demonstrations and tests, which could be easily converted to allow for 3D demonstrations and tests by

Table 1. Experimental conditions as a function of demonstration and test phase

Condition	Demonstration	Test	Transfer analysis 1	Social analysis 2	*n* / Proportion who labeled: 2-years	2.5-years	3-years	Total
3D-3D$_{1,2,3}$	Magnet board	Magnet board	No-transfer	NA	46 0.22	27 0.56	20 0.7	93 0.42
2D-2D$_{3,5}$	Touch screen	Touch screen	No-transfer	High social	NA NA	19 0.21	17 0.59	36 0.39
2D-3D$_{3,5}$	Touch screen	Magnet board	Transfer	High social	NA NA	26 0.56	23 0.78	49 0.67
3D-2D$_{3}$	Magnet board	Touch screen	Transfer	NA	NA NA	19 0.4	18 0.56	37 0.47
Video-3D$_{2,3,4}$	Video	Magnet board	Transfer	NA	26 0.27	26 0.5	12 0.75	64 0.45
Ghost-3D$_{5}$	Touch screen	Magnet board	NA	Low social	NA NA	13 0.23	13 0.54	26 0.38
Ghost-2D$_{5}$	Touch screen	Touch screen	NA	Low social	NA NA	26 0.23	24 0.46	50 0.34

Note. No-transfer conditions involved no change in dimension between demonstration and test; transfer conditions involved a change in dimension (2D to 3D or reverse) between demonstration and test. High social scaffold level involved the experimenter showing the child how to move the puzzle pieces on the screen or magnet board, and low social scaffold level ('ghost' demonstration) involved the virtual pieces on the touch screen moving by themselves. Condition subscripts denote the article(s) in which data for that condition was drawn from: (1) Barr *et al.* (2016), includes only 2-year-olds; (2) Dickerson *et al.* (2013), includes only 2-year-olds; (3) Moser *et al.* (2015); (4) Zimmermann *et al.* (2015); and (5) Zimmermann *et al.* (2016).

sliding a metal magnetic board into the enclosure. The 3D version of the task consisted of three magnetic pieces (5 mm thick) that, when connected correctly, formed a 'fish' or a 'boat'. The puzzle for 13 2-year-olds contained an additional distractor piece that did not influence performance. See Figure 1 for illustration of the puzzle configurations. The corresponding 2D touchscreen stimuli consisted of three virtual puzzle pieces generated from high-resolution photographs of the pieces and rendered to match them in colour and size. Critically, the 2D and 3D pieces could be manipulated (i.e., slid into position by pressing on the surface with one or two fingers) in similar ways despite the differences in dimensionality, but the 3D pieces could also be moved by grasping the edges and picked up and held as 3D objects. Participants received the boat (n = 174) or fish (n = 181) versions of the puzzle.

Procedure

The procedure was identical across all five studies except as noted (see Figure 2). All conditions included three consecutive phases: demonstration, test, and manipulation check. During the *live demonstration*, an adult demonstrator slid each puzzle piece into position to make either the fish or the boat on the 3D magnet board or the 2D touchscreen accompanied by scripted verbal cues (i.e., 'Look at this!' 'What was that?' 'Isn't that fun?'). This demonstration was performed three times and lasted 60 s. The *video demonstration*

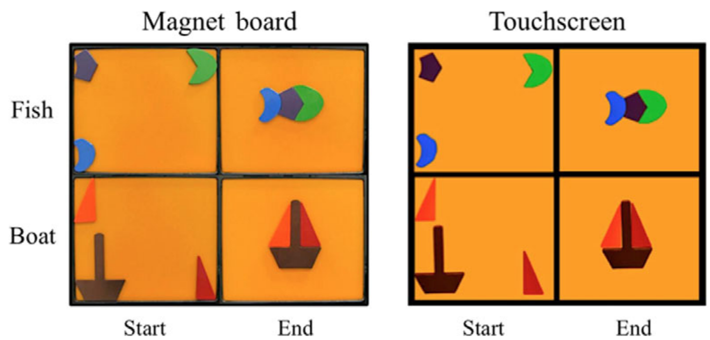

Figure 1. Puzzle configurations for each of the stimuli, fish and boat, shown in their start and end (assembled) positions. The magnetic (left) and touchscreen (right) apparatus and stimuli were well-matched on perceptual cues and functionality. [Colour figure can be viewed at wileyonlinelibrary.com]

Figure 2. Images depict the various demonstration types across all conditions included in the present secondary data analysis. Conditions were equated as much as possible on experimental set-up and perceptual properties. [Colour figure can be viewed at wileyonlinelibrary.com]

consisted of a pre-recorded video of the standard live demonstration on the 3D magnet puzzle board. The *ghost demonstration* was made by screen-capturing the movement of the 2D puzzle pieces (no visible hand) and the corresponding verbal cues (provided via voice-over). During the video and ghost demonstrations, an experimenter sat and watched the demonstration with the child.

A 60 s *test* followed the demonstration and used either the 3D magnet board or the 2D touchscreen. If the child successfully completed the puzzle during the test phase, the experimenter asked the child, 'What did you make?' Finally, a *manipulation check* was conducted to ensure that the child could perform the sliding gestures using the test pieces. During this check, if the child had not labelled the 'boat' or 'fish' during the demonstration or test, the experimenter asked the child, 'What did you make?' Sessions were video-recorded for later coding.

Design variables

In all of the conditions, *transfer condition* was determined by the match between demonstration and test. No-transfer conditions involved no change in dimension between demonstration and test (e.g., live 3D demonstration and 3D test, or live 2D demonstration and 2D test, denoted as 3D-3D and 2D-2D, respectively), while transfer conditions involved a change in dimension between demonstration and test (e.g., 3D, 2D, or video demonstration, and test in the opposite dimension, denoted as 3D-2D, 2D-3D, and Video-3D, respectively). Demonstrations were also categorized by *level of social scaffolding*. Demonstrations involving high social scaffolding (live demonstration; hereafter high social scaffold) entailed the experimenter demonstrating moving the puzzle pieces on the touchscreen or magnet board (including verbal cues). Demonstrations involving low levels of social scaffolding (ghost demonstration; hereafter low social scaffold; denoted as Ghost-2D and Ghost-3D) involved virtual pieces on the touchscreen moving by themselves, accompanied by voice-over verbal cues.

Coding

Goal score

We measured whether children correctly imitated, connecting the puzzle pieces to reach the goal of making a fish or a boat puzzle. The proportion of correctly connected puzzle pieces is referred to as the goal score. Coding was completed using Datavyu (Datavyu.org), a free open-source video coding system. All behaviours were time-stamped. Scoring was consistent across all studies used for the current analysis. A correct connection was scored when a puzzle piece was connected to the central piece in the correct location and orientation. Children received a point for each correct connection of puzzle pieces, for a maximum score of two correct connections, which were then converted to a proportion. Based on 29.6% of all test sessions, inter-rater reliability was above the acceptable level of .70 (Landis & Koch, 1977) in each study; kappa for goal score ranged from .81 to .92.

Assessing self-generated labelling

Some children spontaneously labelled the object during the demonstration or test. For all others, as part of the standard procedure, experimenters asked the child ‘what did you make?’ either after the puzzle was completed during the test phase or during the manipulation check. A small fraction (3.3%) of participants who received the fish puzzle called it a ‘spaceship’ or ‘rocketship’; these children were counted as having correctly identified the object because this label, specific for fish, was produced at far greater frequency compared to other incorrect labels; the fish does, in fact, resemble a spaceship/rocketship. A score of 1 was given if the child labelled the object during demonstration, test, or in response to the question ‘what did you make?’ [see Miller and Marcovitch (2011) for similar method]. Otherwise, the child was given a score of 0, for no response or an incorrect response (e.g., ‘train’ for ‘fish’). For various subject- and experimenter-related reasons, 12 subjects did not receive the question ‘what did you make?’ and were not included in the sample of 355 children. Inter-rater reliability on this measure, generated from 30% of the videos rescored by a second coder, was high at 0.94 (Landis & Koch, 1977).

Parents were also asked to report whether the child’s productive vocabulary contained the task-specific words of boat, fish, and puzzle. The majority of children had the puzzle labels in their productive vocabulary (fish: 93% of 294 participants; boat: 92% of 313 participants), indicating that failure to label the puzzle was probably not due to an inability to produce the label. Finally, there was no difference in latency to touch the puzzle between participants who did ($M = 10.31$ s) and did not ($M = 11.58$ s) self-generate the object label ($t < 1$ after the removal of outliers over 3 *SD*); children who did not label were as likely to interact with the puzzle as those who labelled.

Results

Data analysis plan

The analyses investigated two questions: Analysis 1 investigates the relationship between self-generated labelling and imitation performance on the puzzle task in general. Analysis 2 assesses whether self-generated labelling was associated with the level of social scaffold (high or low). Separate analyses were required due to the presence of unequal sample sizes across ages and conditions or lack of conditions at specific ages (e.g., ghost condition was not run with 2-year-olds).

All statistical analyses were conducted using the R system for statistical computing (R Core Team, 2016; RStudio Team, 2015), along with the appropriate statistical R analysis packages including plyr (Wickham, 2011), Hmisc (Harrell & Dupont, 2016), and ggplot2 (Wickham, 2009).

Analysis 1: self-generated labelling and imitation performance

Samples analysed

Children were drawn from no-transfer (live 2D-2D or 3D-3D demonstration test) or transfer (Video-3D, live 2D-3D, or 3D-2D) conditions, resulting in 129 participants in the no-transfer condition and 150 participants in the transfer condition. Within the no-transfer condition, 53 participants self-generated the object label (10 2-year-olds, 19 2.5-year-olds, and 24 3-year-olds), and 76 participants did not (36 2-year-olds, 27 2.5-year-olds, and 13 3-year-olds). Of those in the transfer condition, 79 participants self-generated the label (seven 2-year-olds, 35 2.5-year-olds, and 37 3-year-olds), and 71 did not (19 2-year-olds, 36 2.5-year-olds, and 16 3-year-olds). See Table 1 for further sample details. Participants who received the ghost demonstration were excluded from this analysis, as 2-year-olds were not run in that condition.

Does self-generated labelling differ across age and transfer condition?

A preliminary analysis examined the pattern of individual differences in self-generated labelling across age and transfer conditions. Labelling was expected to increase with age due to language development across this age range. While no prediction was made regarding transfer, it is possible that transfer may have interfered with labelling by presenting the object in different dimensions between demonstration and test (see Strouse & Ganea, 2017 for discussion). Alternatively, when no transfer was involved, the child was provided with multiple exposures to the object in the same dimension, potentially boosting opportunities for labelling.

A binomial logistic regression was conducted, which included labelling as the dependent measure (1 = yes, 0 = no), and age (2, 2.5, and 3 years, centred and scaled, as a continuous variable), transfer (no-transfer or transfer, as a nominal variable), and the age by transfer interaction as fixed factors. Results revealed a significant effect of age ($\beta = .69$, $SE = 0.19$, $z = 3.55$, $p < .001$), no effect of transfer ($\beta = -.29$, $SE = 0.26$, $z = -1.11$, $p = .27$), and no interaction between age and transfer ($\beta = .03$, $SE = 0.27$, $z = 0.11$, $p = .91$). This analysis indicated, as expected, an age-related increase in self-generation of the object label, such that at 2 years, only 24% of the children self-generated the label, while at 2.5 years almost half (46%) self-generated the object label, and 68% of 3-year-olds self-generated the object label (see Figure 3). Additionally, this analysis indicated no difference in self-generated labelling as a function of transfer condition.

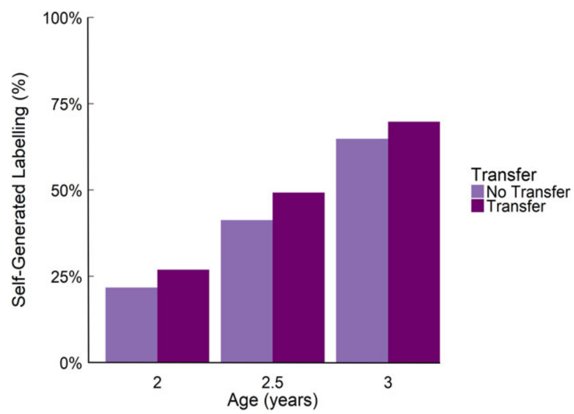

Figure 3. Proportion of subjects who successfully self-generated the object label across age (2-, 2.5-, and 3-year-olds) and transfer (no-transfer and transfer). There are no error bars because labelling is a binary measure. [Colour figure can be viewed at wileyonlinelibrary.com]

Is self-generated labelling associated with imitation performance and transfer of learning?
Imitation performance across age, self-generated labelling, and transfer (see Figure 4) was assessed using linear regression with goal score as the dependent variable (proportions of 0, 0.5, and 1, as a continuous variable) and age (2, 2.5, and 3 years, centred and scaled, as a continuous variable), transfer (transfer or no-transfer, as a nominal variable), self-generated labelling (yes or no, as a nominal variable), and all the interactions entered as fixed effects.

Results revealed an effect of age ($\beta = .15$, $SE = 0.52$, $t = 2.92$, $p = .004$), indicating that as age increased, goal score also increased (2-year-olds: $M = 0.33$; 2.5-year-olds: $M = 0.50$; 3-year-olds: $M = 0.79$). There was also an effect of transfer condition ($\beta = .28$, $SE = 0.07$, $t = 3.99$, $p < .001$), indicating that goal score was significantly higher for the no-transfer ($M = 0.62$) versus the transfer condition ($M = 0.49$). Additionally, the main effect of self-generated labelling was significant ($\beta = .26$, $SE = 0.07$, $t = 3.76$, $p < .001$), such that goal score was higher for those who self-generated the object label ($M = 0.69$), compared to those who did not ($M = 0.42$). The interaction between self-generated labelling and transfer was marginally significant ($\beta = -.17$, $SE = 0.10$, $t = -1.67$, $p = .097$); all other interactions were not significant (t's < 1). The overall model was significant, $F(7, 271) = 11.99$, $p < .0001$, with an adjusted R^2 of 0.22.

These findings indicate that self-generated labelling was associated with better performance on the imitation task. Additionally, the lack of an age by transfer interaction indicates that self-generated labelling was associated with higher goal scores across all age groups (highest variability in performance is in the 2-year-olds; see Figure 4). Despite a trend in the expected direction, self-generated labelling was not associated with differences in imitation performance as a function of transfer.

Analysis 2: self-generated labelling and the social context

Under which experimental conditions is labelling most likely to occur? Multiple social factors have been reported to affect learning from media during early childhood, including social contingency (Myers, LeWitt, Gallo, & Maselli, 2016; Nielsen *et al.*, 2008; Roseberry,

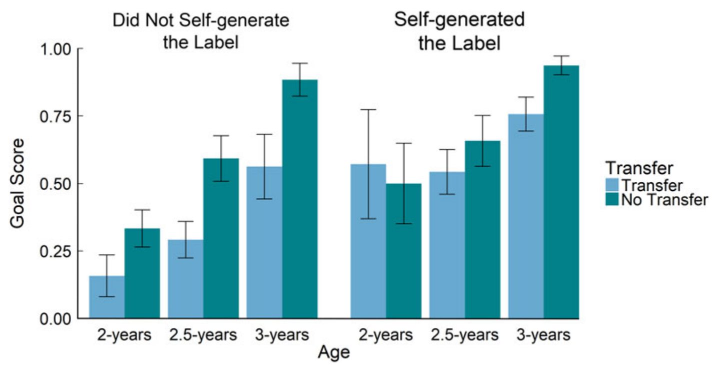

Figure 4. Goal score as a function of age and transfer for those not self-generating the goal object label ('boat', 'fish') on the left and those self-generating the label on the right. Sample sizes are unequal across these groups, as naming is an individual difference. [Colour figure can be viewed at wileyonlinelibrary.com]

Hirsh-Pasek, & Golinkoff, 2014; Troseth, Saylor, & Archer, 2006; Zimmermann *et al.*, 2016), social relevance of the on-screen actor (Krcmar, 2010), and parental scaffolding (Strouse & Troseth, 2014; Zack & Barr, 2016). Why might the social context influence the likelihood of labelling? Labelling is typically embedded in rich, socially reinforcing parent–child interactions (e.g., Goldstein *et al.*, 2010; Meltzoff, 2007; Zammit & Schafer, 2011). Parents use a range of social cues such as eye-gaze, gestures, infant-directed speech, and labelling to draw children's attention to objects in the environment (Callanan, Akhtar, & Sussman, 2014; Graham *et al.*, 2010; Wu, Gopnik, Richardson, & Kirkham, 2011). These social cues act to emphasize important objects and events in the noisy environment (Wu *et al.*, 2011) and facilitate the exchange of information about these objects. Such interactions cement both knowledge about objects in the world and how to name these objects (e.g., Zammit & Schafer, 2011). Considering the presence and frequency of labels in social learning contexts, in parent–child interactions, and in transfer learning experiments, we predicted that self-generation of the object label would increase when a high level of social scaffolding was provided, as compared to when a low level of social scaffolding was provided.

Sample analysed

The analysis examining the relationship between self-generated labelling and level of social scaffolding excluded 2-year-old participants, as they were not tested in the ghost condition. Children who received live 3D demonstrations were also excluded to keep demonstration dimension consistent across social scaffold level. Additionally, children who received the video demonstration were excluded to create a balanced high versus low scaffold design, as video could not be easily delineated as a high or low scaffold condition. Therefore, the high social scaffold condition included participants receiving 2D live demonstrations (2D-3D and 2D-2D), and the low social scaffold condition included participants receiving ghost demonstrations (Ghost-3D and Ghost-2D; see Table 1 and Figure 2), resulting in 85 high social scaffold participants (45 2.5-year-olds and 40 3-year-olds) and 76 low social scaffold participants (39 2.5-year-olds and 37 3-year-olds).

To examine differences in the proportion of self-generated labelling as a function of age (in years, centred and scaled) and social scaffold level (see Figure 5), a binomial logistic regression was conducted, with labelling as the dependent measure, and age (2.5 and 3 years, centred and scaled, as a continuous variable), social scaffold level (high and low, as a nominal variable), and the interaction between age and social scaffold level as the fixed factors. The effect of age was reliable ($\beta = .58$, $SE = 0.23$, $z = 2.53$, $p = .01$), such that 3-year-olds ($M = 0.60$) were more likely to self-generate the label than 2.5-year-olds ($M = 0.33$). Social scaffold level was also reliable ($\beta = -.90$, $SE = 0.34$, $z = -2.63$, $p = .009$), indicating that high-scaffold participants ($M = 0.55$) were more likely to self-generate the label than low scaffold participants ($M = 0.36$). The interaction between age and social scaffold level was not significant ($z < 1$). These findings indicate that, independent of age, labelling was substantially reduced when children did not have a social demonstrator.

General discussion

Imitation performance was higher among children who self-generated the object label. We speculate that object labels generated prior to, during, or following successful completion of the goal indicate that the object concept was retrieved from memory and available during the demonstration and test. Activating this object concept may have several functions that could correspond to better imitation performance, such as promoting goal-directed behaviour by directing attention to the goal state of the task, as opposed to object movement or the demonstrator's gestures. Labelling may also guide attention to the similarities between newly acquired object representations in the child's

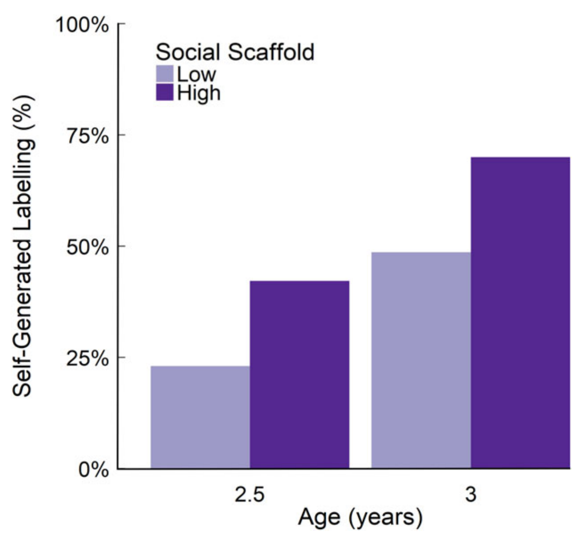

Figure 5. Percentage of children who self-generated the label as a function of age and level of social scaffolding (high or low). There are no error bars because labelling is a binary measure. [Colour figure can be viewed at wileyonlinelibrary.com]

present environment with previously stored representations of objects of the same category (see also Xu, 2002).

Developmentally, labels may help establish abstract and flexible representations of objects (Geraghty *et al.*, 2014; Gerson & Woodward, 2014; Golinkoff *et al.*, 1994; Kiefer & Pulvermüller, 2012; Miller & Marcovitch, 2011; Quinn & Eimas, 2000; Xu, 2002; Zack *et al.*, 2013). Other possibilities and explanations exist as well. Imitation itself may have increased the rate of self-generated labelling; if so, self-generated *actions* may promote symbolic development. That is, with age and experience, memory systems become increasingly flexible, which allows information to be used in multiple contexts (Barr & Brito, 2014), promoting the mapping of self-generated labels. Given the correlational design, the present analysis cannot differentiate between these possibilities. Therefore, it will be important for future research to experimentally determine whether and to what extent manipulated differences between labelling conditions influence transfer of learning.

Why might self-generated labelling be more likely to occur during a highly scaffolded social interaction, as indicated by our second analysis? Children's productive vocabulary is related to frequent and repetitive labelling of objects by parents, both verbally and using gestures (Hoff & Naigles, 2002; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Tamis-LeMonda, Bornstein, & Baumwell, 2001). Imitating a parent who offers these reliable social cues provides a primary avenue through which children learn to direct attention to relevant, information-rich aspects of an object or parts of objects within a complex scene (Meltzoff, 2007). The lack of such social cues in a 2D ghost demonstration context may therefore account for the reduction in labelling in the low social scaffold condition, as well as the reduced ability to learn from 2D sources (see also Nielsen *et al.*, 2008). This finding supports the notion that screen media, like books and other learning tools, work best in a supportive social learning environment (also see Strouse & Troseth, 2014). While an open empirical question, the constellation of social and language cues that typically occur during parent–child interactions during object play (e.g., Meltzoff, 2007) may allow children who imitate high-labelling parents to better self-generate labels themselves. In this way, social interactions, and in particular parent–child interactions involving imitation, may scaffold self-generated labelling and facilitate learning.

While providing important insights, the current study has a number of limitations. Most notably, the study is correlational: Self-generation of the label was not manipulated and sample sizes varied greatly across condition. Further, the goal score, our primary outcome measure, has a small range. The pattern of results seen here warrants further empirical investigation; for example, it would be of particular interest to conduct an experiment in which self-generation of the object label (or at least the probability of self-generation) is manipulated by teaching a novel-object novel-label pairing ahead of time in the home and then testing for self-generation of the label under conditions of transfer and no-transfer. Another factor to consider is when children are asked to generate the label in relation to when they imitate: Asking children to generate before as opposed to following imitation performance would determine the degree to which labelling increases transfer success, as opposed to imitation performance impacting labelling. It would also be useful to test whether labelling is more beneficial as a function of age or task difficulty such that labelling confers a larger advantage under more challenging test conditions or alternatively at a developmental point close to when the child is able to transfer without scaffolding.

Transfer of learning from media provides memory researchers a unique window into the development of memory integration by enabling controlled manipulation of social, perceptual, and linguistic cues. The present study examined these variables using individual differences in self-generated labelling. First, the present study demonstrated that object representations and language are tightly linked. This suggests that object labelling is a marker of the development of a more flexible memory system. How it functions in this capacity needs to be further investigated with regard to learning involving screen media. Additionally, information is not processed in a vacuum, but rather in a social context, which, as shown here, is related to labelling performance. Tests of transfer across development should take into account the participants' level of flexibility as indexed by labelling capability, as well as, the level of social scaffolding within the test.

Acknowledgements

This research was funded by NSF grant no. 1023772 to Rachel Barr and Peter Gerhardstein.

References

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Barr, R. (2010). Transfer of learning between 2D and 3D sources during infancy: Informing theory and practice. *Developmental Review*, *30*, 128–154. https://doi.org/10.1016/j.dr.2010.03.001

Barr, R. (2013). Memory constraints on infant learning from picture books, television, and touchscreens. *Child Development Perspectives*, *7*, 205–210. https://doi.org/10.1111/cdep.12041

Barr, R., & Brito, N. (2014). From specificity to flexibility: Developmental changes during infancy. In P. Bauer & R. Fivush (Eds.), *Wiley-Blackwell handbook on the development of children's memory* (pp. 453–479). Chichester, UK: Wiley.

Barr, R., Moser, A., Rusnak, S., Zimmermann, L., Dickerson, K., Lee, H., & Gerhardstein, P. (2016). The impact of memory load and perceptual cues on puzzle learning by 2-year-olds. *Developmental Psychobiology*, *58*, 817–828. https://doi.org/10.1002/dev.21450

Callanan, M. A., Akhtar, N., & Sussman, L. (2014). Learning words from labeling and directive speech. *First Language*, *34*, 450–461. https://doi.org/10.1177/0142723714553517

Clark, E. V. (2004). How language acquisition builds on cognitive development. *Trends in Cognitive Sciences*, *8*, 472–478. https://doi.org/10.1016/j.tics.2004.08.012

DeLoache, J. S., Simcock, G., & Marzolf, D. P. (2004). Transfer by very young children in the symbolic retrieval task. *Child Development*, *75*, 1708–1718. https://doi.org/10.1111/j.1467-8624.2004.00811.x

Dickerson, K., Gerhardstein, P., Zack, E., & Barr, R. (2013). Age-related changes in learning across early childhood: A new imitation task. *Developmental Psychobiology*, *55*, 719–732. https://doi.org/10.1002/dev.21068

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition*, *105*, 218–228. https://doi.org/10.1016/j.cognition.2006.09.005

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

Gerson, S., & Woodward, A. L. (2014). Labels facilitate infants' comparison of action goals. *Journal of Cognition and Development*, *15*, 197–212. https://doi.org/10.1080/15248372.2013.777842

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., & Edelman, S. (2010). General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences*, *14*, 249–258. https://doi.org/10.1016/j.tics.2010.02.004

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*, 125–155. https://doi.org/10.1017/S0305000900008692

Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development*, *75*, 409–427. https://doi.org/10.1111/j.1467-8624.2004.00683.x

Graham, S. A., Nilsen, E. S., Collins, S., & Olineck, K. (2010). The role of gaze direction and mutual exclusivity in guiding 24-month-olds' word mappings. *British Journal of Developmental Psychology*, *28*, 449–465. https://doi.org/10.1348/026151009X424565

Harrell, F. E., & Dupont, C. (2016). *Hmisc: Harrell miscellaneous*. R package version 3.17-4. Retrieved from https://CRAN.R-project.org/package=Hmisc

Hartshorn, K., Rovee-Collier, C., Gerhardstein, P., Bhatt, R. S., Klein, P. J., Aaron, F., Wondoloski, T. L., & Wurtzel, N. (1998). Developmental changes in the specificity of memory over the first year of life. *Developmental Psychobiology*, *33*, 61–78. https://doi.org/10.1002/(SICI)1098-2302(199807)33:1<61:AID-DEV6>3.0.CO;2-Q

Hayne, H. (2004). Infant memory development: Implications for childhood amnesia. *Developmental Review*, *24*, 33–73. https://doi.org/10.1016/j.dr.2003.09.007

Hayne, H. (2006). Age-related changes in infant memory retrieval: Implications for knowledge acquisition. In Y. Munakata & M. H. Johnson (Eds.), *Processes of brain and cognitive development: Attention and performance XXI* (pp. 209–231). Oxford, UK: Oxford University Press.

Hayne, H., & Herbert, J. (2004). Verbal cues facilitate memory retrieval during infancy. *Journal of Experimental Child Psychology*, *89*, 127–139. https://doi.org/10.1016/j.jecp.2004.06.002

Henderson, A. M. E., & Graham, S. A. (2005). Two-year-olds' appreciation of the shared nature of novel object labels. *Journal of Cognition and Development*, *6*, 381–402. https://doi.org/10.1207/s15327647jcd0603_4

Herbert, J., & Hayne, H. (2000). Memory retrieval by 18–30-month-olds: Age-related changes in representational flexibility. *Developmental Psychology*, *36*, 473–484. https://doi.org/10.1037//0012-1649 36.4.473

Hipp, D., Gerhardstein, P., Zimmermann, L., Moser, A., Taylor, G., & Barr, R. (2017). The dimensional divide: Learning from TV and touchscreens during early childhood. In R. Barr & D. N. Linebarger (Eds.). *Media exposure during infancy and early childhood: The effect of content and context on learning and development* (pp. 33–54). New York, NY: Springer. https://doi.org/10.1007/978-3-319-45102-2

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*, 418–433. https://doi.org/10.1111/1467-8624.00415

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*, 236–248. https://doi.org/10.1006/pmed.1998.0301

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, *48*, 805–825. https://doi.org/10.1016/j.cortex.2011.04.006

Krcmar, M. (2010). Can social meaningfulness and repeat exposure help infants and toddlers overcome the video deficit? *Media Psychology*, *13*, 31–53. https://doi.org/10.1080/15213260903562917

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. https://doi.org/10.2307/2529310

Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental Psychology*, *48*, 530. https://doi.org/10.1037/a0025913

Mareschal, D., & Quinn, P. C. (2001). Categorization in infancy. *Trends in Cognitive Sciences*, *5*, 443–450. https://doi.org/10.1016/S1364-6613(00)01752-6

Meltzoff, A. N. (2007). 'Like me': A foundation for social cognition. *Developmental Science*, *10*, 126–134. https://doi.org/10.1111/j.1467-7687.2007.00574.x

Miller, S. E., & Marcovitch, S. (2011). Toddlers benefit from labeling on an executive function search task. *Journal of Experimental Child Psychology*, *108*, 580–592. https://doi.org/10.1016/j.jecp.2010.10.008

Moser, A., Zimmermann, L., Dickerson, K., Grenell, A., Barr, R., & Gerhardstein, P. (2015). They can interact but can they learn? Toddlers' transfer learning from touchscreens and television. *Journal of Experimental Child Psychology*, *137*, 137–155. https://doi.org/10.1016/j.jecp.2015.04.002

Myers, L. J., LeWitt, R. B., Gallo, R. E., & Maselli, N. M. (2016). Baby FaceTime: Can toddlers learn from online video chat? *Developmental Science*, *20*, e12430. https://doi.org/10.1111/desc.12430

Nakao, K., & Treas, J. (1992). Computing the 1989 Socioeconomic Index. General Social Survey Methodological Reports, No. 70. Chicago, IL: National Opinion Center.

Nakao, K., & Treas, J. (1994). Updating occupational prestige and socioeconomic scores: How the new measures measure up. *Sociological Methodology*, *24*, 1–72. https://doi.org/10.2307/270978

Nielsen, M., Simcock, G., & Jenkins, L. (2008). The effect of social engagement on 24-month-olds' imitation from live and televised models. *Developmental Science*, *11*, 722–731. https://doi.org/10.1111/j.1467-7687.2008.00722.x

Osina, M. A., Saylor, M. M., & Ganea, P. A. (2013). When familiar is not better: 12-month-old infants respond to talk about absent objects. *Developmental Psychology*, *49*, 138. https://doi.org/10.1037/a0027903

Quinn, P. C., & Eimas, P. D. (2000). The emergence of category representations during infancy: Are separate perceptual and conceptual processes required? *Journal of Cognition and Development*, *1*, 55–61. https://doi.org/10.1207/S15327647JCD0101N_6

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Development*, *85*, 956–970. https://doi.org/10.1111/cdev.12166

RStudio Team. (2015). *RStudio: Integrated development for R*. Boston, MA: RStudio. Retrieved from http://www.rstudio.com/.

Samuelson, L. K., & Smith, L. B. (2005). They call it like they see it: Spontaneous naming and attention to shape. *Developmental Science*, *8*, 182–198. https://doi.org/10.1111/j.1467-7687.2005.00405.x

Simcock, G., & Hayne, H. (2002). Breaking the barrier? Children fail to translate their preverbal memories into language. *Psychological Science*, *13*, 225–231. https://doi.org/10.1111/1467-9280.00442

Smith, L. B. (2013). It's all connected: Pathways in visual object recognition and early noun learning. *American Psychologist*, *68*, 618. https://doi.org/10.1037/a0034185

Strouse, G. A., & Ganea, P. A. (2017). Toddlers' word learning and transfer from electronic and print books. *Journal of Experimental Child Psychology*, *156*, 129–142. https://doi.org/10.1016/j.jecp.2016.12.001

Strouse, G. A., & Troseth, G. L. (2014). Supporting toddlers' transfer of word learning from video. *Cognitive Development*, *30*, 47–64. https://doi.org/10.1016/j.cogdev.2014.01.002

Suddendorf, T. (2003). Early representational insight: Twenty-four-month-olds can use a photo to find an object in the world. *Child Development*, *74*, 896–904. https://doi.org/10.1111/1467-8624.00574

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, *72*, 748–767. https://doi.org/10.1111/1467-8624.00313

Taylor, G., Liu, H., & Herbert, J. S. (2016). The role of verbal labels on flexible memory retrieval at 12-months of age. *Infant Behavior and Development*, *45*, 11–17. https://doi.org/10.1016/j.infbeh.2016.08.002

Troseth, G. L. (2010). Is it life or is it Memorex? Video as a representation of reality. *Developmental Review*, *30*, 155–175. https://doi.org/10.1016/j.dr.2010.03.007

Troseth, G. L., Bloom Pickard, M. E., & DeLoache, J. S. (2007). Young children's use of scale models: Testing an alternative to representational insight. *Developmental Science*, *10*, 763–769. https://doi.org/10.1111/j.1467-7687.2007.00625.x

Troseth, G. L., & DeLoache, J. S. (1998). The medium can obscure the message: Young children's understanding of video. *Child Development*, *69*, 950–955. https://doi.org/10.1111/j.1467-8624.1998.tb06153.x

Troseth, G. L., Saylor, M. M., & Archer, A. H. (2006). Young children's use of video as a source of socially relevant information. *Child Development*, *77*, 786–799. https://doi.org/10.1111/j.1467-8624.2006.00903.x

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis* (p. 2009). New York, NY: Springer. https://doi.org/10.1007/978-0-387-98141-3

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*, 1–29. https://doi.org/10.18637/jss.v040.i01

Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology*, *47*, 1220. https://doi.org/10.1037/a0024023

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*, 223–250. https://doi.org/10.1016/S0010-0277(02)00109-9

Yee, M., Jones, S. S., & Smith, L. B. (2012). Changes in visual object recognition precede the shape bias in early noun learning. *Frontiers in Psychology*, *3*, 533. https://doi.org/10.3389/fpsyg.2012.00533

Zack, E., & Barr, R. (2016). The role of interactional quality in learning from touch screens during infancy: Context matters. *Frontiers in Psychology*, *7*, 1264. https://doi.org/10.3389/fpsyg.2016.0126

Zack, E., Gerhardstein, P., Meltzoff, A., & Barr, R. (2013). 15-month-olds' transfer of learning between touch screen and real-world displays: Language cues and cognitive loads. *Scandinavian Journal of Psychology*, *54*, 20–25. https://doi.org/10.1111/sjop.12001

Zammit, M., & Schafer, G. (2011). Maternal label and gesture use affects acquisition of specific object names. *Journal of Child Language*, *38*, 201–221. https://doi.org/10.1017/S0305000909990328

Zimmermann, L., Moser, A., Grenell, A., Dickerson, K., Yao, Q., Gerhardstein, P., & Barr, R. (2015). Do semantic contextual cues facilitate transfer learning from video in toddlers? *Frontiers in Psychology*, *6*, 561. https://doi.org/10.3389/fpsyg.2015.00561

Zimmermann, L., Moser, A., Lee, H., Gerhardstein, P., & Barr, R. (2016). The ghost in the touchscreen: Social scaffolds promote learning by toddlers. *Child Development*, *88*, 2013–2025. https://doi.org/10.1111/cdev.12683

Received 27 July 2017; revised version received 31 May 2018