View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of East Anglia digital repository

Title: Lexical acquisition through category matching: 12-month-old infants associate words to visual categories

Authors: Barbara Pomiechowska[1] & Teodora Gliga[2]

Affiliations:

[1] Cognitive Development Center, Department of Cognitive Science, Central European University, Hungary

[2] School of Psychology, University of East Anglia, Norwich, United Kingdom

Corresponding author: Barbara Pomiechowska, Cognitive Development Center, Department of Cognitive Science, Central European University, Nádor u. 9, Budapest 1051, Hungary, email: PomiechowskaB@ceu.edu

Word count: 140 words (abstract) + 2177 words (main text; does not include Methods and Results sections)

Abstract

Although it is widely recognized that human infants build a sizeable conceptual repertoire before mastering language, it remains a matter of debate whether and to what extent early conceptual and category knowledge contributes to language development. We addressed this question by investigating whether 12-month-olds used preverbal categories to discover the meanings of new words. We showed that infants (N = 18) readily extended novel labels to previously unseen exemplars of preverbal visual categories after only a single labeling episode, while struggling to do so when taught labels for unfamiliar categories (N = 18). These results suggest that infants expect labels to denote categories of objects and are equipped with learning mechanisms responsible for matching prelinguistic knowledge structures with linguistic inputs. This ability is in line with the idea that our conceptual machinery provides building blocks for vocabulary and language acquisition.

Keywords: word learning, word extension, categorization, cognitive development, infancy

Language not only pervades human communication and social interactions, but also provides us with powerful cognitive tools. In particular, language enables the development of new systems of knowledge (Spelke, 2003) and their transmission (Gelman & Roberts, 2017). Important concepts are lexicalized (e.g. “coffee”, “computer”, “DNA”) and labels are used in generic statements to convey semantic knowledge that extends beyond the here and now (e.g. “Coffee keeps people awake”). The development of these devices of cultural transmission starts in the form of word learning in early infancy. Although infants witness only particular things being named for them, their mental lexicon is not a catalogue of sounds paired with those items. Rather, it contains labels standing for categories of objects, such as body parts, food items, or artefacts (Bergelson & Swingley, 2012; Parise & Csibra, 2012). To date, however, it remains unknown how word meanings enabling the storage and generalization of cultural knowledge develop. Two main answers have been suggested.

On one hand, some have argued that the discovery of word meanings is initially guided by attentional biases (Smith, Jones, & Landau, 1992) that make children attend to salient perceptual features and interpret words as names for objects sharing these features (Smith, 2003). For example, toddlers tend to generalize new words to objects of the same shape (a phenomenon known as *shape bias*, Colunga & Smith, 2005). However, experience with language is needed for this strategy to emerge (Smith et al., 2002; Smith & Samuelson, 2006), and children below 2 years of age struggle to use it (Son, Smith, & Goldstone, 2008). This observation is in stark contrast with the evidence that preverbal infants successfully generalize familiar words (Bergelson & Swingley, 2012; Parise & Csibra, 2012), suggesting that before shape bias develops, other cognitive mechanisms must support the acquisition of word meanings.

On the other hand, it has been proposed that language builds on preexisting conceptual knowledge (Clark, 2017; Mandler, 2004). Indeed, it has been widely documented that infants have access to some innate concepts (e.g. object, agent, Carey, 2009) and readily learn a variety of novel categories, both with help from communication (Ferguson & Waxman, 2016; Ferry, Hespos, & Waxma, 2013) and, more importantly, without it (Mandler, 2004; Mareschal & Quinn, 2001). Preverbal categories could become the meanings of new words thanks to a specialized word-learning mechanism, referred to as *category matching* (Macnamara, 1982; Nelson, 1973). According to this account, infants seek to map new words onto categorical representations already available in their conceptual repertoire. Once the new word becomes linked to a specific category, word generalization beyond the labeling context comes for free as all objects identified as members of this category fall under the label describing it. Such a mechanism would ensure fast word learning from minimal input. Despite the consensus that language must, to some degree, be built upon our prelinguistic representations (Clark, 2004), to date there is no direct evidence that category knowledge can be directly used as a source of word meanings in infancy.

Here, we asked whether human infants possess special cognitive mechanisms, such as *category matching*, that ensure efficient word learning and open the way for the acquisition of knowledge conveyed through language. More specifically, across three experiments, we investigated whether preverbal categories can be accessed by infants and interpreted as word meanings when only one of the category tokens is named. First, we tested whether 12-month-olds use preverbal category knowledge to identify the meanings of new nouns and extend them to previously unseen category exemplars (Experiment 1). Then, we explored whether preverbal categories were necessary and sufficient to explain the observed word extension effects (Experiments 2 and 3).

Experiment 1

Twelve-month-olds viewed objects from two new categories, either in a category-training procedure or in a manner not conducive to category learning. Then, they saw a single exemplar of each category being labeled. We reasoned that, when given the opportunity to learn categories before labeling, infants would engage in word learning through category matching and extend the new labels to previously unseen exemplars of these categories.

Methods

Participants. In total, 36 healthy, full-term 12-month-olds participated in the experiment. Eighteen infants were assigned to the blocked-categories group (10 females, average age 12.11 months, range 11.64 to 12.76 months) and 18 to the interleaved-categories group (8 females, average age 12.06 months, range 11.73 to 12.65 months). An additional 7 infants were tested but excluded from the final sample as they did not provide enough data (*n* = 4; see Data Analysis) or failed to complete the experiment (*n* = 3). Families were recruited on a voluntary basis through advertisement in the local press. All infants were raised in monolingual English-speaking families and written informed consent was received from all caregivers. The target sample size of 36 was determined a priori based on previous studies that have investigated word-learning and word-extension in infancy using eye tracking (e.g. Yin & Csibra, 2015). We acknowledge, however, that more optimal approaches for determining sample sizes (i.e. taking into account publication biases) have been recently suggested (e.g. Anderson, Kelley, & Maxwell, 2017).

Apparatus. Infants' binocular gaze data were acquired using a Tobii TX 300 eye-tracker (Tobii Technology AB, Danderyd, Sweden) with a 23-inch integrated monitor (sampling rate: 120 Hz, resolution: 1920 x 1080 px). The sound was delivered via external stereo speakers. The experiment was administered using custom-built Matlab scripts with the Psychophysics Toolbox for stimuli presentation (Brainard, 1997) and the Tobii Pro Analytics software development kit for data collection (https://www.tobiipro.com/product-listing/tobii-pro-sdk/).

Stimuli

Pictorial stimuli. We selected two types of objects that were unfamiliar to the infants: coffee makers and staplers (Fig. 1). Prior to the experimental session, the caregivers confirmed that the infants had not been previously exposed to these objects. Multiple color photographs (*n* = 15) were selected for each type of object. The depicted items differed in color, orientation, and shape, but had a similar surface area. All stimuli were presented on a white background around 15 cm x 15 cm in size that subtended approximately 14º of the visual angle. For each object category, a subset of 13 images was randomly selected (different for every participant): 8 images for the familiarization, 1 for the word learning, and 4 for the word recognition test.

Speech stimuli. The speech stimuli were two phonetically distinct pseudowords that conformed to English phonotactics: *rif* and *toma*. As infants process labels that are presented in sentence frames more efficiently (Fenell & Waxman, 2009), the words were embedded in carrier phrases (see Design and Procedure). All speech stimuli were recorded by a female native speaker of British English, using infant-directed speech.

Design and Procedure. The experiment consisted of three phases administered without pause: familiarization, word training, and word-extension test, for a total length of about 4 minutes. Infants were randomly assigned to one of the two groups, which differed only with respect to the type of familiarization, i.e. *blocked* or *interleaved categories*. Infants were tested individually in a soundproof, dimly lit room. They sat in a car seat approximately 60 cm away from the monitor. To obtain a reliable eye-tracking signal, a five-point calibration sequence was used prior to the recording and was repeated until all points were successfully calibrated.

Familiarization. Each infant was exposed to a set of objects consisting of 16 visual items, 8 tokens per category. Images were introduced by sliding into the geometrical center of the picture located within a preselected area (1580 x 620 pixels) positioned in the middle of the display. Each image was presented for 1.5 s, zooming in and out, before leaving the display by sliding out.

In both groups, infants received two familiarization streams containing 8 items each. In the *blocked-categories* group, one familiarization stream consisted of 8 items from category A and the other of 8 items from category B (Fig. 1). The order of item appearance and the order of category presentation (A v. B) were randomized. In the *interleaved-categories* group, both familiarization streams included 4 items from one category interleaved with 4 items from the other category. The items were randomized with the constraint that no more than two tokens of the same category could be presented in succession.

Younger and Fearing (1999) demonstrated that, when presented with items from two distinct real-world categories (i.e. CAT and HORSE) interleaved within a single familiarization stream, infants had difficulty forming two differentiated categories. In contrast, infants excel at forming categories when familiarized with items of a single real-world category (i.e. CAT and HORSE, see Eimas & Quinn, 1994).

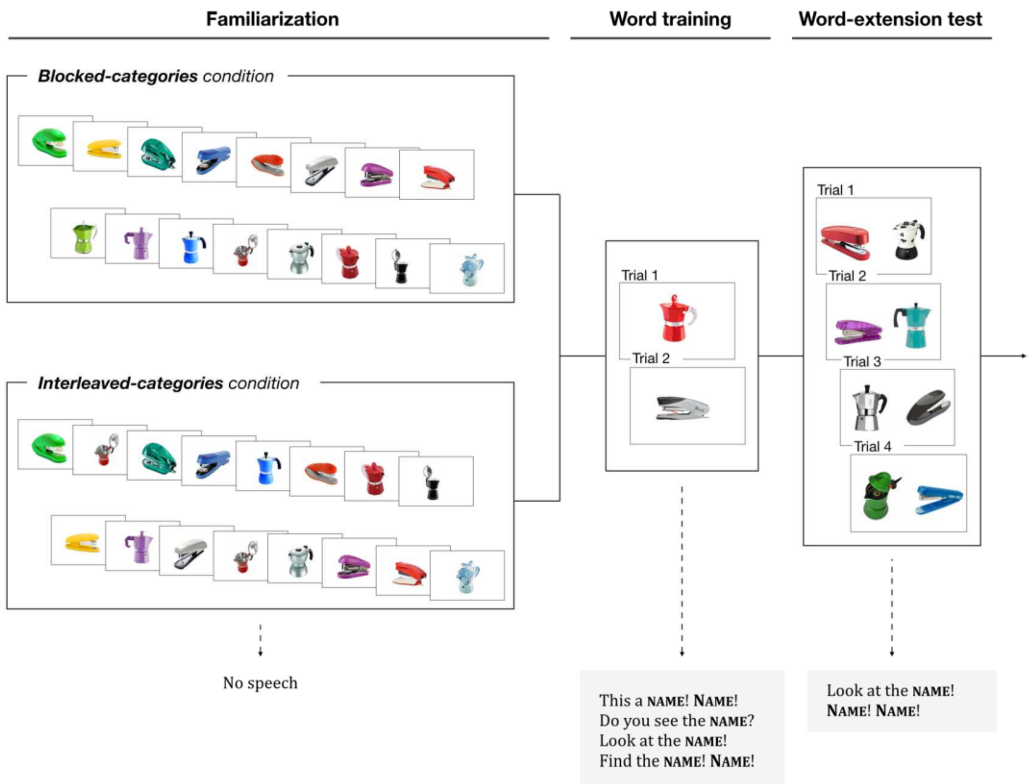

Fig. 1. Schematic visualization of the experimental design in Experiment 1. The task had three phases: familiarization, word training, and word-extension test. Only the familiarization phase differed between the conditions with respect to whether the tokens of the two target categories were blocked into separate familiarization streams (the *blocked-categories* group) or not (the *interleaved-categories* group). Speech stimuli were administered only during the word training and word-extension test (see bottom row). The depicted object images represent a sample stimuli sequence corresponding to one experimental session. Each category token was presented only once.

Moreover, grouping or highlighting various exemplars of the same category is a strategy employed by teachers when they intend to demonstrate category- as opposed to exemplar-specific properties (Shafto, Goodman & Griffiths, 2014). Therefore, the presentation of category A and category B tokens in two separate familiarization streams should lead to formation of two distinct categories, whereas a presentation of tokens from category A interleaved with tokens from category B should lead to weaker categorization performance or a lack thereof (see also Experiment 3).

Word training. Following familiarization, infants were introduced to two novel words, each paired with a previously unseen exemplar of one of the target categories. Each object was labeled six times: "*This a NAME! NAME! Do you see the NAME? Look at the NAME! Find the NAME! NAME!*" The pictures were kept still during the delivery of the auditory stimuli, but moved to the four corners and the center of the screen (randomized) in between the labeling phrases to maximize the infants' interest in the procedure. The label-category pairings and the order of presentation of a particular label-category pair during the training were counterbalanced across infants.

Word-extension test. The test phase comprised 4 trials; 2 per word. All test trials had the same structure: after fixating on a centrally displayed attention getter, infants were exposed to two novel images (one from each category), different at each of the four test trials. The images were displayed side-by-side for 10 seconds. Only one label was provided during each test trial. The label was uttered three times: first, embedded in a carrier phrase and then twice as a single word with short periods of silence between presentations, i.e. "*Look at the NAME! NAME! NAME!*"

Images were presented in silence for 2 seconds before the onset of speech so that we could measure the baseline preference for the two objects. The three tokens of the target label were delivered after 2.8 seconds, 5 seconds, and 7.5 seconds, respectively. The side on which the images appeared and the order of word and image presentation were counterbalanced across participants. There were two possible word orders: ABAB and BABA.

A short attention getter was presented in the center of the display before every familiarization stream, every word-training trial, and every word-extension test. The attention getter was gaze-controlled by the infants. That is, its duration was not predefined, but it depended on the infants' looking behavior. The attention getter was displayed on the screen until infants fixated on it continuously for 500 ms.

Data analysis. Infants had to fulfill two criteria to be included in the final sample: (1) minimum 50% attendance to the screen during each of the three experimental phases, (2) contribution of a minimum of one valid test trial for each word (i.e. a trial with more than 50% attendance to the test stimuli computed across both pre- and post-naming). Infants in both groups spent a comparable amount of time attending to the stimuli (blocked-categories group: M = 77% of total session time, SD = 11%, interleaved-categories group: M = 78%, SD = 9%, $t(34) = .09$, $p = .930$) and provided a comparable amount of valid test trials (blocked-categories group: M = 3.61 trials, SD = 0.70, R = 2 to 4; interleaved-categories group: M = 3.78 trials, SD = 0.55, R = 2 to 4, $t(34) = .80$, $p = .430$).

To analyze infants' looking behavior during test trials, we defined two regions in the test display, one corresponding to the target and one to the distracter object. Following the terminology used in the eye-tracking literature, we refer to these regions as areas of interest (AOI): the target AOI and the distractor AOI, respectively. We investigated two indices of word comprehension: the *proportion of target looking* and the *longest look difference scores*. The *proportion of target looking* (PTL) was computed by dividing the amount of time spent in the target AOI by the total amount of time spent in the target and distracter AOIs. Using proportional rather than absolute looking times ensure that the effects are not driven by infants who display overall longer looking times. The longest look was defined as the longest time spent within a single visit to an AOI (a visit corresponds to the sum of fixations within the AOI not separated by a fixation outside of the AOI). The *longest look difference scores* (LLDS) were calculated by subtracting the longest look to the target from the longest look to the distractor and then dividing this difference by the sum of longest looks to both the target and distractor. The difference scores ranged from -1 to 1, with positive values indicating that infants directed their longest look to the target.

To assess the impact of the speech stimuli delivered during the test on the infants' looking behavior, test trials were divided into pre-naming and post-naming segments and both word-comprehension measures were derived separately for each segment (for the time course of infants' target fixation see Figure S1 in the supplemental material available online). The post-naming segment started 367 ms after the onset of the first token of the target word (Swingley, Pinto & Fernald, 1999) and lasted until the end of the trial (i.e. the pre-naming lasted for 3.167 seconds, whereas the post-naming lasted 6.833 seconds). An increase from pre- to post-naming in the looking toward the target as indexed by the PLT or LLDS was taken as evidence for word extension.

Results

To assess infants' word extension performance, the mean proportions of target looking time during the pre- and post-naming segments of the test were entered into a two-way mixed ANOVA with group (blocked-categories familiarization v. interleaved-categories familiarization) as a between-subject factor and segment (pre- v. post-naming) as a within-subject factor (Fig. 2). This analysis yielded a significant main effect of segment, $F(1,34) = 4.29$, $p = .046$, $\eta_p^2 = .11$, and an interaction between group and segment, $F(1,34) = 12.39$, $p = .001$, $\eta_p^2 = .27$. In line with our predictions, follow-up paired-samples *t* tests revealed that this interaction was due to the increased target looking from the pre- to the post-naming segment in the blocked-categories group, *t*(17) = 3.69, *p* = .002, *d* = .87, 95% CI = [.04, .14] (pre-naming: M = .51, SD = .08, post-naming: M = .60, SD = .12), with 15/18 infants showing this effect, but not in the control group, *t*(17) = 1.11, *p* = .283, *d* = .26, 95% CI = [-.02, .07] (pre-naming: M = .52, SD = .17, post-naming: M = .49, SD = .16). Moreover, following naming, only infants who received category training looked at the named object significantly more than expected by chance (.50), *t*(17) = 3.58, *p* = .002, *d* = .84, 95% CI = [.54, .66] (interleaved-categories group: *t*(17) = .23, *p* = .82, *d* = .05, 95% CI = [.41, .57]). Prior to naming, neither group displayed a preference for either of the test stimuli (blocked-categories group: *t*(17) = .60, *p* = .558, *d* = .14, 95% CI = [.47, .55], interleaved-categories group: *t*(17) = .41, *p* = .690, *d* = .09, 95% CI = [.43, .60]). Please note that for two-samples tests we report 95% confidence intervals (CI) for the difference in the mean of the dependent variable and for one-samples tests we report 95% CIs for the mean.

A two-way mixed ANOVA on the average longest look difference scores yielded a significant interaction between group and segment: *F*(1,34) = 15.37, $p < .001$, $\eta_p^2 = .31$. Infants' looks towards the target increased in duration after labeling in the blocked-categories group, *t(17)* = 4.06, *p* = .001, *d* = .96, 95% CI = [.13, .42] (pre-naming: M = -.01, SD = .20, post-naming: M = .26, SD = .26), with 16/18 infants displaying this pattern of results, but not in the interleaved-categories group, *t*(17) = 1.49, *p* = .154, *d* = .35, 95% CI = [-.04, .25] (pre-naming: M = .06, SD = .35; post-naming: M = -.04, SD = .37). Before labeling, the difference scores were at chance in both groups (blocked-categories group: *t*(17) = .29, *p* = .776, *d* = .07, 95% CI = [-.11, .09], interleaved-categories group: *t*(17) = .76, *p* = .458, *d* = .18, 95% CI = [-.11, .24]). After labeling, the longest look difference scores increased above chance in the blocked-categories group, $t(17) = 4.32$, $p < .001$, $d = 1.02$, 95% CI = [.13, .39], but not in the interleaved-categories group, $t(17) = .46$, $p = .65$, $d = .11$, 95% CI = [-.22, .14].

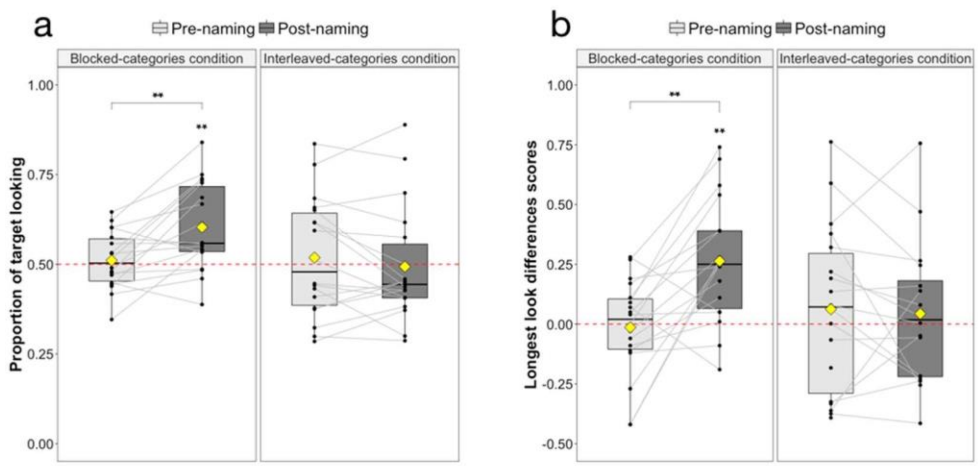

Fig. 2. Results of Experiment 1: (a) mean proportions of target looking time and (b) mean longest look difference scores during test, split by pre- v. post-naming, for the blocked- and interleaved-category groups. Diamonds indicate means and black horizontal lines indicate medians. The bottom and the top of the boxes represent the first and the third quartiles. Whiskers extend from the middle quartiles to the smallest and largest values within 1.5 times the interquartile range. Black points connected across boxes represent the data sets of individual participants. Dotted lines represent the chance level: (a) 0.5, (b) 0. Significant effects are indicated with asterisks: ** $p < .01$. Positive values in (b) indicate that longest looks were directed at the target and negative values that they were directed at the distracter.

Discussion

Our results suggest that preverbal categories are used by infants as meanings of newly encountered labels and enable them to determine the extensions of these labels after only a single naming episode. However, two alternative explanations of our findings should be considered. First, although unlikely (Son, Smith, & Goldstone, 2008), it remains possible that infants could generalize new words without prior category knowledge. We addressed this possibility in Experiment 2 by testing whether removing the category training would affect infants' word extension performance. Second, the chance performance in the interleaved-category group might have been due to learning a single superordinate-like category that included both coffee makers and staplers. Children are reluctant to attach multiple names to a single object (a phenomenon known as *mutual exclusivity,* Markman & Watchel, 1988), so they might have disregarded the naming events because objects from the single superordinate category were given two different names. We investigated this issue in Experiment 3, by testing whether the interleaved-category presentation of two kinds of objects leads to formation of a single category.

Experiment 2

To establish whether word generalization depends on category knowledge, we modified the task used in Experiment 1 by removing the familiarization phase. Furthermore, we administered an additional word-recognition test that followed the word-extension test. By using the same objects as presented during the labeling events, this test was meant to investigate whether it is specifically the ability to generalize words, and not the ability to map words onto the specific objects, that is impaired without category knowledge.

Methods

Participants. Eighteen healthy full-term 12-month-olds (6 females, average age 12.08 months, range 11.5 to 12.9 months) participated in the experiment. Thirteen other infants were excluded from the analysis because they did not provide enough data (*n* = 10; following the same exclusion criteria as in Experiment 1), because the experiment was interrupted (*n* = 1), because of a technical failure (*n* = 1), or because of being born preterm (*n* = 1). Families were recruited and gave consent to participate as in Experiment 1. Sample size was selected to match that of Experiment 1.

Apparatus and stimuli. The experimental setup and stimuli were identical to those used in Experiment 1. For each object category (coffee makers and staplers) a subset of 5 pictures was randomly selected (i.e. different for each participant) that included 1 picture for the word learning phase and the word recognition test and 4 for the word extension test. The order of image presentation was randomly determined for each participant.

Design and procedure. The design and procedure were the same as in Experiment 1, except as follows: first, the word training phase was not preceded by any familiarization. Instead, to ensure that the infants attended to the screen before the word training started, we presented them with a five-second video clip depicting flower buds opening to the sound of soft music. Second, we administered two types of word-learning tests of 4 trials each to independently assess two dimensions of infants' word-learning performance: *word extension* and *word recognition.* The word extension test was identical to the one in Experiment 1 and consisted of two measurement periods: pre- and post-naming (see Experiment 1).

The word recognition test had the same structure as the word extension test, except that instead of using novel category exemplars, we used the same tokens as presented during the word training test. The order of presentation of the two tests was fixed to enable direct comparisons with Experiment 1, i.e. the word extension test was always administered immediately after the word training and before the word recognition test. Word-training and test trials were presented after a short gaze-controlled attention getter (minimum duration: 500 ms).

Data analysis. The participant inclusion criteria and data analysis were identical to those in Experiment 1. We conducted two separate analyses for each type of test. Word-extension analysis consisted of 18 data sets, while word-recognition analysis consisted of 15 data sets as three participants did not contribute at least two valid word-recognition trials. On average, the included data sets contained 3.78 valid word-extension test trials (SD = .43, R = 3 to 4) and 3.33 valid word-recognition test trials (SD = .82, R = 2 to 4).

Results

Word extension. Unlike in Experiment 1, infants did not increase their looking to the named category exemplars from pre- to post-naming, as revealed by a paired-samples *t* test comparing the proportions of target looking between pre- and post-naming, $t(17) = .36$, $p = .723$, $d = .08$, 95% CI = [-.06, .08] (pre-naming: M = .49, SD = .08, post-naming: M = .50, SD = .10). Their looking level remained at chance throughout the test trial (pre-naming: $t(17) = .62$, $p = .545$, $d = .14$, 95% CI = [.45, .53], post-naming: $t(17) = .02$, $p = .98$, $d < .01$, 95% CI = [.45, .55]). The duration of the infants' looks directed to the target and distractor objects was not affected by labeling, $t(17) = .86$, $p = .402$, $d = .20$, 95% CI = [-.07, .16] (pre-naming: M = -.05, SD = .13; post-naming: M = -.01, SD = .21), with difference scores remaining at the chance level during both pre-naming, $t(17) = 1.73$, $p = .101$, $d = .41$, 95% CI = [.-12, .01], and post-naming, $t(17) = .14$, p = .888, $d = .03$, 95% CI = [-.11, .10]. Thus, in this experiment, infants did not generalize novel labels. This result is in line with the literature documenting difficulties in word extension without relevant category knowledge in older children (Son, Smith, & Goldstone, 2008).

Word-extension results were compared across experiments using mixed-model ANOVAs with group (blocked-categories familiarization in Experiment 1 v.

interleaved-categories familiarization in Experiment 1, v. no familiarization in Experiment 2) as a between-subject factor and segment (pre v. post-naming) as a within-subject factor. The analysis on the proportions of target looking yielded a significant interaction between group and segment, $F(2,51) = 4.51$, $p = .016$, $\eta_p^2 = .15$. This interaction was due to the fact that speech affected infants' looking patterns differently across groups. That is, before naming the duration of infants' target looking was comparable across groups, $F(2,51) = .30$, $p = .741$, $\eta_p^2 = .01$, while after naming it varied across groups, $F(2,51) = 4.19$, $p = .021$; $\eta_p^2 = .14$). Post-hoc comparisons indicated that, after naming, blocked-categories familiarization led to longer target looking than interleaved-categories familiarization, $p < .036$ (Bonferroni corrected), and marginally longer than no familiarization, $p < .061$ (Bonferroni corrected). The analysis of longest look differences scores yielded a similar pattern of results: the interaction between group and segment, $F(2,51) = 8.96$, $p < .001$, $\eta_p^2 = .26$ was due to differences in how naming affected the duration of longest looks across groups (pre-naming: $F(2,51) = 1.04$, $p = .362$, $\eta_p^2 = .04$; post-naming: $F(2,51) = 5.97$, $p = .005$, $\eta_p^2 = .19$). Blocked-categories familiarization resulted in longer looks to the target than each of the control conditions (v. interleaved-categories familiarization, $p = .003$; v. no familiarization $p < .007$, both values Bonferroni corrected). These results confirm that there is a difference in how infants with different levels of category knowledge extended novel words: namely, infants who received category training (i.e. blocked-categories familiarization, Experiment 1) performed better than infants who did not receive such training (i.e. interleaved-categories familiarization, Experiment 1, and no familiarization, Experiment 2).

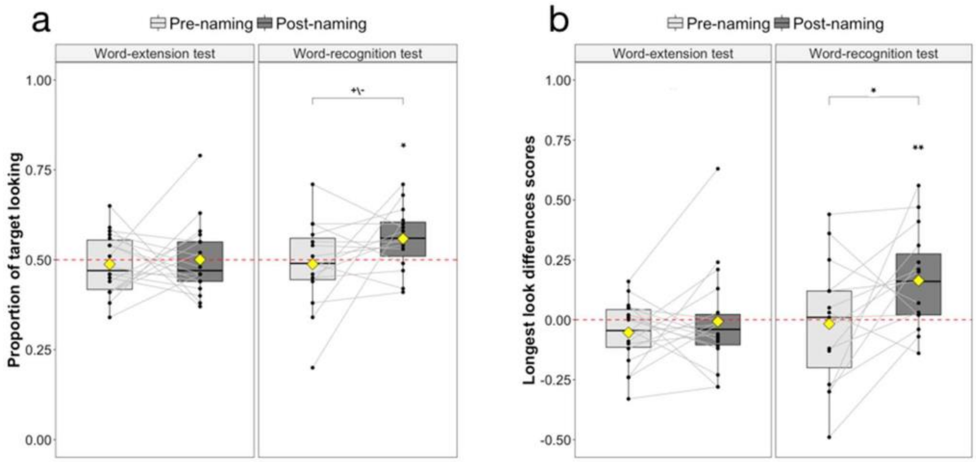

Fig. 3. Results of Experiment 2: (a) mean proportions of target looking time and (b) mean longest look difference scores during test split by pre- v. post-naming as a function of test type (word-extension v. word-recognition test). Diamonds indicate means and black horizontal lines indicate medians. The bottom and the top of the boxes represent the first and the third quartiles. Whiskers extend from the middle quartiles to the smallest and largest values within 1.5 times the interquartile range. Black points connected across boxes represent the data sets of individual participants. Dotted lines represent the chance level: (a) 0.5, (b) 0. Significant effects are indicated with asterisks: +/- $p \leq .10$, * $p < .05$, ** $p < .01$. Positive values in (b) indicate that longest looks were directed at the target and negative values that they were directed at the distracter.

Word recognition. Infants displayed a tendency to increase their target looking between pre- and post-naming, $t(14) = 1.75$, $p = .101$, $d = .45$, 95% CI = [-.02, .16] (pre-naming: M = .49, SD = .12, post-naming: M = .56, SD = .09). Eleven out of 15 infants showed this effect. In addition, the proportion of target looking was significantly above chance during the post-naming period, $t(14) = 2.62$, $p = .020$, $d = .68$, 95% CI = [.51, .61], but not during the pre-naming period, $t(14) = .38$, $p = .710$, $d = .10$, 95% CI = [.42, .56]. The duration of infants' longest looks toward to the target object increased significantly from the pre- to post-naming, $t(14) = 2.27$, $p = .039$, $d = .59$, 95% CI = [.01, .35] (pre-naming: M = -.02, SD = .26; post-naming: M = .16, SD = .21), with 12/15 participants displaying this effect. The difference scores were significantly above chance during the post-naming, $t(14) = 3.06$, $p = .008$, $d = .79$, 95% CI = [.05, .28], but not during the pre-naming, $t(14) = .26$, $p = .800$, $d = .07$, 95% CI = [-.16, .13].

Discussion

Without nonverbal category knowledge, infants failed to extend novel labels to further exemplars of the labeled category, while showing a tendency to identify by name the specific objects that were used during labeling. Although we cannot conservatively conclude that infants mapped the trained words onto the particular objects that were labeled during the training, as the increase in the proportion of target looking after labeling failed to reach significance, the longest look results suggest that such an association has been formed. This discrepancy between the two measures may be explained by the fact that the longest look measure is not affected by the decrease of attention over time, as is the case for total duration of looking (Schafer & Plunket, 1998). Note also that the word-recognition test was always administered following the word-extension test, meaning that infants had to maintain the newly formed label-object mappings over the short delay required to complete the first test phase and to handle possible interference with processing it.

To sum up, word extension does not occur spontaneously following the exposure to labeling and suggests that word extension is independent from word mapping. In particular, witnessing an unfamiliar object being labeled did not provide infants with sufficient information to reliably and rapidly extend the label to other objects.

Experiment 3

As we did not directly assess category learning in Experiment 1, it remains possible that infants only formed the new categories during the labeling, prompted by the presence of labels. Furthermore, infants in the interleaved-categories condition, who saw staplers and coffee makers mixed in a single familiarization stream, might have formed one extensive category containing both kinds of objects. These possibilities were tested in Experiment 3.

Methods

Participants. Thirty-six 12-month-old infants were included in the final sample. Infants were randomly assigned to two experimental conditions, 18 to the blocked familiarization condition (10 females, average age: 12.11 months, range: 11.57 to 12.49 months) and 18 to the interleaved familiarization condition (8 females, average age: 12.15 months, range: 11.62 to 12.51 months). An additional 9 infants were excluded from the analysis because they did not complete the calibration routine (n = 1), failed to provide enough data (n = 6), or failed to complete the experiment (n = 2). Sample size was selected to match those of Experiments 1 and 2. Families were recruited and gave consent to participate as in Experiments 1 and 2.

Apparatus. Experiment 3 was conducted in a different laboratory than Experiments 1 and 2. The experimental set up was the same as for Experiments 1 and 2, except as follows: infants' gaze data were acquired using a Tobii Pro X2-60 eye-tracker (sampling rate: 60 Hz, Tobii Technology AB, Danderyd, Sweden) and the visual stimuli were displayed on an external 23-inch monitor (resolution: 1920 x 1080 px).

Stimuli. In addition to the two types of objects used in the previous experiments (familiarized categories: staplers and coffee makers), we selected one more type of object that was unfamiliar to the infants (novel category: garlic press) as confirmed through the parental reports. For the familiarized categories, we used the same photographs as in Experiment 1, out of which a subset of 10 was randomly selected for each participant: 8 images for the familiarization and 2 images for the categorization test. We selected 4 additional photographs depicting exemplars of the novel category from which a subset of 2 images was used during the test (randomly determined). The additional pictures varied in color, orientation, and shape, but were matched in size and surface area to the pictures used in Experiment 1.

Design and procedure. The task included only two phases: familiarization and categorization test, which were administered without pause. The **familiarization** phase was identical to that in Experiment 1, with half of the infants tested on the blocked-categories familiarization and the other half on the interleaved-categories familiarization. During the **categorization test** phase, infants were presented with two pairs of test pictures displayed side by side in silence: a previously unseen token from the familiarized category (stapler and coffee maker, in a randomized order) and a token of a novel category (garlic press). Each test pair was displayed for 6 seconds and then repeated with the objects locations (left v. right) swapped, which was in accordance with the designs used in the infant categorization literature (e.g. Plunkett, Hu, & Cohen, 2008). Each familiarization stream and each test trial were preceded by a short gaze-controlled attention getter (minimum duration: 500 ms). The procedure was the same as for Experiment 1.

Data analysis. The inclusion criteria were the same as in Experiment 1. Note, however, that the categorization test trials were shorter than the word-extension word trials, such that 50% of on-screen time corresponded to 3 seconds. Infants had to provide at least 2 valid test trials to be included in the final sample; one trial per familiarization category. Infants in both groups attended equally to the display (blocked-categories group: M = 83% of total session time, SD = 8%; interleaved-categories group: M = 82% of total session time, SD = 12%, $t(34) = .02$, $p = .984$) and provided a comparable amount of valid test trials (blocked-categories group: M = 3.89, SD = .32, R = 3 to 4; interleaved-categories group: M = 3.94, SD = .24, R = 3 to 4, $t(34) = .59$, $p = .559$).

Following the methodology established in the field of infant categorization, we assessed category formation by measuring infants' preference for the novel category tokens operationalized as the proportion of time spent looking at the novel category relative to the total looking time. Novelty preference is considered to be an index of category formation (e.g. Ferry, Hespos, & Waxman, 2013; Hu, Plunkett, & Cohen, 2008; Younger & Ferring, 1999). This reflects the fact that infants discriminate between the test images and perceive one of them as familiar and the other as novel based on their freshly formed category knowledge and orient preferentially to the novel stimulus. To determine whether infants reliably displayed a looking preference for the novel category, we divided the amount of time they spent in the novel category's AOI by the total amount of time spent in the novel and familiarized categories' AOIs (please refer to the on-line supplemental material for an additional analysis of the longest looks).

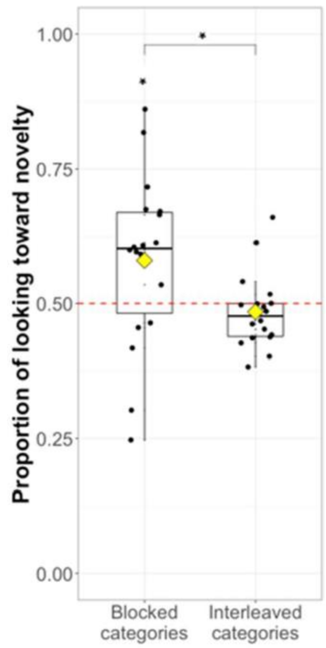

Fig. 4. Results of Experiment 3: average proportions of looking toward novel category tokens during test. Diamonds indicate means and black horizontal lines indicate medians. The bottom and the top of the boxes represent the first and the third quartiles. Whiskers extend from the middle quartiles to the smallest and largest values within 1.5 times the interquartile range. Dotted lines represent the chance level: 0.5. Significant effects are indicated with asterisks: * $p < .05$.

Results

Infants in the blocked-categories group looked longer at the novel category tokens (M = 0.58, SD = 0.16) than infants in the interleaved-categories group (M = 0.48, SD = 0.06), $t(23.25) = 2.34$, $p = .028$, $d = 0.78$, 95% CI = [.01, .18] by a Welch's *t* test. Only in the blocked-categories group did infants look towards the novel object more than expected by chance, $t(17) = 2.14$, $p = .047$, $d = .50$, 95% CI = [.51, .66] (interleaved-categories group: $t(17) = .96$, $p = .351$, $d = .23$, 95% CI = [.45, .52]).

Discussion

First, category formation occurred in the absence of language in the blocked-categories group under conditions identical to those in Experiment 1, indicating that in Experiment 1 categorical representations were available to infants before labeling. Second, there was no evidence for category formation in the interleaved-categories group. It is therefore unlikely that infants' failure to generalize new words in Experiment 1 could be explained by the formation of a single superordinate category spanning both target categories (staplers and coffee makers), which interfered with the word mapping processes.

General Discussion

The current results provide the first experimental evidence that preverbal category knowledge bootstraps the discovery of word meanings. We found that 12-month-olds rapidly grasped the meanings of two novel labels, each introduced with an object representing a different basic-level category. Infants' understanding of new words was revealed via their word extension performance: they succeeded at extending the new labels to previously unseen tokens of the referent categories. Importantly, infants could do so only when given the chance to learn the two categories before the new labels were introduced (i.e. blocked-categories familiarization, Experiment 1). Infants who did not have relevant category knowledge before labeling, either because they were exposed to the exemplars of labeled categories in a manner not conducive to category formation (i.e. interleaved-categories familiarization, Experiment 1), or because they were presented with the word training without any prior exposure to the category (Experiment 2), failed to extend novel words (see also, Son, Smith, &

Goldstone, 2008).

Our results suggest that specialized cognitive mechanisms allow preverbal infants to rapidly acquire basic linguistic devices of cultural transmission. In particular, infants rely on category-matching to identify the meanings of new labels. We demonstrated that even newly formed visual categories can be linked to novel words as their meanings and can guide the infants' word extension. Our findings by no means exclude the possibility that, later in development, attentional biases developing in children due to their experience with language (Gershkoff-Stowe & Smith, 2004) play a role in word learning (Son, Smith & Goldstone, 2008; Smith et al., 2002), but highlight the fact that knowledge of preverbal categories contributes to word-generalization processes before that happens. Note that, in the absence of relevant category knowledge, infants successfully associate new labels with particular objects (Pruden et al., 2006; Woodward, Markman, & Fitzsimmons, 1994; suggestive evidence in Experiment 2). Therefore, nonverbal category knowledge is not a prerequisite for associating labels and the specific objects used during labeling, but enables infants to go beyond these associations.

How is category matching achieved? One possibility is that infants expect objects introduced through communication to represent categories. According to Csibra and Shamsudheen (2015), the labeled object is treated as a symbol standing for its category and the new label is directly attached to a category-level representation. In this scenario, category knowledge is used *during* labeling. Alternatively, the category knowledge might only be used at the stage of word extension. Initially, infants associate the label with a particular object, then upon hearing the new label without its original referent at test, they retrieve the category to which this object belongs and extend the label along the boundary of this category. Further research is needed to establish which of these two strategies is at play.

We propose that the category matching mechanism is available for word learning in the real world, but remain aware that the categories used by infants in our study were taught in a lab in the somewhat artificial manner of assembling category tokens. However, category tokens also occur together outside labs (e.g. crossing a London street crowded with cars, strolling in a park full of dogs and trees, or just walking through a supermarket filled with orderly presented products). This is sometimes due to the structuring of our environment for learning purposes. For example, caregivers who are asked to teach the properties of a category automatically select numerous category exemplars rather than a single one to highlight the category-diagnostic properties (Rhodes, Gelman, & Brickman, 2009). Infants, when allowed to freely manipulate objects, tend to explore exemplars of only one category in a sequence rather than alternating between exemplars of different categories (i.e. sequential touching, Rakison & Butterworth, 1998), thus selectively sampling the information in their environment. Even if this particular category learning strategy was never available in real life, our finding still stands as category-matching mechanisms should operate regardless of the way in which categories are learnt (Yin & Csibra, 2015; Waxman & Booth, 2003).

While in the past emphasis has been placed on demonstrating that labeling guides category learning (Ferguson & Waxman, 2017), our study provides the first evidence that visual categories themselves are directly exploited during word learning as sources of word meanings. One important implication of this observation is that individual differences in category learning may be a source of individual differences in the rate of language acquisition. For example, poor category learning in certain developmental disorders, such as autism (e.g. Davis & Plaisted-Grant, 2015), may explain the slower rate of vocabulary growth (Hudry et al., 2014).

To conclude, the current results not only constitute the earliest evidence for the generalization of newly learnt words, but also indicate the cognitive mechanisms underlying this ability. Infants use nonverbal category knowledge as a source of word meanings and in its earliest stages, the cognitive machinery responsible for category formation is a key component of language development. Therefore, nonverbal conceptual resources provide a stepping stone to language and culture transmission.

Author Contributions

B. Pomiechowska and T. Gliga conceived and designed the study, collected and analyzed the data, and wrote the manuscript.

Acknowledgements

We thank Katarina Begus, Gergely Csibra, Johannes Mahr, Eugenio Parise, and Denis Tatone for comments on the manuscript, the editor and two reviewers for helpful suggestions, Sinead Rocha for recording the speech stimuli, Eszter Körtveylesi, Krisztina Andrasi, Maria Toth, and Iti Arora for research assistance, and Chloe Taylor for proofreading the manuscript.

Funding

This research was supported by the European Comisison Marie Marie Curie Initial Training Networks (264301) and the UK Medical Research Council (G0701484).

References

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28, 1547-1562.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, USA*, 109, 3253–3258.

Booth, A. E., & Waxman, S. R. (2002). Word learning is 'smart': evidence that conceptual information affects preschoolers' extension of novel words. *Cognition*, 84, B11-B22.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433-436.

Clark, E. V. (2017). Semantic categories in acquisition. In H. Cohen & C. Lefebvre (Eds.), *Handbook of Categorization in Cognitive Science* (pp. 397-421). Amsterdam, Netherlands: Elsevier.

Clark, E. V. (2004). How language acquisition builds on cognitive development. *Trends in Cognitive Sciences*, *8*, 472-478.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, 112, 347-382.

Csibra, G., & Shamsudheen, R. (2015). Nonverbal generics: Human infants interpret objects as symbols of object kinds. *Annual Review of Psychology*, *66*, 689.

Davis, G., & Plaisted-Grant, K. (2015). Low endogenous neural noise in autism. *Autism*, 19, 351-362.

Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, *65*, 903-917.

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, 81, 1376-1383.

Ferguson, B., & Waxman, S. (2017). Linking language and categorization in infancy. *Journal of Child Language*, 44, 527-552.

Ferguson, B., & Waxman, S. R. (2016). What the [beep]? Six-month-olds link novel communicative signals to meaning. *Cognition*, 146, 185-189.

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2013). Nonhuman primate vocalizations support categorization in very young human infants. *Proceedings of the National Academy of Sciences*, 110, 15231-15235.

Gelman, S. A., & Roberts, S. O. (2017). How language shapes the cultural inheritance of categories. *Proceedings of the National Academy of Sciences*, 201621073.

Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, 75, 1098-1114.

Hudry, K., Chandler, S., Bedford, R., Pasco, G., Gliga, T., Elsabbagh, M., ... & Charman, T. (2014). Early language profiles in infants at high-risk for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 44, 154-167.

Kemler Nelson, D. G. K. (1999). Attention to functional properties in toddlers' naming and problem-solving. *Cognitive Development*, *14*, 77-100.

Kemler Nelson, D. G. K. (1995). Principle-based inferences in young children's categorization: Revisiting the impact of function on the naming of artifacts. *Cognitive Development*, *10*, 347-380.

Macnamara, J. (1982). *Names for things: A study of human learning*. Cambridge, MA: MIT Press.

Mandler, J. M. (2004). Thought before language. *Trends in Cognitive Sciences*, 8, 508-513.

Mareschal, D., & Quinn, P. C. (2001). Categorization in infancy. *Trends in Cognitive Sciences, 5*, 443-450.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, 20, 121-157.

Nelson, K. (1973). Some evidence for the cognitive primacy of categorization and its functional basis. *Merrill-Palmer Quarterly of Behavior and Development*, *19*, 21-39.

Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science, 23*, 728–733.

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, 106, 665-681.

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The Birth of Words: Ten-Month-Olds Learn Words Through Perceptual Salience. *Child Development*, 77, 266-280.

Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development*, 69, 309-320.

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, *71*, 55-89.

Smith, L. B. (2003). Learning to recognize objects. *Psychological Science*, 14, 244-250.

Smith, L. B., & Samuelson, L. (2006). An attentional learning account of the shape bias: Reply to Cimpian and Markman (2005) and Booth, Waxman, and Huang (2005). *Developmental Psychology*, 42, 1339-1343.

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, 13-19.

Son, J. Y., Smith, L. B., & Goldstone, R. L. (2008). Simplicity and generalization: Short-cutting abstraction in children's object categorizations. *Cognition*, *108*, 626-638.

Spelke, E. S. (2003). What makes us smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought*, (pp. 277-311). Cambridge, MA: MIT Press.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, 71, 73-108.

Waxman, S. R. (2003). Links between object categorization and naming. In D. H. Rakison & L. M. Oakes (Eds.) *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 213-241), New York, NY: Oxford University Press.

Waxman, S., & Booth, A. (2003). The origins and evolution of links between word learning and conceptual organization: new evidence from 11-month-olds. *Developmental Science*, *6*, 128-135.

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13-and 18-month-olds. *Developmental Psychology*, *30*, 553.

Yin, J., & Csibra, G. (2015). Concept-based word learning in human infants. *Psychological Science, 26*, 1316-1324.

Younger, B. A., & Fearing, D. D. (1999). Parsing items into separate categories: Developmental change in infant categorization. *Child Development*, 70, 291-303.