Journal of Experimental Child Psychology 172 (2018) 149–167

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Infants use linguistic group distinctions to chunk items in memory

Aimee E. Stahl [a,*], Lisa Feigenson [b]

[a] Department of Psychology, The College of New Jersey, Ewing, NJ 08628, USA
[b] Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD 21218, USA

ARTICLE INFO

Article history:
Received 8 August 2017
Revised 14 March 2018

Keywords:
Infants
Memory
Working memory
Chunking
Language
Social cognition

ABSTRACT

Although the capacity of infants' working memory is highly constrained, infants can overcome this limit via chunking; for example, they can use spatial cues to group individual objects into sets, thereby increasing memory efficiency. Here we investigated the use of abstract social knowledge as a basis for chunking. In four experiments, we asked whether 16-month-olds can use their sensitivity to distinctions between languages to efficiently chunk an array. Infants saw four identical dolls hidden in a box. Without chunking cues, infants in previous experiments fail to remember this number of items in such arrays. In Experiment 1, infants saw two of the four dolls each produce an utterance in a familiar language (English) prior to hiding and saw the other two dolls each produce an unfamiliar language (German or Mandarin). Infants successfully remembered all four dolls. Next we asked whether infants could chunk using linguistic group distinctions even when all dolls spoke unfamiliar languages. Infants failed to chunk speakers of unfamiliar languages when each doll within a pair produced a unique utterance (Experiment 2), but they succeeded when each doll within a pair produced the same utterance (Experiment 3). Infants' performance was not driven by low-level acoustical cues in the utterances given that infants failed to chunk when the dolls' speech was played backward (Experiment 4). Together, these results suggest that infants can leverage their early sensitivities to linguistic distinctions to hierarchically reorganize their memory representations, thereby overcoming working memory limits.

* Corresponding author.
E-mail address: stahla@tcnj.edu (A.E. Stahl).

https://doi.org/10.1016/j.jecp.2018.03.005
0022-0965/

Introduction

The ability to temporarily store information in memory over brief durations is limited in both adults and infants. For example, adults can remember three or four items in working memory at a given time, but they fail to remember larger arrays (e.g., Alvarez & Cavanagh, 2004; Brady, Konkle, & Alvarez, 2011; Cowan, 2001; Sperling, 1960). A variety of methodologies have revealed similar limits in infants, indicating that they too typically can remember up to but no more than three items in working memory (e.g., Barner, Thalwitz, Wood, & Carey, 2007; Feigenson & Carey, 2003, 2005; Feigenson, Carey, & Hauser, 2002; Feigenson & Halberda, 2004; Oakes, Hurley, Ross-Sheehy, & Luck, 2011; Ross-Sheehy, Oakes, & Luck, 2003; Zosh, Halberda, & Feigenson, 2011). For instance, 12- to 14-month-olds will persistently search a hiding location after three objects are seen hidden and only a subset (i.e., one or two objects) is retrieved, showing that they successfully represented the three hidden objects and detected a mismatch with the number of retrieved objects. However, when four objects are hidden, infants fail to keep searching after *any* subset is retrieved (e.g., Barner et al., 2007; Feigenson & Carey, 2003, 2005).

Despite this signature limit on working memory in adults and infants, the process of "chunking" representations can allow observers to overcome typical memory capacity limits. Chunking involves hierarchically reorganizing memory representations by binding representations of individual items into sets so that the observer represents both the higher-order set (the "chunk") and the individual items within the set. This strategy allows observers to maintain more items in working memory than they otherwise could. Adults can employ a host of different cues for chunking, including low-level perceptual features, category membership, semantic relatedness, and statistical co-occurrences between items (e.g., Bower, 1972; Brady, Konkle, & Alvarez, 2009; Chase & Ericcson, 1982; Chase & Simon, 1973; Cowan, 2001; Ericsson, Chase, & Faloon, 1980; Gobet & Clarkson, 2004; Hitch, Burgess, Towse, & Culpin, 1996; Mathy & Feldman, 2012; Miller, 1956; Simon, 1974).

More recent research shows that chunking does not require extensive experience, language, or explicit instruction; even infants spontaneously chunk items, indicating that chunking is likely a fundamental aspect of human memory (e.g., Feigenson & Halberda, 2004, 2008; Kibbe & Feigenson, 2016; Rosenberg & Feigenson, 2013; Stahl & Feigenson, 2014). For example, although 14-month-olds typically fail to store representations of four identical objects concurrently, they succeed if the objects were observed as two spatially separated groups of two prior to hiding (Feigenson & Halberda, 2004; Rosenberg & Feigenson, 2013) or if perceptually distinctive objects reliably co-occurred in pairs (e.g., red circle, blue cross) throughout the experiment (Kibbe & Feigenson, 2016). Infants also can use conceptual knowledge to chunk items. For instance, infants remember a four-object array containing two toy cats and two toy cars, but they fail to remember arrays of four identical cats or four unique cats (Feigenson & Halberda, 2008). Even younger infants (7-month-olds) can chunk visual arrays if provided with multiple redundant grouping cues (Moher, Tuerk, & Feigenson, 2012).

More recent evidence suggests that in addition to using spatiotemporal cues and knowledge of object kinds to support chunking, infants also can use their knowledge in the social domain. In this previous work (Stahl & Feigenson, 2014), an infant saw four identical dolls equally spaced. In one condition, the dolls initially faced each other in pairs. The experimenter then turned the dolls toward the infant and made each doll greet the infant by saying "hello." In the other condition, the four dolls initially faced the infant. The experimenter then turned the dolls toward each other in pairs and made them greet each other by saying "hello." Infants in both conditions then watched all four dolls hidden, saw just two of the four dolls retrieved, and then were allowed to search the box. Infants remembered all four dolls (i.e., persisted in searching for the two remaining dolls) only when the dolls had behaved in affiliative pairs by turning to greet each other prior to hiding. This cue apparently caused infants to represent the four identical dolls as two social dyads, and this mental reorganization improved infants' working memory performance.

Infants' use of social knowledge for chunking is impressive because information about social groups often is not directly available to perception and instead must be inferred from abstract knowledge and subtle cues. In the studies by Stahl and Feigenson (2014), it was the observed interaction of dolls with each other that formed the basis for infants creating representations of social groups. However, individuals also can be construed as members of groups even when they have never been observed to interact. Social categorization can be motivated by seeing that individuals share a socially relevant trait even in cases when the individuals are not present in the same place at the same time. A prime example of such a socially relevant trait is language.

The language individuals speak marks their membership in an important social category. It can indicate where people are from, what their preferences might be, and with whom they can effectively communicate. Infants are sensitive to differences among languages from early in life. In addition to preferring to listen to their own native language from birth (e.g., Byers-Heinlein, Burns, & Werker, 2010; Mehler et al., 1988; Moon, Cooper, & Fifer, 1993), infants also show language-based social preferences. They prefer to look at native speakers over foreign speakers, and they prefer to accept toys from and offer toys to native speakers over foreign speakers (Kinzler, Dupoux, & Spelke, 2007, 2012). Similarly, infants selectively learn novel actions from and trust native speakers over foreign speakers (Buttelmann, Zmyj, Daum, & Carpenter, 2013; Howard, Henderson, Carrazza, & Woodward, 2015; Shutts, Kinzler, McKee, & Spelke, 2009) and show attentional biases toward stimuli introduced by native speakers over foreign speakers (Soley & Sebastián-Gallés, 2015). Recent evidence further suggests that infants might prefer native speakers because they expect to be able to learn from them (Begus, Gliga, & Southgate, 2016). Finally, infants use linguistic information to make inferences about others' behavior. They expect individuals who share a language to interact prosocially, whereas they expect individuals who speak different languages to interact antisocially (Liberman, Woodward, & Kinzler, 2016), and they expect individuals who share a language to share similar preferences (Liberman, Woodward, Sullivan, & Kinzler, 2016).

Together, these results suggest that infants use spoken language as a cue to social group membership. However, it is unknown whether infants can further use these linguistic group distinctions to organize memory representations as in chunking. Can infants represent a speaker both as an individual entity and as a member of a language-defined set? If so, this type of hierarchical organization would potentially allow infants to expand the contents of their working memory. In the current experiments, we sought to characterize the ability of 16-month-olds to chunk items in memory using linguistic cues. In Experiment 1, we asked whether infants could chunk four identical dolls into two sets of two when two of the dolls spoke infants' own native language (English) and two of the dolls spoke an unfamiliar language (either German or Mandarin). In Experiments 2 and 3, we asked whether infants could chunk arrays when two dolls spoke one unfamiliar language (German) and two dolls spoke a different unfamiliar language (Mandarin). Finally, in Experiment 4, we tested the possibility that infants' performance in our experiments merely reflected sensitivity to low-level acoustical differences; we asked whether they would chunk four-object arrays when the two of the dolls produced the same "utterance" in backward speech and the other two dolls produced a different backward utterance.

Experiment 1

Across all experiments, we assessed infants' working memory using a manual search task (Barner et al., 2007; Feigenson & Carey, 2003, 2005; Feigenson & Halberda, 2004, 2008; Kibbe & Feigenson, 2014; Rosenberg & Feigenson, 2013; Stahl & Feigenson, 2014; Van de Walle, Carey, & Prevor, 2000). We showed infants arrays of identical dolls equidistantly positioned so as to avoid spatial grouping cues. Infants next watched all of the dolls placed out of sight in an opaque box, then saw either all or just a subset of the dolls retrieved from the box. The dependent measure was whether infants continued searching the box when there were missing dolls. If so, this would show that infants had successfully represented the hidden objects in working memory.

On some trials infants saw arrays that were well within their working memory capacity (one or two dolls), whereas on other trials they saw arrays that were beyond their typical working memory capacity (four dolls). In many previous experiments, infants who saw one or two dolls hidden searched the box appropriately (e.g., continuing to search for the missing doll after seeing two dolls hidden and only one doll retrieved), whereas infants who saw four objects hidden and any subset retrieved failed to continue searching for the remaining dolls (Barner et al., 2007; Feigenson & Carey, 2003, 2005; Feigenson & Halberda, 2004, 2008; Rosenberg & Feigenson, 2013; Stahl & Feigenson, 2014). In the current experiment, we asked whether language cues would help infants to remember these larger arrays. Before hiding, infants saw half of the dolls in the array produce a short utterance in a familiar language, English, and saw the other half produce an utterance in an unfamiliar language, either German or Mandarin. For example, when four dolls were presented, two dolls spoke English and two spoke German. We hypothesized that if infants can use language as a social grouping cue, and if this enables them to represent the four-object arrays as two sets of two, then they should successfully remember the objects in the hidden arrays even when the number of objects hidden surpasses their typical working memory limits.

Method

Participants

The participants were 16 healthy, full-term infants between 15 and 17 months of age (range = 15 months 18 days to 17 months 2 days, M = 16 months 14 days; 7 female). We focused on this age range because previous studies have found that infants between 12 and 20 months show an upper limit of three objects in working memory tasks (Barner et al., 2007; Feigenson & Carey, 2003, 2005; Feigenson & Halberda, 2004, 2008; Rosenberg & Feigenson, 2013; Stahl & Feigenson, 2014). An additional 5 infants were excluded due to fussiness (n = 1), parental interference (n = 1), or refusal to search in the box on more than 50% of measurement periods[1] (a criterion used in previous studies employing manual search; e.g., Stahl & Feigenson, 2014) (n = 3). In our final sample, 5 infants were reported to have had exposure to languages other than English from either one or both parents at least 5% of the time (2 infants were exposed to Spanish and 1 infant each was exposed to Korean, Japanese, and Italian).

Stimuli

Infants were presented with a black box made of foam core (40.5 × 25 × 12 cm). The front of the box had a rectangular opening covered in blue spandex (13 × 7.5 cm) that had a horizontal slit that allowed infants to reach through and retrieve objects but prevented them from seeing inside. The rear face of the box had a concealed opening covered with a black felt flap that allowed the experimenter to surreptitiously reach into the box and withhold objects. The stimulus objects were Lego Duplo dolls (6.5 cm high). Four of these were identical dolls with brown hair and a tan outfit; the other four were identical dolls with gray hair, glasses, and a blue and pink outfit. The contrast between doll types was included in order to maintain infants' interest throughout the experiment.

Prerecorded auditory stimuli were played from a hidden computer that the experimenter controlled with a remote. All of the stimuli were recorded by the same speaker, who was fluent in English, German, and Mandarin; this prevented infants from being able to use differences in vocal properties as a basis for chunking the stimuli. The utterances were "How are you doing?" and "Let's play a game!" in English, "Wie geht es dir?" and "Las uns spielen!" in German, and "Ni3 jin1-tian1 hao3-ma1 (你 今天 好吗)?" and "Lai2 he2 wo3 yi4-qi3 wan2 (来 和 我 – 起 玩)!" in Mandarin. The German and Mandarin utterances are approximate translations of the English.

Design

All infants participated in two blocks of trials. In one block (1 vs. 2 comparison), infants saw arrays that contained numbers of objects expected to be well within their working memory capacity. In the other block (2 vs. 4 comparison), they saw numbers of objects expected to be outside of their working memory capacity. Different dolls were used for each block. Which dolls were used in which block and the order of the trial blocks were counterbalanced across infants. In addition, to ensure that any observed effects were not specific to the unfamiliar language being used, English was paired with

[1] Across all experiments, infants who failed to search in the box on more than 50% of trials were equally likely to refuse to search when arrays were within typical working memory capacity (one or two dolls) and when arrays exceeded typical working memory capacity (four dolls). Hence, their refusal to search the box likely reflects task disengagement rather than failure to represent the hidden versus retrieved objects.

German in one block of trials and was paired with Mandarin in the other block. Which language was presented in each block was counterbalanced across infants, as was which language was presented first, the order of the sentences, and the number of objects hidden first within a block.

Procedure

Infants sat in a high chair at a low table, with the experimenter kneeling on the floor next to them. The session began by familiarizing infants to the box and encouraging them to practice reaching inside. Infants watched the experimenter hide a set of toy keys through the spandex-covered opening in the box's front face and were encouraged to retrieve them. Once they did, the experimenter said "Great job!" and hid the keys again, but this time secretly withheld them for 5 s in the back of the box (by reaching through the concealed opening in the box's rear face). After infants had searched inside the box, the experimenter showed infants as she reached inside through the front of the box and pulled out the keys, saying, "Was I hiding those? Look at what I found!" This was to show infants that sometimes objects could still be in the box but out of reach. This familiarization process was repeated again with a wooden block before the experiment began.

The 1 versus 2 comparison. In this block, we measured infants' searching following the hiding of one or two dolls. The block contained three different measurement periods, each presented twice: one object (none remaining), two objects (more remaining), and two objects (none remaining).

On one-object (none remaining) trials, the experimenter placed a single doll atop the center of the box, facing infants, while saying, "Watch this!" The experimenter then looked to ensure that infants were attending to the doll (if not, she attracted their attention by calling their name or tapping the box); this was done for all object presentations on both trial blocks. Then the experimenter moved the doll up and down slightly with one hand as she clicked the remote under the table with her other hand; this activated the auditory stimuli such that infants heard either "How are you doing?" or "Let's play a game!" in English, German, or Mandarin. Thus, it appeared that the doll was "speaking." If infants heard English on the first one-object trial (e.g., "How are you doing?"), they heard either German ("Wie geht es dir?") or Mandarin ("Ni3 jin1-tian1 hao3-ma1?") on the second trial. After the utterance had played, the experimenter pointed to the doll and said, "Look at this!" The experimenter let infants view the doll for approximately 2 s and then inserted it into the box while saying, "In we go!" She then pushed the box to the edge of the table and said, "What's in there?" Nearly all infants successfully reached in and retrieved the object. On the rare occasion that infants did not, the experimenter again asked, "What's in there?" If infants were still hesitant, the experimenter reached in and partially pulled out the object so that a small portion was visible through the spandex slit. If infants still did not retrieve the object, the experimenter retrieved it and momentarily gave it to them. Regardless of how the object was retrieved, once infants had seen it emerge from the box and had handled it briefly, the experimenter promptly took the doll away and placed it out of sight under the table.

At this point, a 10-s measurement period began during which infants' searching of the box was measured; this measurement period was called one object (none remaining) because infants had seen one object hidden and seen one object retrieved from the box, and so no objects remained inside. Throughout the measurement period the experimenter looked down and remained silent (so as not to influence infants' searching). After 10 s had passed, the experimenter said "Great job!" to end the trial. If infants were still searching at 10 s, the experimenter kept her head bowed until infants removed their hand from the box.

On two-objects (more remaining) trials (Fig. 1A), the experimenter placed two identical dolls atop the center of the box, facing infants, while saying, "Watch this!" Then she moved the first doll up and down slightly as she clicked the hidden remote to play the first utterance. For example, the doll said "Let's play a game!" (or "How are you doing?") in English. The experimenter then put that doll down, picked up the other doll, and activated the second utterance. The second utterance played in a foreign language and contained the same sentence but in either German (e.g., "Las uns spielen!") or Mandarin (e.g., "Lai2 he2 wo3 yi4-qi3 wan2!"). After the utterance had finished playing, the experimenter put down the second doll and then pointed to both dolls while saying, "Look at this!" Infants could view both dolls for approximately 2 s, and then the experimenter inserted them into the box simultaneously (so as to match the number of movements with the one-object trial) while saying, "In we

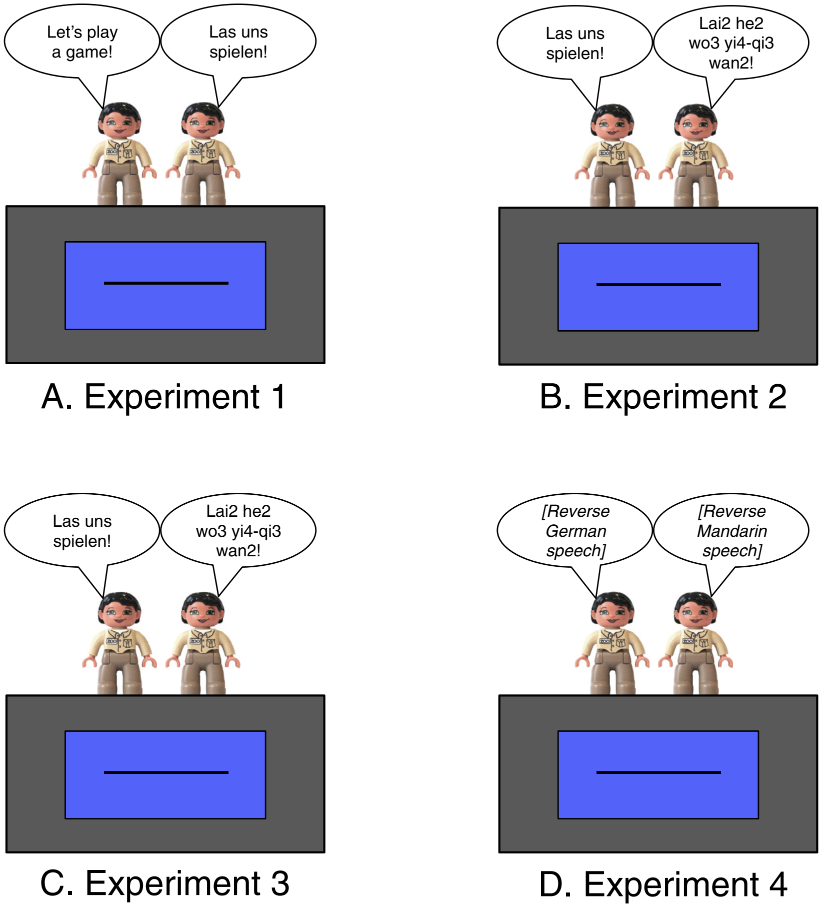

Fig. 1. Examples of two-object arrays and utterances in Experiments 1 through 4. Each doll was lifted in turn while the utterance played from underneath the table so as to appear that the doll was speaking. (A) In Experiment 1, one doll produced a token of English and one doll produced a token of either German (pictured) or Mandarin. (B) In Experiment 2, one doll produced a token of German and one doll produced a token of Mandarin. (C) In Experiment 3, one doll produced a token of German and one doll produced a token of Mandarin. (D) In Experiment 4, one doll produced a token of reversed German speech and one doll produced a token of reversed Mandarin speech.

go!" As the experimenter placed both dolls inside through the box's front face, she secretly reached with her other hand through the concealed opening in the back of the box to withhold one of the two dolls. She pushed the box toward infants and said, "What's in there?" Infants were then able to reach in and retrieve the doll that was not being withheld; nearly all infants spontaneously did so. The experimenter then immediately took the doll away, and a 10-s measurement period began (two objects, more remaining). After 10 s, the experimenter said "What else was hiding in here?" and showed infants as she reached through the front face of the box and "found" the second doll, showing it to infants briefly before placing it under the table. A final 10-s measurement period followed (two objects, none remaining) during which the box was again empty because all objects had been retrieved. After the measurement period ended, the experimenter said "Great job!" and pulled the box out of infants' reach.

Half of the infants heard the dolls say "How are you doing?" (in English and German or Mandarin) in this block of trials, and the other half heard the dolls say "Let's play a game!" (in English and

German or Mandarin); half of the infants heard English versus German, and the other half heard English versus Mandarin.

The 2 versus 4 comparison. In this block we measured infants' searching following the hiding of two or four dolls. The block contained three different measurement periods, each presented twice: two objects (none remaining), four objects (more remaining), and four objects (none remaining). Half of the infants heard English versus German, and the other half heard English versus Mandarin.

On two-objects (none remaining) trials, the experimenter placed two dolls atop the center of the box one at a time, facing infants, while saying "Watch this!" for each doll. Then the experimenter moved the first doll up and down slightly as she clicked the hidden remote to play the first utterance. For example, the doll appeared to say "How are you doing?" in English. The experimenter then put that doll down and picked up the other doll in the same manner. The second utterance played in a foreign language and said the same sentence but in either German ("Wie geht es dir?") or Mandarin ("Ni3 jin1-tian1 hao3-ma1?"). After the second utterance finished playing, the experimenter put down the second doll and then pointed and said "Look at this!" for each doll. Infants could view both dolls for approximately 2 s, and then the experimenter inserted them into the box one at a time (so as to match the number of movements with the four-object trial) while saying "In we go!" for each. The experimenter then pushed the box toward infants and said, "What's in there?" Infants reached in and retrieved a doll, which the experimenter took from them, and the experimenter then immediately reached into the box while saying "And look what else!" and retrieved the second remaining doll. She then showed infants both dolls before placing them out of sight under the table. A 10-s measurement period followed (two objects, none remaining). After 10 s, the experimenter said "Great job!" and pulled the box out of infants' reach.

On four-objects (more remaining) trials (Fig. 2A), the experimenter placed four dolls atop the box, two at a time, while saying "Watch this!" for each pair. The dolls all were evenly spaced facing infants. The experimenter then moved the first doll up and down slightly as she clicked the hidden remote to play the first utterance. For example, the doll said "How are you doing?" in English. The experimenter then put that doll down and picked up the second doll in the same manner, but this doll produced a different English utterance (i.e., "Let's play a game!"). This process was repeated for the remaining two dolls, each of whom spoke unique tokens of the same foreign language (either German ["Wie geht es dir?" followed by "Las uns spielen!"] or Mandarin ["Ni3 jin1-tian1 hao3-ma1?" followed by "Lai2 he2 wo3 yi4-qi3 wan2!"]). Each of the four sentences played approximately 1 s apart. The experimenter then pointed to each pair of dolls while saying "Look at this!" for each pair. Infants could view all four dolls for approximately 2 s before the experimenter inserted each pair of dolls into the box while saying "In we go!" for each pair. As she placed the dolls inside, the experimenter secretly reached through the opening in the back of the box with her other hand to withhold two of the four dolls. The experimenter then pushed the box toward infants and asked, "What's in there?" Infants were able to reach in and retrieve a doll, which the experimenter took from them. The experimenter then immediately reached into the box while saying "And look what else!" and retrieved a second remaining doll. She showed infants both dolls before placing them out of sight under the table. A 10-s measurement period began (four objects, more remaining). After 10 s, the experimenter said "What else was hiding in here?" and showed infants as she reached through the front of the box and "found" the remaining two dolls, showing them to infants briefly before placing them under the table. A final 10-s measurement period followed (four objects, none remaining) during which the box was again empty because all objects had been retrieved. After the measurement period, the experimenter said "Great job!" and pulled the box out of infants' reach.

Coding

Infants were defined as searching when one or both hands were inserted through the slit in the spandex up to or past the knuckle closest to the palm. Occasionally infants began actively searching inside the box but then became disengaged and left their hand inside without further moving it (e.g., infants left one hand in the box but looked down at their other hand while they played with the high chair). On these trials, infants were coded as searching the box until their hand visibly stopped moving inside and they had clearly disengaged with the task. An observer, blind to trial type,

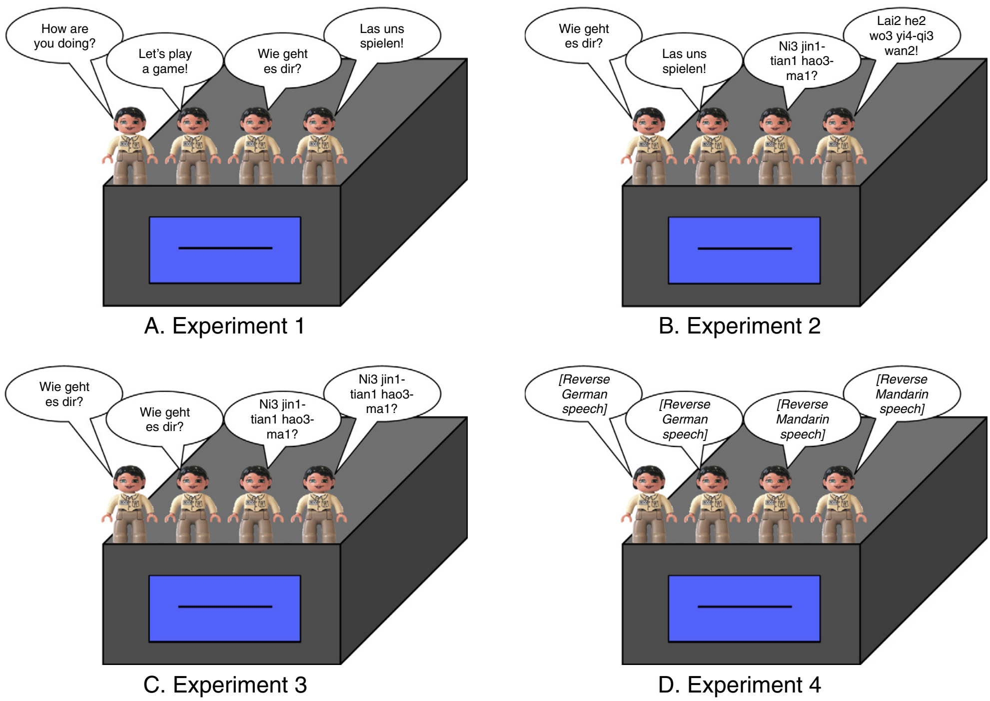

Fig. 2. Examples of four-object arrays and utterances in Experiments 1 through 4. Each doll was lifted in turn while the utterance played from underneath the table so as to appear that the doll was speaking. (A) In Experiment 1, two dolls each produced unique tokens of English and two dolls produced unique tokens of either German (pictured) or Mandarin. (B) In Experiment 2, two dolls each produced unique tokens of German and two dolls produced unique tokens of Mandarin. (C) In Experiment 3, two dolls each produced identical tokens of German and two dolls produced identical tokens of Mandarin. (D) In Experiment 4, two dolls each produced identical tokens of reversed German speech and two dolls produced identical tokens of reversed Mandarin speech.

coded infants' searching offline, frame by frame, using Preferential Looking Coder software (Libertus, 2011). A second observer recoded all trials from a randomly selected 25% of participants across all experiments, and interobserver agreement averaged .92.

Results

The 1 versus 2 comparison

We first compared infants' searching on the two measurement periods when no more objects remained in the box, and so searching was predicted to be minimal. We found no difference in searching on one-object (none remaining) measurement periods ($M = 1.37$, $SD = 0.93$) and two-objects (none remaining) measurement periods ($M = 1.28$, $SD = 0.63$), $t(15) = 0.37$, $p = .71$. Therefore, we averaged infants' searching on these two measurement periods ($M = 1.33$, $SD = 0.66$) as in previous studies using the manual search task (e.g., Feigenson & Carey, 2003, 2005; Feigenson & Halberda, 2004, 2008; Rosenberg & Feigenson, 2013; Stahl & Feigenson, 2014). We next compared this baseline searching with searching on the two-objects (more remaining) measurement periods, where two objects had been hidden and only one object retrieved ($M = 2.09$, $SD = 1.23$). We calculated a difference score by subtracting infants' searching on the averaged none remaining measurement periods from searching on the more remaining measurement periods. As predicted, this difference score differed significantly from chance, $t(15) = 2.62$, $p = .02$, Cohen's $d = 0.66$ (Fig. 3). Thus, infants successfully stored representations of one and two hidden objects in working memory. They persistently searched after seeing two objects hidden and only one object retrieved, and they appropriately decreased their searching after seeing either one object hidden and one retrieved or two objects hidden and both retrieved.

The 2 versus 4 comparison

We again first compared infants' searching on the two measurement periods when no more objects remained in the box, and so searching was predicted to be minimal. We found no difference in searching on two-objects (none remaining) measurement periods (M = 1.71, SD = 1.51) and four-objects (none remaining) measurement periods (M = 1.57, SD = 1.58), $t(15)$ = 0.28, p = .78. Therefore, we averaged infants' searching on these two measurement periods (M = 1.64, SD = 1.20) and subtracted this from infants' searching on the four-objects (more remaining) periods, where four dolls had been hidden and only two retrieved (M = 2.21, SD = 1.27). We found this difference score to differ significantly from zero, $t(15)$ = 2.44, p = .03, Cohen's d = 0.61 (Fig. 3). That is, infants successfully continued searching the box after seeing four dolls hidden and only two retrieved.

Next we compared infants' performance on the 1 versus 2 comparison with their performance on the 2 versus 4 comparison. As predicted, these difference scores did not differ significantly from each other, $t(15)$ = 0.67, p = .51. Therefore, infants were equally successful at representing four items in working memory as they were at representing fewer than four items.

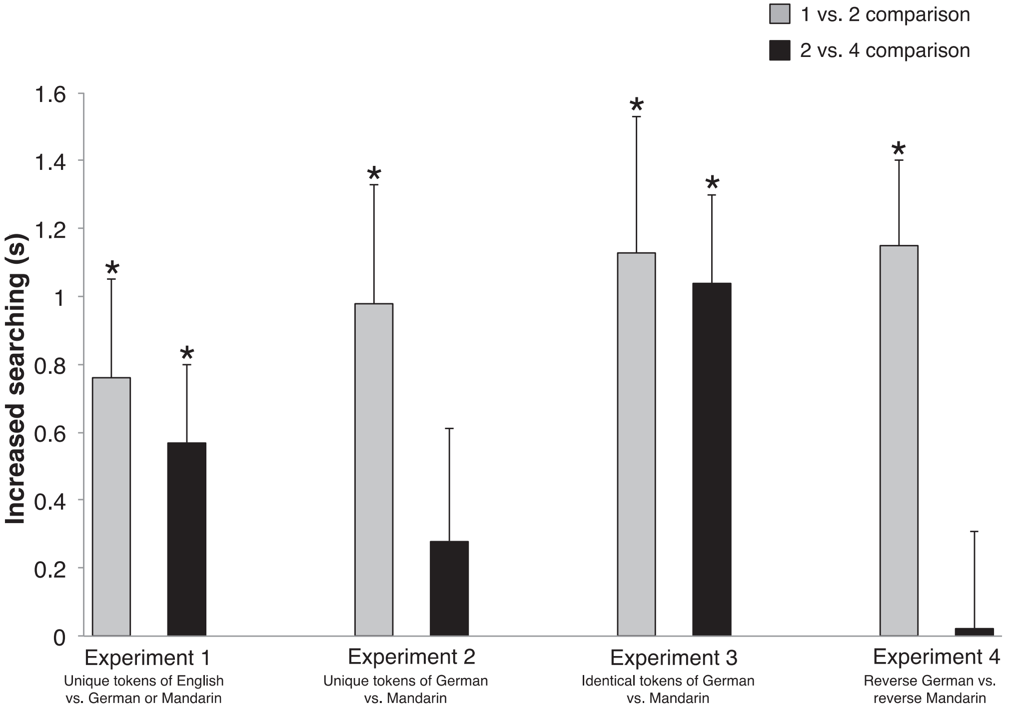

Fig. 3. Results from Experiments 1 through 4. Bars represent average difference scores (infants' searching when there were more items remaining in the box minus infants' searching when the box was empty). In the 1 versus 2 comparison, we compared infants' searching when two dolls were hidden and only one was doll retrieved with that when the box was empty (one doll was hidden and one retrieved and two dolls were hidden and two retrieved). In the 2 versus 4 comparison, we compared infants' searching when four dolls were hidden and two retrieved with that when the box was empty (two dolls were hidden and two retrieved and four dolls were hidden and four retrieved). Error bars represent standard errors of the mean. $^{*}p$ < .05.

Did infants' success at using language distinctions for chunking depend on having had exposure to more than one language? Although parents confirmed that none of the participants in Experiment 1 had exposure to either German or Mandarin, it is possible that *some* experience with two or more languages was required for infants' success regardless of the particular languages heard. Infants with multilingual exposure might be better at using language to group objects into sets given other social and cognitive benefits that bilingual infants demonstrate (e.g., Kovács & Mehler, 2009a, 2009b; Liberman, Woodward, Keysar, & Kinzler, 2016; Sebastián-Gallés, Alabareda-Castellot, Weikum, & Werker, 2012). To test this possibility, we compared the difference scores from the 2 versus 4 comparison of infants who were exposed only to English with those who were exposed to English plus at least one other language. We found no difference in the difference scores of infants from monolingual versus multilingual households, $t(14) = 0.01$, $p = .99$. Thus, infants used differences between languages as a chunking cue regardless of their own language exposure.

Discussion

In many previous studies, infants failed to concurrently represent four hidden objects (Barner et al., 2007; Feigenson & Carey, 2003, 2005; Feigenson & Halberda, 2004, 2008; Rosenberg & Feigenson, 2013; Stahl & Feigenson, 2014). Yet when provided with cues to parse a four-object array into sets using spatial separation between objects, perceptual similarity, shared kind membership, or statistical co-occurrences between objects, infants successfully represent the array as two sets of two (Feigenson & Halberda, 2004, 2008; Kibbe & Feigenson, 2016; Rosenberg & Feigenson, 2013). Here we extend these results by showing that infants can use the language an agent speaks as a basis for chunking. After seeing four identical dolls, two of whom "spoke" in English and two of whom "spoke" in an unfamiliar language, infants successfully represented all four objects, thereby overcoming the typical three-item limit in working memory. They did so despite the absence of any visual cues for chunking the array.

Experiment 2

The results of Experiment 1 show that infants can hierarchically reorganize working memory representations using linguistic group distinctions. When two dolls produced an utterance in English and two other dolls produced an utterance in a foreign language (German or Mandarin), infants represented all four objects. This raises the question of what language-based distinctions supported success in our task. One possibility is that infants represented one set of dolls as social agents who belonged to infants' own linguistic in-group (because English was the primary language spoken in our participants' homes) and represented the other set of dolls as social agents who belonged to an out-group. Another possibility, not mutually exclusive, is that infants grouped the dolls into sets based on the distinction between the two languages heard irrespective of whether one of those languages was the infants' own native language. If infants recognize that speaking the same language is an important socially relevant trait, this perceived similarity between two of the dolls (and dissimilarity to the other two dolls) may have motivated infants to represent them as two sets of two. To test these possibilities, in Experiment 2 we asked whether infants could use linguistic group distinctions for chunking even when both sets of speakers produced unfamiliar languages (i.e., were in two distinct linguistic out-groups).

Method

Participants

The participants were 16 healthy, full-term infants between 15 and 17 months of age (range = 15 months 4 days to 16 months 30 days, M = 16 months 10 days; 10 female). An additional 4 infants were excluded due to fussiness ($n = 3$) or refusal to search in the box on more than 50% of measurement periods ($n = 1$). In our final sample, 7 infants were reported to have had exposure to languages other than English more than 5% of the time from either one or both parents (1 infant each was exposed to

French, Dutch, and German, and 2 infants each were exposed to American Sign Language and Mandarin[2]).

Stimuli
The stimuli were the same as those in Experiment 1 except that only German and Mandarin were played.

Procedure
The procedure was identical to that in Experiment 1 with the exception of dolls producing only German and Mandarin utterances.

Each infant participated in a 1 versus 2 comparison block and a 2 versus 4 comparison block. The 1 versus 2 comparison block contained three measurement periods, each presented twice. On one-object (none remaining) measurement periods, infants saw a single doll produce an utterance in either German or Mandarin, saw the object hidden, retrieved the object, and then were able to search the box. On two-objects (more remaining) measurement periods (Fig. 1B), infants saw one doll produce an utterance in German and the other doll produce an utterance in Mandarin, saw both objects hidden, retrieved one of the objects, and then were able to search the box. On two-objects (none remaining) measurement periods, infants saw the missing object from the preceding measurement period retrieved and then were able to search the box.

The 2 versus 4 comparison block also contained three measurement periods, each presented twice. On two-objects (none remaining) measurement periods, infants saw one doll produce an utterance in German and the other doll produce an utterance in Mandarin, saw both objects hidden, retrieved both of the objects, and then were able to search the box. On four-objects (more remaining) measurement periods (Fig. 2B), infants saw two dolls each produce a distinct utterance in German ("Wie geht es dir?" and "Las uns spielen!") and saw the other two dolls each produce a distinct utterance in Mandarin ("Ni3 jin1-tian1 hao3-ma1?" and "Lai2 he2 wo3 yi4-qi3 wan2!"), saw all four objects hidden, retrieved two of the objects, and then were able to search the box. On four-objects (none remaining) measurement periods, infants saw the two missing objects from the preceding measurement period retrieved and then were able to search the box.

Whether infants participated in the 1 versus 2 comparison block first or second, the number of objects presented first within a block, and the order of languages uttered by the dolls within a block all were counterbalanced across infants.

Results

One infant reached for 3 standard deviations longer than the mean on a one-object (none remaining) trial (this was the only infant to do so across all experiments), and this reach was excluded from the analyses. In addition, due to experimenter error, 1 infant did not receive a two-objects (none remaining) measurement period in the 1 versus 2 comparison block (the last measurement of that infant's testing session).

The 1 versus 2 comparison
We first compared infants' searching on trials when no more objects remained in the box. We found no difference in infants' searching on one-object (none remaining) measurement periods ($M = 1.42$, $SD = 1.24$) and two-objects (none remaining) measurement periods ($M = 1.12$, $SD = 0.82$), $t(15) = 1.11$, $p = .28$, and so we averaged infants' searching on these two types of none remaining periods ($M = 1.27$, $SD = 0.89$). We compared this with infants' searching in the two-objects (more remaining) measurement periods in which two objects were hidden and only one was retrieved ($M = 2.25$, $SD = 1.40$). As in Experiment 1, we calculated a difference score by subtracting infants' search-

[2] Three infants were reported to have had some exposure to German or Mandarin, the languages we presented to infants that were intended to be "unfamiliar." The results remain unchanged with these infants excluded from the statistical analyses; when these infants are excluded, infants still succeeded in remembering one and two dolls but failed to remember four dolls (see Results).

ing on none remaining measurement periods from searching on more remaining measurement periods. As predicted, this difference score differed significantly from chance, $t(15) = 2.84$, $p = .01$, Cohen's $d = 0.71$ (Fig. 3).

The 2 versus 4 comparison

As in the 1 versus 2 comparison, we found no difference in searching on two-objects (none remaining) measurement periods ($M = 1.25$, $SD = 0.97$) and four-objects (none remaining) measurement periods ($M = 1.55$, $SD = 1.77$), $t(15) = -0.94$, $p = .36$. Therefore, we averaged infants' searching on these two measurement periods ($M = 1.40$, $SD = 1.27$) and subtracted this from infants' searching on the four-objects (more remaining) measurement periods in which four dolls had been hidden and only two retrieved ($M = 1.68$, $SD = 1.06$). We found that this difference score did not differ significantly from zero, $t(15) = 0.87$, $p = .40$, Cohen's $d = 0.22$ (Fig. 3). That is, infants did not continue searching when four dolls had been hidden and only two retrieved.

Next we compared the difference scores from the 2 versus 4 comparison of infants who had been exposed only to English with those of infants who had been exposed to English plus at least one other language. We found no difference in the performance of these two groups, $t(14) = -0.78$, $p = .45$.

Lastly, we compared infants' performance in the 1 versus 2 comparison with that in the 2 versus 4 comparison. We found that these difference scores did not differ significantly from one another, $t(15) = 1.42$, $p = .18$. In addition, infants' difference scores from the 2 versus 4 comparison in Experiment 2 did not differ significantly from those of infants in Experiment 1 (in which two dolls spoke English and two spoke a foreign language), $t(30) = 0.71$, $p = .48$.

Discussion

Whereas infants in Experiment 1 successfully used the contrast between their native language and an unfamiliar language to represent a four-object array, infants in Experiment 2 failed to do so when the contrast involved two unfamiliar languages. Infants who saw four identical dolls hidden in a box, two that spoke one unfamiliar language (German) and two that spoke another unfamiliar language (Mandarin), failed to continue searching for the missing dolls after only two of the four had been retrieved. This suggests that infants failed to represent the four-object array, and so they were unable to use the language cues provided as a chunking cue to aid memory. Importantly, hearing two unfamiliar languages was not in itself disruptive to infants' engagement in our task. Infants succeeded in representing objects that each produced a different unfamiliar language as long as the arrays were within infants' working memory limits; hence, infants' failure in the 2 versus 4 comparison block reflects a failure to use language as a chunking cue rather than a more general failure caused by the increased novelty of the auditory stimuli.

A possible interpretation of Experiments 1 and 2 is that infants can chunk only using linguistic group distinctions that involve representing a linguistic in-group and a linguistic out-group. On this account, infants from English-speaking households can recognize the English utterances as similar to each other, but distinct from the non-English utterances, as the basis for chunking an array. But another possibility is that infants can in fact chunk using the contrast between two non-native languages if the task is made easier. Experiment 2 was challenging for infants because it required them to abstract over two unique utterances from an unfamiliar language and to recognize these as being tokens of the same language. For example, when infants heard "Wie geht es dir?" and "Las uns spielen!" they needed to recognize these as utterances of the same language and then use this knowledge to chunk the two speakers. Although infants did not statistically succeed at doing this in the 2 versus 4 comparison block of Experiment 2, their performance with four-object arrays also was not statistically different from their performance with two-object arrays; similarly, the 2 versus 4 performance of infants in Experiment 2 did not statistically differ from that of infants in Experiment 1. This hints that infants may have had some ability to use the distinction between German and Mandarin to form chunks in working memory, even when the utterances contained different token sentences, but that their success was not robust. In Experiment 3, we asked whether infants would succeed with an easier version of the task, namely chunking objects into sets using the distinction between two unfamiliar languages when each language was presented using a single token utterance.

Experiment 3

The design of Experiment 3 was identical to that of Experiment 2 except that, for four-object arrays, two dolls produced identical tokens of a German utterance and two dolls produced identical tokens of a Mandarin utterance.

Method

Participants

The participants were 16 healthy, full-term infants between 15 and 17 months of age (range = 15 months 9 days to 16 months 26 days, M = 16 months 0 days; 7 female). An additional 9 infants were excluded because of parental interference (n = 1), refusing to search in the box on more than 50% of measurement periods (n = 7), or refusing to retrieve objects from the box throughout the experiment (n = 1). In our final sample, 4 infants were exposed to languages other than English more than 5% of the time (2 were exposed to Spanish and 1 infant each was exposed to Arabic and Russian).

Stimuli

The stimuli were the same as those in Experiment 2.

Procedure

The procedure was identical to that in Experiment 2 with one exception: When four dolls were presented, two dolls produced identical tokens of German (e.g., "Wie geht es dir?") and the other two dolls produced identical tokens of Mandarin (e.g., "Ni3 jin1-tian1 hao3-ma1?"). Each infant participated in a 1 versus 2 comparison block composed of a one-object (none remaining) measurement period, a two-objects (more remaining) measurement period (Fig. 1C), and a two-objects (none remaining) measurement period (each presented twice), and each infant participated in a 2 versus 4 comparison block composed of a two-objects (none remaining) measurement period, a four-objects (more remaining) measurement period (Fig. 2C), and a four-objects (none remaining) measurement period (each presented twice).

Results

The 1 versus 2 comparison

We first compared infants' searching on trials when no more objects remained in the box. We found no difference in infants' searching on one-object (none remaining) measurement periods (M = 1.11, SD = 1.08) and two-objects (none remaining) measurement periods (M = 1.55, SD = 1.36), $t(15) = -1.62$, p = .13, and so we averaged infants' searching on these two types of none remaining periods (M = 1.33, SD = 1.10). We compared this with infants' searching in the two-objects (more remaining) measurement periods in which two objects were hidden and only one object retrieved (M = 2.47, SD = 1.58). As in Experiments 1 and 2, we calculated a difference score by subtracting infants' searching on none remaining periods from searching on more remaining measurement periods. As predicted, this difference score differed significantly from chance, $t(15) = 2.83$, p = .01, Cohen's d = 0.71 (Fig. 3).

The 2 versus 4 comparison

As in the 1 versus 2 comparison, we found no difference in searching on two-objects (none remaining) measurement periods (M = 1.43, SD = 1.50) and four-objects (none remaining) measurement periods (M = 1.48, SD = 1.65), $t(15) = -0.23$, p = .82. Therefore, we averaged infants' searching on these two measurement periods (M = 1.46, SD = 1.49) and subtracted this from infants' searching on the four-objects (more remaining) measurement periods in which four dolls had been hidden but only two retrieved (M = 2.50, SD = 2.14). We found that this difference score differed significantly from zero, $t(15) = 3.95$, p = .001, Cohen's d = 0.99 (Fig. 3). That is, infants kept searching when four dolls had been hidden and only two retrieved.

Next, we compared the difference scores from the 2 versus 4 comparison of infants who had been exposed only to English with those of infants who had been exposed to English plus at least one other language. We found no difference between these two groups, $t(14) = 1.22$, $p = .24$. Because our sample of multilingual infants was small in each experiment, we next pooled all of the multilingual infants in Experiments 1 through 3 to compare their performance with that of monolingual infants. We found that, just as within each experiment, multilingual infants did not perform any differently in the 2 versus 4 comparison than monolingual infants, $t(46) = 0.34$, $p = .73$. A Fisher's exact test confirmed that multilingual infants were not more likely to search longer on more remaining measurement periods than on none remaining measurement periods in the 2 versus 4 comparison relative to monolingual infants ($p = 1.00$), indicating that multilingual exposure yielded no benefit in this task.

Lastly, we compared infants' searching behavior in the 1 versus 2 comparison with their searching behavior in the 2 versus 4 comparison. Given infants' success in each comparison block, we found that, as predicted, their difference scores for each did not differ significantly from one another, $t(15) = 0.21$, $p = .84$. We also compared infants' performance in the 2 versus 4 comparison of the current experiment (Experiment 3) with that of Experiment 2 (when two dolls produced unique tokens of German and two dolls produced unique tokens of Mandarin). This revealed a trend toward infants in Experiment 3 showing greater success than infants in Experiment 2, $t(30) = 1.81$, $p = .08$.

Discussion

When presented with four identical dolls, two of which produced identical utterances in one unfamiliar language and two of which produced identical utterances in a different unfamiliar language, infants successfully represented the array—that is, infants continued searching after four objects had been hidden and only two retrieved. Given that in the absence of chunking cues infants fail to represent four-object arrays, the results of Experiment 3 suggest that infants used the difference between the German and Mandarin utterances to chunk the array into two sets of two. Although they were unable to do this when the two speakers of each language produced a different token utterance (Experiment 2), infants' success when the task was made easier suggests that infants can use language as a basis for memory reorganization even when no in-group/out-group language distinctions are available.

Experiment 4

Our findings from Experiment 3 show that infants successfully chunked social objects into two sets of two if pairs of objects produced identical tokens of a foreign language. These results suggest that infants inferred that social agents who produced the same language shared an important trait, even when those languages were unfamiliar languages, and used this inference to chunk those agents into sets to overcome working memory limits. However, an alternative explanation is that infants' chunking was supported by low-level acoustical cues, not abstract social knowledge. On this account, infants detected that the sound properties associated with two of the objects were identical, and distinct from the identical sound properties associated with the other two objects, without representing the dolls' utterances as language related or socially relevant. To test this possibility, in Experiment 4 we presented infants with four dolls that produced the same speech as in Experiment 3—two identical tokens of German and two identical tokens of Mandarin—but played in reverse. Reverse speech is an ideal perceptual control because all of the low-level perceptual features that distinguished German from Mandarin utterances in Experiment 3 are preserved. However, many studies have shown that young children process reverse speech differently than forward speech such that it is no longer perceived as containing linguistically relevant information (e.g., Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002; Kinzler et al., 2007; Mehler et al., 1988; Ramus, Hauser, Miller, Morris, & Mehler, 2000). If low-level acoustical cues drove infants' chunking, infants in Experiment 4 should again succeed at representing all four dolls. However, if infants are using the social cue of language to chunk the dolls, infants should fail to remember all four dolls when those dolls produce reverse speech.

Method

Participants

The participants were 16 healthy, full-term infants between 15 and 17 months of age (range = 15 months 9 days to 16 months 26 days, M = 15 months 28 days; 8 female). An additional 7 infants were excluded for refusing to search in the box on more than 50% of measurement periods (n = 5) or equipment failure (n = 2). In our final sample, 4 infants were exposed to a language other than English more than 5% of the time by one or both parents (all 4 infants were exposed to Spanish).

Stimuli

The stimuli were the same as those in Experiment 3 except that the speech utterances were played in reverse.

Procedure

The procedure was identical to that in Experiment 3 with the exception of reverse speech being played on all trials. On one-object trials, infants heard one of the utterances in either reverse German or reverse Mandarin. On two-object trials (Fig. 1D), one doll produced reverse German and the other doll produced reverse Mandarin. When four dolls were presented, two dolls produced identical reverse tokens of German and the other two dolls produced identical reverse tokens of Mandarin (Fig. 2D).

Results

Due to equipment failure, 1 infant did not receive half of the trials in the 2 versus 4 comparison: a two-objects (none remaining) trial, a four-objects (more remaining) trial, and a four-objects (none remaining) trial. This infant still received one trial of each type in this block of trials and so was included in subsequent analyses.

The 1 versus 2 comparison

We first compared infants' searching on trials when no more objects remained in the box. We found no difference in infants' searching on one-object (none remaining) measurement periods (M = 1.05, SD = 1.21) and two-objects (none remaining) measurement periods (M = 0.87, SD = 0.96), $t(15) = 0.53$, $p = .60$, and so we averaged infants' searching on these two types of none remaining periods (M = 0.96, SD = 0.85). We compared this with infants' searching in the two-objects (more remaining) measurement periods in which two objects were hidden and only one object was retrieved (M = 2.11, SD = 1.42). As in Experiments 1 through 3, we calculated a difference score by subtracting infants' searching on none remaining measurement periods from searching on more remaining measurement periods. As predicted, this difference score differed significantly from chance, $t(15) = 4.64$, $p < .001$, Cohen's d = 1.16 (Fig. 3). Critically, these results demonstrate that infants were attentive during the experiment and were not distracted by the reverse speech.

The 2 versus 4 comparison

As in the 1 versus 2 comparison, we found no difference in searching on two-objects (none remaining) measurement periods (M = 1.30, SD = 1.21) and four-objects (none remaining) measurement periods (M = 1.96, SD = 1.50), $t(15) = -1.80$, $p = .09$. Therefore, we averaged infants' searching on these two measurement periods (M = 1.63, SD = 1.15) and subtracted this from infants' searching on the four-objects (more remaining) measurement periods in which four dolls had been hidden but only two retrieved (M = 1.65, SD = 1.25). We found that this difference score did not differ significantly from zero, $t(15) = 0.05$, $p = .96$, Cohen's d = 0.02 (Fig. 3). That is, infants did not continue searching when four dolls had been hidden and only two retrieved.

Next we compared the difference scores from the 2 versus 4 comparison of infants who had been exposed only to English with those from infants who had been exposed to English plus at least one other language. We found no difference in the performance of these two groups, $t(14) = 0.58$, $p = .58$.

Lastly, we compared infants' performance in the 1 versus 2 comparison and the 2 versus 4 comparison. As predicted, infants' difference scores on the 1 versus 2 comparison were significantly greater than those on the 2 versus 4 comparison, $t(15) = 3.75$, $p = .002$. In addition, as predicted, infants in Experiment 3 (where dolls produced human speech) outperformed infants in the current experiment (where dolls produced reversed speech) on the 2 versus 4 comparison, $t(30) = 2.60$, $p = .01$.

Discussion

Infants did not chunk objects when presented with four identical dolls, two of which produced identical tokens of backward German speech and two of which produced identical tokens of backward Mandarin speech. This suggests that the success observed in Experiment 3, when the same utterances were played in typical forward speech, hinged on infants' recognition of the utterances as instances of two different languages. That infants in Experiment 4 successfully represented one- and two-object arrays of dolls that produced backward speech shows that hearing reversed speech did not, in and of itself, disrupt infants' performance. Rather, reversed speech did not provide a foundation for chunking.

General discussion

In four experiments, we investigated infants' ability to harness their knowledge about language and social agents to surpass their typical three-item limit in working memory. In Experiment 1, infants successfully represented four dolls when two of the dolls produced unique tokens of English and two of the dolls produced unique tokens of either German or Mandarin. In Experiment 2, we asked whether infants could use language as a chunking cue when all of the dolls produced an unfamiliar language. We found that when two dolls produced unique tokens of German and two dolls produced unique tokens of Mandarin, infants failed to chunk the dolls into sets. Experiment 3 probed whether this failure may have been due to the challenge of abstracting across unique tokens of an unfamiliar foreign language. When this challenge was eliminated by having two dolls produce identical tokens of German and two dolls produce identical tokens of Mandarin, infants succeeded at representing all four dolls. Hence, when the task of identifying tokens of an unfamiliar language was made easier, infants successfully used the distinction between two unfamiliar languages to chunk a larger array into sets. Finally, in Experiment 4, we ruled out an alternative explanation for our results, showing that infants failed to remember all four dolls when two produced identical tokens of reverse German speech and two produced identical tokens of reverse Mandarin speech. Therefore, infants were not chunking based on low-level acoustical cues but rather clunking based on their inferences about social agents who speak different languages.

Together, these results suggest that infants can use distinctions among languages as a basis for forming meaningful chunks, remembering individual items when those items are distinguished by linguistic group membership—even when infants are unfamiliar with the languages denoting that membership. This finding adds to the growing body of evidence that language is a meaningful marker of social groups even well before infants become competent speakers themselves. Not only do infants use language to determine with whom to interact and from whom to learn (e.g., Buttelmann et al., 2013; Kinzler et al., 2007), but they also use language cues to efficiently restructure their memory representations. Other studies have shown that social information can bias what infants store in memory (Kibbe & Leslie, in press; Southgate, Csibra, Kaufman, & Johnson, 2008; Yoon, Johnson, & Csibra, 2008) and how they store that information (Stahl & Feigenson, 2014). The current experiments continue to highlight how social knowledge and memory interact.

This work raises questions for future inquiry. For example, we found that infants successfully chunked representations of social agents into sets based on the language the agents produced. However, it is unknown whether social agents per se must produce the language or whether any object associated with the speech tokens would suffice. That is, if infants were presented with four nonsocial objects (e.g., blocks) that were each lifted as each speech token played (in the same manner as the current experiments), could infants still use the distinct languages to chunk the objects into sets? It might be that merely hearing instances of two distinct languages is enough to induce chunking and remember four items in working memory. Alternatively, because language is inherently social, those language tokens might need to be produced by and mapped to particular social agents to be a meaningful chunking cue.

A second question concerns the subtlety of the distinctions to which infants are sensitive. Here we found that infants use the contrasts between two quite different sounding languages (German and Mandarin, or English and Mandarin, or English and German) to motivate chunking an array into sets. In discrimination preference studies, younger infants also distinguish between languages that are more similar (such as two rhythmically similar Romance languages, Spanish and Catalan; Bosch & Sebastián-Gallés, 2001). Furthermore, 5- and 6-month-olds distinguish between native speakers of their own language and speakers who produce accented speech (preferring to look at speakers with a native accent) (Kinzler et al., 2007). It is an open question whether this early sensitivity to subtle language distinctions also supports categorizing the world into social groups.

In addition, it remains unknown what other types of social cues infants can use to chunk memory representations. It is possible that any social trait, even a seemingly superficial one, might suffice to induce chunking. For example, perhaps infants in Experiments 1 and 3 were chunking the dolls within a pair not based on the fact that they produced the same language but rather based on the fact that they shared *something* in common (e.g., infants in Experiment 1 recognized that two English-speaking dolls shared something in common that was different from the other two dolls who did not speak English; infants in Experiment 3 recognized that dolls within a pair merely produced the same utterance). If so, then infants might also succeed at chunking four identical English-speaking dolls into sets based on the particular English utterance produced (e.g., two dolls say "How are you doing?" and two dolls say "Let's play a game!"). Alternatively, infants might ignore superficial distinctions like these and chunk only using distinctions more likely to signal a deeper, more permanent similarity among individuals (including, in addition to the language one speaks, the preferences one exhibits and the place where one lives).

Interestingly, whereas social preferences concerning language emerge early, this is not true of race; infants are equally likely to interact with own-race and other-race individuals (Kinzler & Spelke, 2011). Thus, infants do not appear predisposed to form social groups based on just any available social category marker. Rather, they might be specifically attuned to linguistic differences between individuals from very early in life, whereas sensitivity to other types of social distinctions may develop only with experience. Although race-based preferences do emerge later in childhood (e.g., Kinzler & Spelke, 2011), these older children still privilege language over race in their social decisions. For example, children indicate a desire to be friends with a native speaker of a different race over a foreign speaker of their own race (Kinzler, Shutts, DeJesus, & Spelke, 2009). Young children also incorrectly believe that language—not race—is a stable trait; they find it more plausible that a White English-speaking child will grow up to be a Black English-speaking adult than a White French-speaking adult (Kinzler & Dautel, 2012).

Although infants do not show social preferences based on race, they do exhibit early emerging social preferences based on evaluations of social agents' behavior (e.g., Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2007; Hamlin, Wynn, Bloom, & Mahajan, 2011), intentions (Dunfield & Kuhlmeier, 2010; Hamlin, 2013; Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013), and preferences (Hamlin, Mahajan, Liberman, & Wynn, 2013; Mahajan & Wynn, 2012). Thus, it would be interesting to ask whether such socially relevant traits can influence memory organization in the same way as language.

Furthermore, it would be interesting to investigate precisely what infants remember about the items that are chunked based on social information. Typically as the number of items to remember increases, the resolution with which infants remember those items decreases (Kibbe, 2015; Kibbe & Feigenson, 2016; Kibbe & Leslie, 2013; Zosh & Feigenson, 2012). Recent evidence shows that infants can maintain conceptual information about an item in working memory even when they lose perceptual information (Kibbe & Leslie, in press; see also Southgate et al., 2008). It is unknown whether infants better remember items that were chunked based on social information, and language in particular, than items chunked based on nonsocial information. Similarly, it is an open question whether infants would be more likely to remember details about agents who produce their native language over a foreign language, given their preference for native speakers (e.g., Kinzler et al., 2007), or whether they would remember precisely how many dolls had produced each language on retrieval. Consideration of how social context influences memory storage and resolution can inform research on both social cognition and working memory.

In sum, the current experiments underscore the intersection between social knowledge and memory during infancy. Despite their limited working memory capacity, infants can use their expectations about language and social agents to overcome typical limits on working memory; from early on, they carve up the world based on the language speakers produce. Infants are remarkably sensitive to language from birth onward (e.g., Mehler et al., 1988), showing strong preferences for native speakers and making sophisticated inferences about social groups based on linguistic group distinctions (e.g., Buttelmann et al., 2013; Kinzler et al., 2007; Liberman et al., 2016). Our results demonstrate yet another way in which infants can harness these sensitivities, here to benefit working memory.

Acknowledgments

We thank Stephanie Caronna for assistance with data collection and thank Jingjing (Jenny) Wang for recording the auditory stimuli.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*, 106–111.
Barner, D., Thalwitz, D., Wood, J., & Carey, S. (2007). On the relation between the acquisition of singular–plural morpho-syntax and the conceptual distinction between one and more than one. *Developmental Science, 10*, 365–373.
Begus, K., Gliga, T., & Southgate, V. (2016). Infants' preferences for native speakers are associated with an expectation of information. *Proceedings of the National Academy of Sciences of the United States of America, 113*, 12397–12402.
Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy, 2*, 29–49.
Bower, G. H. (1972). Perceptual groups as coding units in immediate memory. *Psychonomic Science, 27*, 217–219.
Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138*, 487–502.
Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11*. https://doi.org/10.1167/11.5.4.
Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. *Child Development, 84*, 422–428.
Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science, 21*, 343–348.
Chase, W. G., & Ericcson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.). *The psychology of learning and motivation* (Vol. 16, pp. 1–58). New York: Academic Press.
Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–81.
Cowan, N. (2001). The magical number 4 in short term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.
Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science, 298*, 2013–2015.
Dunfield, K. A., & Kuhlmeier, V. A. (2010). Intention-mediated selective helping in infancy. *Psychological Science, 21*, 523–527.
Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science, 208*, 1181–1182.
Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science, 6*, 568–584.
Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*, 295–313.
Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science, 13*, 150–156.
Feigenson, L., & Halberda, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition, 91*, 173–190.
Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 9926–9930.
Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four ... or is it two? *Memory, 12*, 732–747.
Hamlin, J. K. (2013). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition, 128*, 451–474.
Hamlin, J. K., Mahajan, N., Liberman, Z., & Wynn, K. (2013). Not like me = bad: Infants prefer those who harm dissimilar others. *Psychological Science, 24*, 589–594.
Hamlin, J. K., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science, 16*, 209–226.
Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development, 26*, 30–39.
Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450*, 557–559.
Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 19931–19936.

Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *The Quarterly Journal of Experimental Psychology A, 49*, 116–139.
Howard, L. H., Henderson, A. M. E., Carrazza, C., & Woodward, A. L. (2015). Infants' and young children's imitation of linguistic in-group and out-group informants. *Child Development, 86*, 259–275.
Kibbe, M. M. (2015). Varieties of visual working memory representation in infancy and beyond. *Current Directions in Psychological Science, 24*, 433–439.
Kibbe, M. M., & Leslie, A. M. (in press). Conceptually rich, perceptually sparse: Object representations in six-month-old infants' working memory. *Psychological Science*.
Kibbe, M. M., & Feigenson, L. (2014). Developmental origins of recoding and decoding in memory. *Cognitive Psychology, 75*, 55–79.
Kibbe, M. M., & Feigenson, L. (2016). Infants use temporal regularities to chunk objects in memory. *Cognition, 146*, 251–263.
Kibbe, M. M., & Leslie, A. M. (2013). What's the object of object working memory in infancy? Unraveling "what" and "how many". *Cognitive Psychology, 66*, 380–404.
Kinzler, K. D., & Dautel, J. (2012). Children's essentialist reasoning about language and race. *Developmental Science, 15*, 131–138.
Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 12577–12580.
Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2012). "Native" objects and collaborators: Infants' object choices and acts of giving reflect favor for native over foreign speakers. *Journal of Cognition and Development, 13*, 67–81.
Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition, 27*, 623–634.
Kinzler, K. D., & Spelke, E. S. (2011). Do infants show social preferences for people differing in race? *Cognition, 119*, 1–9.
Kovács, A. M., & Mehler, J. (2009a). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 6556–6560.
Kovács, A. M., & Mehler, J. (2009b). Flexible learning of multiple speech structures in bilingual infants. *Science, 325*, 611–612.
Liberman, Z., Woodward, A. L., Keysar, B., & Kinzler, K. D. (2016). Exposure to multiple languages enhances communication skills in infancy. *Developmental Science*. Advance online publication. http://doi.org/10.1111/desc.12420.
Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2016). Preverbal infants infer third-party social structure based on linguistic group. *Cognitive Science*. Advance online publication. http://doi.org/10.1111/cogs.12403.
Liberman, Z., Woodward, A. L., Sullivan, K., & Kinzler, K. D. (2016). Early emerging system or reasoning about the social nature of food. *Proceedings of the National Academy of Sciences of the United States of America, 113*, 9480–9485.
Libertus, K. (2011). *Preferential looking coder* (Version 1.3.3) [computer software]. Baltimore, MD: Johns Hopkins University.
Mahajan, N., & Wynn, K. (2012). Origins of "us" versus "them": Prelinguistic infants prefer similar others. *Cognition, 124*, 227–233.
Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition, 122*, 346–362.
Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143–178.
Miller, G. M. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.
Moher, M., Tuerk, A. S., & Feigenson, L. (2012). Seven-month-old infants chunk items in working memory. *Journal of Experimental Child Psychology, 112*, 361–377.
Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development, 16*, 495–500.
Oakes, L. M., Hurley, K., Ross-Sheehy, S., & Luck, S. J. (2011). Developmental changes in infants' visual short-term memory for location. *Cognition, 118*, 293–305.
Ramus, F., Hauser, M., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*, 349–351.
Rosenberg, R. D., & Feigenson, L. (2013). Infants hierarchically organize memory representations. *Developmental Science, 16*, 610–621.
Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development, 74*, 1807–1822.
Sebastián-Gallés, N., Alabareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science, 23*, 994–999.
Shutts, K., Kinzler, K. D., McKee, C., & Spelke, E. S. (2009). Social information guides infants' selection of foods. *Journal of Cognition and Development, 10*, 1–17.
Simon, H. (1974). How big is a chunk? *Science, 183*, 482–488.
Soley, G., & Sebastián-Gallés, N. (2015). Infants prefer tunes previously introduced by speakers of their native language. *Child Development, 86*, 1685–1692.
Southgate, V., Csibra, G., Kaufman, J., & Johnson, M. H. (2008). Distinct processing of objects and faces in the infant brain. *Journal of Cognitive Neuroscience, 20*, 741–749.
Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*, 1–29.
Stahl, A. E., & Feigenson, L. (2014). Social knowledge facilitates chunking in infancy. *Child Development, 85*, 1477–1490.
Van de Walle, G. A., Carey, S., & Prevor, M. (2000). Bases for object individuation in infancy: Evidence from manual search. *Journal of Cognition and Development, 1*, 249–280.
Yoon, J. M., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 13690–13695.
Zosh, J. M., & Feigenson, L. (2012). Memory load affects object individuation in 18-month-old infants. *Journal of Experimental Child Psychology, 113*, 322–336.
Zosh, J. M., Halberda, J., & Feigenson, L. (2011). Memory for multiple visual ensembles in infancy. *Journal of Experimental Psychology: General, 140*, 141–158.