COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 34 (2010) 719–751

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2010.01109.x

A One-to-One Bias and Fast Mapping Support Preschoolers' Learning About Faces and Voices

Mariko Moher, Lisa Feigenson, Justin Halberda

Department of Psychological and Brain Sciences, Johns Hopkins University

Received 5 June 2009; received in revised form 23 January 2010; accepted 25 January 2010

Abstract

A multimodal person representation contains information about what a person looks like and what a person sounds like. However, little is known about how children form these face-voice mappings. Here, we explored the possibility that two cognitive tools that guide word learning, a one-to-one mapping bias and fast mapping, also guide children's learning about faces and voices. We taught 4- and 5-year-olds mappings between three individual faces and voices, then presented them with new faces and voices. In Experiment 1, we found that children rapidly learned face-voice mappings from just a few exposures, and furthermore spontaneously mapped novel faces to novel voices using a one-to-one mapping bias (that each face can produce only one voice). In Experiment 2, we found that children's face-voice representations are abstract, generalizing to novel tokens of a person. In Experiment 3, we found that children retained in memory the face-voice mappings that they had generated via inference (i.e., they showed evidence of fast mapping), and used these newly formed representations to generate further mappings between new faces and voices. These findings suggest that preschoolers' rapid learning about faces and voices may be aided by biases that are similar to those that support word learning.

Keywords: Inference; Mutual exclusivity; Person identification; Social cognition; Fast mapping; Domain general; Faces; Voices; Disjunctive syllogism

1. Introduction

Daily social interactions present us with a constant flux of visual and auditory information. We catch a glimpse of a familiar face on the sidewalk; we hear a few words spoken by a familiar voice behind us. In order to support successful interaction with other social agents, these two perceptually dissimilar sources of information must be combined into unified representations of other people. This capacity to map information about what someone

Correspondence should be sent to Mariko Moher, Department of Psychological and Brain Sciences, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218. E-mail: mariko.moher@jhu.edu

looks like to information about what he or she sounds like has been studied in adults, animals, and infants, from both behavioral and neuropsychological perspectives (for review, see Campanella & Belin, 2007). This work has helped to elucidate the role of integrated face-voice representations in speech processing and affect perception, and it has begun to identify the neural substrates underlying these abilities.

The importance of person perception early in development (e.g., the need to recognize caregivers) suggests that face-voice processing is likely an early-developing skill. However, little is known about how children come to form mappings between faces and voices, and in particular, whether any cognitive constraints help guide the initial formation of these mappings. This stands in contrast to other domains in which integration of visual and auditory information has been explored in more depth. A prominent example is that of word learning, in which researchers have devoted considerable effort to understanding the constraints that influence children's early mappings between objects and labels (e.g., Carey, 1978; Golinkoff, Mervis, & Hirsh-Pasek, 1994; Halberda, 2006; Markman, 1992; Markman & Hutchinson, 1984; Mervis, 1987; Wilkinson & Mazzitelli, 2003). It has been suggested that such constraints narrow the range of possible mappings children will consider, enabling children to quickly and accurately determine which of many objects is the referent of a new word. For example, children have been suggested to rely on an assumption that each object has only one name (Clark, 1990; Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Markman & Wachtel, 1988; Merriman & Bowman, 1989). This one-to-one mapping bias empowers children to learn new words in ambiguous naming situations (Carey & Bartlett, 1978): When faced with a novel object for which they do not know a label and a familiar object with a known label, children assume that a novel label refers to the novel object. In addition, children learning new words have been shown to rapidly form a new object-word pairing after just a single exposure to a novel word and to retain this mapping in memory—an ability called ''fast mapping'' (Carey & Bartlett, 1978; Heibeck & Markman, 1987; Jaswal & Markman, 2003). Fast mapped pairings between words and objects are robust enough to themselves support further learning of new words (Golinkoff et al., 1992).

In some ways, the challenges that children face in the domain of word learning parallel the challenges they face in the domain of social cognition. In both domains, children must form mappings between information from different sensory modalities (in word learning: between objects and words; in social cognition: between faces and voices). In both domains, children are often faced with many possible candidates for each pairing (e.g., they see many objects when hearing a new word, or see many faces when hearing a new voice). And in both domains, children must acquire new mappings in the absence of explicit, ostensive input aimed at teaching them to pair a particular sound with a particular image. Although we do not argue for a direct analogy between face-voice mapping and object-word mapping (e.g., the symbolic reference involved in word-object mapping does not apply in the case of faces and voices), these similarities raise the possibility that some of the constraints that empower children's rapid word learning have analogs that support the rapid learning of integrated face-voice representations.

Previous research suggests that some characteristics of word learning do apply to children's learning in other domains. For example, children and adults can fast map facts to objects (e.g., ''This is the one her uncle gave her''; Markson & Bloom, 1997) and novel actions to objects (Childers & Tomasello, 2003). But it remains unknown whether children exhibit an analogous ability to fast map within the social domain, learning new face-voice pairs from a single exposure. Furthermore, because Markson & Bloom and Childers & Tomasello provided children with ostensive evidence of which object was the ''one my uncle gave me'' or the object with which to perform the specific action, children did not have to make an inference about which fact or action mapped to which object. It is therefore unknown whether a one-to-one assumption, combined with the ability to rapidly learn from ambiguous input, constrains children's mappings in problem spaces other than word learning. For this reason, investigating children's use of such constraints in learning face-voice pairings informs the role these constraints may play in a more domain general manner throughout cognition.

In the present work, we explored the hypothesis that children rely on a one-to-one bias and fast mapping to help solve the face-voice mapping problem, in a similar way to that in which they solve the word-object mapping problem.[1] We focused on three questions. First, we asked whether children form face-voice mappings rapidly, or whether children instead require large amounts of exposure to form integrated face-voice representations. Second, we asked whether the assumption of a one-to-one mapping between faces and voices helps guide children's learning. Third, we asked whether children retain newly formed face-voice representations in memory and use them as input for further inferences about other faces and voices. In approaching each of these questions, we relied on methods adapted from published experiments in word learning.

Before turning to our experiments, we begin by briefly reviewing previous findings on children's competence at person identification using visual, auditory, and integrated sources of information.

1.1. Visual person identification

Before asking when and how children integrate visual information about faces with auditory information about voices, it is important to consider how children represent each of these pieces of information individually. A precursor to visual person identification is the ability to discriminate between the faces of different individuals. This ability is in place early in development. Infants preferentially attend to faces over other visual stimuli by 2 months of age or earlier (Easterbrook, Kisilevsky, Hains, & Muir, 1999; Maurer & Barrera, 1981) and discriminate between two unfamiliar faces from as early as a few days old (Barrera & Maurer, 1981; Bahrick, Hernandez-Reif, & Flom, 2005; Bahrick, Lickliter, Vaillant, Shuman & Castellanos, 2004; Cohen & Strauss, 1979; Cornell, 1974; Fagan, 1976; Pascalis & de Schonen, 1994). Face discrimination abilities are subtly tuned during infancy. Although newborn infants can distinguish between faces that do not belong to their social in-group (i.e., between two faces from a race other than their parents' race, or between two faces from a non-human primate species), they become less adept at these discriminations over the first year of life as their discrimination sensitivity within their own in-group increases (Kelly et al., 2007; Pascalis et al., 2005).

Beyond the ability to make simple discriminations between faces, young infants also remember the faces of specific individuals. One-day-old infants increase their sucking rates in order to view a photograph of their mother's face rather than a stranger's face matched for hair and eye color, complexion, and hairstyle (Walton, Bower, & Bower, 1992; see also Pascalis, de Schonen, Morton, Deruelle, & Fabre-Grenet, 1995 for results from a preferential looking paradigm that support similar conclusions). Two- to four-day-old infants express a visual preference for their mother's face even after a 15-min delay from their last exposure to their mother, and this preference emerges after as little as 5–6 hours of exposure to the mother's face (Bushnell, 2001).

As young children become exposed to increasing numbers of familiar people their stored representations of individuals must expand to incorporate this widening social circle. Indeed, 3-year-old children shown photographs of their preschool classmates correctly identify approximately 70% of their classmates by name, and by 6 years they are 90% correct. Seven- to fourteen-year-olds are at ceiling, performing from 96% to 100% correct (Brooks & Goldstein, 1963; see also Diamond & Carey, 1977). Hence by the time they are of preschool age, children can visually identify multiple familiar individuals to whom they have had significant social exposure. In laboratory tasks, children can also recognize faces to which they have had much less exposure. This ability to recognize faces following short periods of exposure improves from 6 years of age through adolescence (Carey, Diamond, & Woods, 1980).

1.2. Auditory person identification

Like faces, voices offer a stable source of information for person identification. And as with faces, evidence suggests that infants can perform simple discriminations between two voices from an early age (DeCasper & Prescott, 1984; Floccia, Nazzi, & Bertoncini, 2000). For example, 2- and 3-month-old infants who were habituated to one woman's voice reading a nursery rhyme dishabituated to a different woman's voice reading the same nursery rhyme (Bahrick, Lickliter, Shuman, Batista, & Grandez, 2003; Bahrick et al., 2005). Children also can perform explicit discriminations between voices: 5-year-old children who heard two utterances were above chance at identifying whether the utterances were produced by the same speaker (Cleary, Pisoni, & Kirk, 2005).

Beyond mere discriminations, infants also perform person identification based on voice information. Infants as young as a few days old are more likely to suck in order to hear their mother's voice than to hear a female stranger's voice (DeCasper & Fifer, 1980; Mehler, Bertoncini, Barrière, & Jassik-Gerschenfeld, 1978; Mills & Melhuish, 1974). Some evidence suggests that this preference is in place even before birth (Kisilevsky et al., 2003). Furthermore, such preferences appear to play a role in learning. Seven-month-old infants are better able to learn words presented amidst noise when their mother is the speaker than when an unfamiliar woman is the speaker (Barker & Newman, 2004).

However, identifying individuals by voice appears to be more difficult than identifying by face (Bartholomeus, 1973). Whereas preschool-aged children are above chance at naming the familiar classmate whose verbal utterance they have just heard (Bartholomeus, 1973), they have a harder time identifying an unfamiliar person by voice. In a delayed match-to-sample task in which children heard an utterance spoken by an unfamiliar person and then had to identify which of several new utterances was produced by the same speaker, 6-year-olds performed near chance and did not rise above chance until 8–10 years old (Mann, Diamond, & Carey, 1979).

1.3. Multimodal integration of faces and voices

Although the evidence reviewed above suggests that infants and children do form long-term memories of particular faces and voices, it is less clear to what extent this reflects abstract person representation rather than maintenance of modality-specific information. One source of evidence that children's person representations are abstract would come from demonstrations that children bind visual and auditory information into a single, unified representation. For example, can children recall an image of a person's face while hearing that person's voice over the phone, or recall an auditory memory of a person's voice while seeing their photograph?

Such multimodal representations of individuals have been demonstrated in nonhuman species. In one experiment domesticated horses saw a target herd-mate walk past their stall and then immediately heard a recorded vocalization—the vocalization had either been produced by the target horse or by a different, unseen horse. Subjects were faster to turn their heads in the direction in which the target horse had disappeared, and remained oriented for longer, when the visual image of the passing horse was incongruent with the auditory vocalization (i.e., when Horse 1 walked by, but Horse 2's neigh was heard). This suggests that horses can detect the mismatch of visual and auditory identity (Proops, McComb, & Reby, 2009).

Human infants also demonstrate some multimodal representations of social agents. Three-and-a-half-month old infants who heard a parent's voice from a central location and saw their mother's and their father's faces oriented toward the face associated with the voice (Spelke & Owsley, 1979). This suggests that infants had access to long-term representations of each parent that included both visual and auditory information. Infants also can form multimodal representations of individuals whose faces and voices are learned in the laboratory (Bahrick et al., 2005; Brookes et al., 2001). Four- and six-month-old infants saw alternating videos of two unfamiliar women or two unfamiliar men reciting a nursery rhyme for up to 20 min. Following this learning phase, infants dishabituated to videos depicting novel face-voice pairings (Bahrick et al., 2005). These results suggest that over the course of the testing session infants formed representations linking a particular person's face and voice. Subsequent work points to the period from 2 to 4 months of age for the emergence of this ability to form these face-voice pairings (Bahrick et al., 2005). However, to date no research has asked about the processes that underlie the construction of these important multimodal representations.

1.4. The present experiments

The results from the word learning literature reviewed above motivate several questions about how children might form multimodal representations of people, and whether the process of forming these person representations might show similar characteristics to the process of word learning. First, children form mappings between objects and words rapidly and with only minimal input. Can children also rapidly learn mappings between faces and voices, or does this require more extensive exposure? To date, infants in laboratory tasks require up to 20 min of exposure in order to form successful mappings (Bahrick et al., 2005), and older children do not succeed at face-voice pairing tasks that require explicit judgments until 6 years of age (Mann et al., 1979). Both of these previous sources of evidence suggest that children cannot rapidly form face-voice mappings. Second, a one-to-one mapping bias leads young word learners to expect each object to have just one label (Markman & Wachtel, 1988). Is children's learning about face-voice pairings aided by a similar one-to-one mapping bias, such that children expect each person to have one unique face and one unique voice? Third, young word learners have been shown to retain new word-object mappings in memory and to use these to bootstrap further learning about new words (Golinkoff et al., 1992). Can children also store a new face-voice mapping in memory and use this as input for further learning about other faces and voices?

To address these three questions we relied on a methodology borrowed from the word learning literature. We focused on preschool-aged children because we anticipated that they would be able to learn multiple face-voice pairs over a single testing session (whereas infants might not), thereby allowing us to investigate what kinds of novel inferences these newly learned mappings would support. In Experiment 1, we addressed the first two questions by asking whether 4- to 5-year-old children could learn three face-voice pairings over the course of a brief training session. We then probed for the presence of a one-to-one mapping bias between faces and voices by presenting children with a recently learned face and a completely novel face, and playing a novel voice. We predicted that children would spontaneously rule out the recently learned face as the correct match to the novel voice, and would thereby generate a new mapping between the completely novel face and completely novel voice. In Experiment 2, we attempted to replicate the results of Experiment 1, and we also tested the abstractness of children's face-voice pairings by asking whether the face-voice pairings they had formed would support recognition of novel images of a learned person (i.e., multiple views of each face). Finally, in Experiment 3, we addressed the last question by testing fast mapping and the productive nature of children's ability to make inferences about new faces and voices. We predicted that, having just formed a representation of a new face-voice pairing from a single trial, children would retain this integrated representation and be able to use it to support further inferences about new faces and voices on a later trial.

2. Experiment 1

In our first experiment, we asked whether children can rapidly form new face-voice mappings, and whether they use a one-to-one mapping bias to guide their learning about new faces and voices.[2] We based our method on a procedure used to study children's fast mapping and one-to-one biases in word learning (e.g., Halberda, 2006). First, we presented children with a brief training phase in which we taught them mappings between three novel faces and three novel voices. During this phase children saw an Introduction Trial for each face-voice pair in which a single static face appeared onscreen and a single voice was played. Then children saw a series of trials in which two of the three learned faces appeared side by side onscreen and one voice played over an audio speaker. Children had to choose which face had ''spoken.'' These trials provided a measure of children's ability to rapidly learn new face-voice pairs. Then on critical test trials children were presented with a ''known face'' (i.e., one of the three learned faces for which they had learned the voice during the Training phase) and a completely novel face. A novel voice was then played. These trials provided a measure of children's ability to use a one-to-one mapping bias to learn about new faces and voices. If children maintain a one-to-one mapping bias for faces and voices, and if they can use this bias productively to learn new mappings, they should reject the known face as the possible speaker and instead choose the novel face.

If children succeed, this result would add to the body of literature suggesting that these particular learning constraints may not be specific to word learning (Childers & Tomasello, 2003; Markson & Bloom, 1997). As our methods are adapted from the literature on word learning, they present children with a single image of each face and a single spoken phrase from each voice over multiple trials. As such, this first foray into testing face-voice pairings does not present children with the full-blown discrimination problem of seeing dynamic moving faces and changing voices across multiple contexts. However, the reasoning involved, within the domains of individual faces and individual voices, require children to discriminate different individuals within two dimensions relevant to constructing multimodal person representations, though these representations will be richer and more complicated in ecologically natural settings.

2.1. Methods

2.1.1. Participants

Sixteen children (mean age: 4 years 8 months, range: 4 years 0 months to 5 years 6 months; 10 boys) were included in the final data analysis. All children were recruited from the Baltimore, MD, area and surrounding suburbs. Twelve additional children were excluded for reasons potentially related to the method. Two children demonstrated a side bias (defined as choosing the face on either the right or the left side of the screen on more than 75% of all trials), and 10 children persisted in a premature behavioral response unrelated to the actual trial (i.e., pointed at one of the two stimulus faces prior to hearing the voice on more than 50% of all trials).[3] Their behavior suggests that these children failed to take into account all of the relevant information before making their response, and thus their response fails to inform our hypothesis. Two additional children were excluded for parental interference (1) and for losing interest during the task (1).

2.1.2. Stimuli

Over the course of the experiment children saw 11 different 30×42 cm color photographs, each of a different woman's face (see Fig. 1A). Hands were also included in some of the images to highlight the difference between images and to maintain children's interest

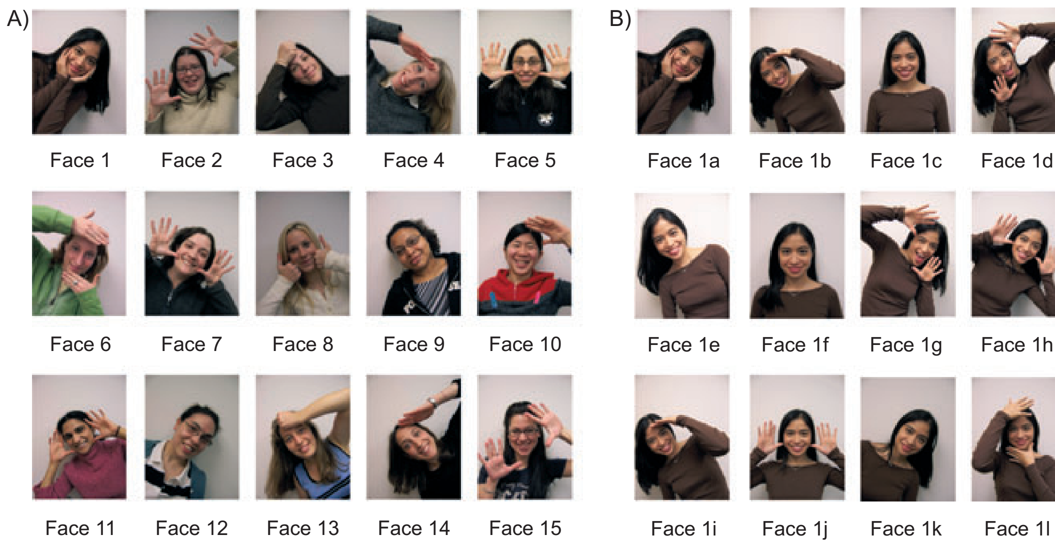

Fig. 1. (A) Face stimuli used in Experiments 1–3. Faces 1–11 were used for Experiments 1 and 2. Faces 1–15 were used for Experiment 3. (B) Examples of face stimuli used in Experiment 2. Children saw each photograph only once.

throughout the task. The faces all had smiling expressions, were approximately matched in age, and had varying hair color and skin tone.

Speech recordings from seven women served as the auditory stimuli. All of the recordings contained the identical utterance using child-directed speech: ''Can you touch my nose?'' The intonation and prosody of these recordings were intentionally varied between individuals. Only one recording from each woman was used, and each recording lasted approximately 1.5 s. The matching of a specific face to a specific voice was arbitrary (i.e., it was never the case that a woman whose photograph was displayed was the actual speaker of the utterance) but was consistent across all participants. These parameters, a single photograph of each woman and a single auditory speech sample, were chosen as an initial test of whether children could learn these mappings. In Experiment 2, we expand our method to include multiple views of each woman such that no single photograph was ever repeated during the course of the experiment.

2.1.3. Procedure

Children sat before a large screen flanked by two speakers. Parents or caregivers sat behind children and were instructed not to interact with the child during the task. Stimulus presentation was controlled by an experimenter seated out of sight in the back of the room.

2.1.3.1. Training phase: During the Training phase children were exposed to three face-voice pairings (i.e., three faces each paired with a unique voice; Fig. 2). There were four Introduction trials in which just a single face appeared onscreen and a single voice was played, thereby making the correct face-voice mapping unambiguous. Two of these

Fig. 2. Example Training trials for Experiments 1–3, with the target face outlined in bold. In Experiment 2, each time a Known Face appeared, it was a single unique photograph that was shown only once during the experiment.

Introduction trials presented Known Face 1 (Trials 1 and 2),[4] one presented Known Face 2 (Trial 4), and one presented Known Face 3 (Trial 6). On each of these Introduction trials the single face appeared and 3 s later a voice played (the voice always played from both speakers to eliminate any spatial cues). The voice always produced the same utterance, ''Can you touch my nose?'' Children received a wand with which to touch the face and were instructed to ''Use the wand to touch the person who is talking.'' Because these Introduction trials presented only one face onscreen, all children touched the wand to the correct face.

Interspersed with these Introduction trials were 12 Training Pair trials in which two faces appeared simultaneously and one voice played, and children had to choose between the faces (Fig. 2; see Table 1 in the Appendix for the full trial order). On these Training Pair trials the faces and voices were always the same three that had been seen during the Introduction trials, and were therefore referred to as ''Known Faces'' and ''Known Voices.'' On all of the Training Pair trials a known target (Known Face 1, Known Face 2, or Known Face 3) and a known distractor (one of the other two Known Faces) appeared side by side onscreen and a Known Voice played. For example, a Training Pair trial might present Known Face 1 on the left and Known Face 2 on the right, while Known Voice 1 played. This was called a ''known-target-with-known-distractor'' trial because both the face that had been paired with the voice (i.e., the target face) and the face that had not been paired with the voice (i.e., the distractor face) had already been presented during the Introduction trials.

Children received computerized feedback on all Training Pair trials. If they chose the correct face, a pleasant chime sounded and both faces simultaneously disappeared from the screen. If they chose the incorrect face, a buzzer sounded, both faces remained onscreen while the same voice played again, and children received another chance to choose the correct face. We allowed two consecutive incorrect choices within the same trial; after the second error, the experimenter verbally prompted the child to select the other face (10% of all Training Pair trials across all children, range 0–25%). If children were hesitant to choose between the two faces, the experimenter encouraged them without revealing the correct answer. All children included in the final sample willingly completed all of the Training Pair trials and, crucially, required no prompting during the subsequent Test trials.

2.1.3.2. Test phase: After the Training phase the experiment immediately proceeded to the 14 Test trials in which children received no explicit feedback about their choices. After every Test trial, regardless of whether children chose correctly or incorrectly, their choice was followed by a neutral tone and the simultaneous disappearance of both faces. During this Test phase children received six additional known-target-with-known-distractor trials as well as two new types of trials (four trials of each; see Appendix, Table 1).

On ''known-target-with-novel-distractor'' trials one Known Face and one Novel Face (i.e., a new face that had not appeared during the Training phase) appeared, and a Known Voice (that had been paired with the Known Face during the Training phase) was played (Fig. 3). On these trials the Known Face was the correct choice. These trials examined children's memory for the face-voice pairings they had learned during the Training phase, and their ability to select the Known Face in the presence of a potentially distracting Novel Face. On ''novel-target-with-known-distractor'' trials one Known Face and one Novel Face appeared, and a Novel Voice (that had never been heard before) was played. On these trials the Novel Face was the correct choice. These trials examined children's ability to use their recently acquired knowledge to make an inference about a further face-voice mapping. If children assume a one-to-one mapping between faces and voices, and if they can use their knowledge of Known Face-Known Voice pairings to reason productively about new faces and voices, they should reject the Known Face and correctly pair the Novel Voice with the Novel Face on these trials.

Fig. 3. Example of Test trials for Experiments 1 and 2, with the target face outlined in bold.

Each Novel Face and Novel Voice appeared on only a single test trial (either as a target or a distractor) and was never repeated. Thus, children never saw the same Novel Face or heard the same Novel Voice twice. Throughout both the Training and Test phases of the experiment, the correct face appeared equally often on either side of the screen. The side on which the correct face was presented was pseudo-randomized across trials, with the restriction that it could not appear on the same side of the screen on more than two consecutive trials.

Children's choices were later coded from video by a coder who was blind to which side was correct. Children were judged to have made a choice if they touched the wand to one of the two faces after the voice had begun playing. Choices made before the voice had played were not scored, because they could not have reflected an attempt to match the face with the voice (approximately 3% of all trials). Children were given approximately 5 s after the voice stopped speaking to make a choice, after which the experimenter encouraged them to make a choice (5% of all trials). A secondary coder who was also blind to the correct side coded the videos for six randomly selected children and agreed with the primary coder on 100% of all trials.

2.2. Results

2.2.1. Training phase

We first asked whether children were successful at learning the face-voice pairings of the three faces presented during Training. Children were correct on 77.1% of the known-target-with-known-distractor trials, on which two Known Faces appeared and one Known Voice played, $t(15) = 6.45$, $p < .0001$, Cohen's $d = 1.61$, two-tailed t test against the chance level of 50% (Fig. 4). This suggests that children successfully learned the three Known Face-

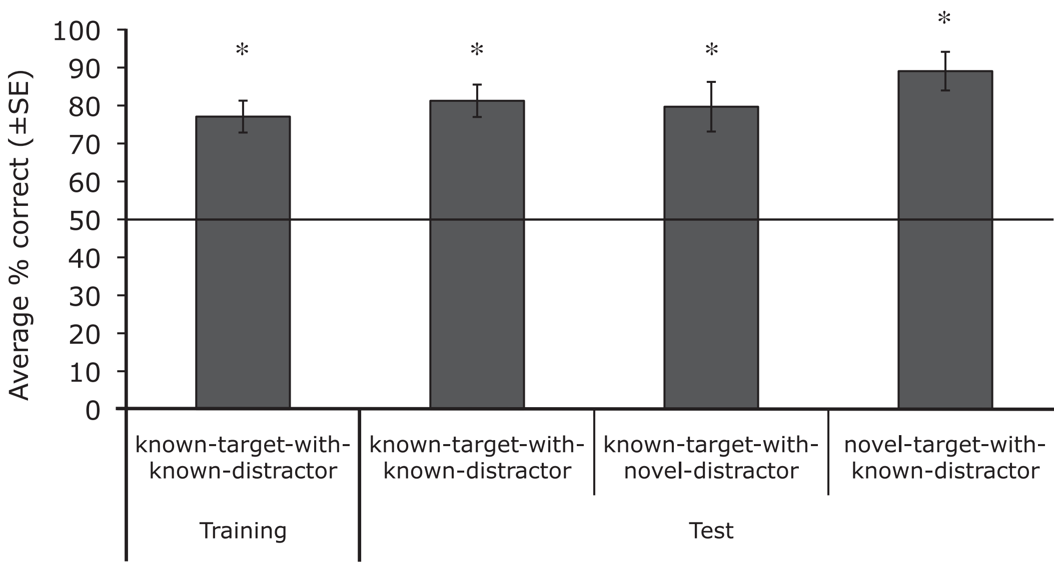

Fig. 4. Children's performance in Experiment 1. Solid line denotes chance performance (50%). Asterisks denote $p < .05$.

Known Voice pairings. Although there was some improvement in performance from the first six training trials (70.8% correct) to the second six training trials (83.3%), this difference was not significant—$t(15) = 1.54$, $p = .14$, two-tailed paired sample t test—suggesting that children learned the pairings rapidly.

Boys and girls did not perform differently—two-tailed two-sample t test, $t(14) = 0.63$, $p = .54$—and there was no difference between the performance of younger (mean age = 54 months) versus older children (mean age = 62 months), as revealed by a split-median two-tailed two-sample t test, $t(14) = 0.060$, $p = .95$.

2.2.2. Test phase

We next examined children's memory retention of the face-voice pairings learned during Training (via known-target-with-known-distractor and known-target-with-novel-distractor trials), and their ability to map new faces to new voices (via novel-target-with-known-distractor trials). We averaged children's performance for each Test trial type and entered these averages into a 3 (Trial Type: known-target-with-known-distractor, known-target-with-novel-distractor, novel-target-with-known-distractor) × 2 (Gender) ANOVA. Children performed equally well on all three trial types, $F(2,14) = 1.08$, $p = .35$ (Fig. 4). There were no differences between boys and girls, $F(1,14) = 0.38$, $p = .55$, nor any interaction between Trial Type and Gender, $F(2,28) = 0.54$, $p = .59$.

Planned 2-tailed t tests against chance (50%) revealed that children succeeded on all three trial types: known-target-with-known-distractor, 81.3% correct, $t(15) = 7.32$, $p < .0001$, Cohen's $d = 1.83$; known-target-with-novel-distractor, 79.7% correct, $t(15) = 4.54$, $p < .0001$, Cohen's $d = 1.13$; novel-target-with-known-distractor, 89.1% correct, $t(15) = 7.68$,

$p < .0001$, Cohen's $d = 1.92$. Children's success on the known-target-with-known-distractor and known-target-with-novel-distractor trials suggests that they retained the Known Face-Known Voice mappings they had learned during the Training phase of the experiment. In addition, children made inferences about new faces and voices. Their success on novel-target-with-known-distractor trials shows that children correctly paired Novel Faces with Novel Voices in the absence of explicit instruction or feedback. This was not simply due to a bias to choose the novel face, as shown by children's success on known-target-with-novel-distractor trials, in which they correctly chose a Known Face over a Novel Face.

2.3. Discussion

The data from Experiment 1 support two conclusions. First, children's success at choosing the correct Known Face upon hearing a Known Voice (during both the Training and Test phases) reveals that 4- and 5-year-old children learned specific mappings between individual faces and voices over the course of a short training session. Furthermore, they did so using only arbitrary intermodal information, as there were no spatial or temporal cues linking the faces and voices. Hence preschoolers appear able to rapidly integrate visual and auditory features from relatively sparse input.

Second, children's success on the novel-target-with-known-distractor trials suggests that they maintain a one-to-one mapping bias between faces and voices (i.e., that children expect each unique face to be mapped to a single unique voice). Children were able to use this bias to infer a pairing between a novel face and a novel voice given just a single exposure and no explicit instruction or feedback. A similar principle has been demonstrated in word learning (e.g., Halberda, 2006; Markman & Wachtel, 1988). To our knowledge, this is the first demonstration of such an inference in the domain of social cognition.

However, we temper our claims by noting that the visual stimuli we presented were quite different from the actual experiences of people that children have in daily life. In particular, children in Experiment 1 saw just a single static photograph of each person, whereas outside of the laboratory, children experience people from multiple different viewpoints, with different physical postures and different emotional expressions. Hence from Experiment 1, the robustness of the face-voice mappings formed by children is unclear. Did children form abstract representations that were resilient against changes in posture or viewpoint? Or were children's representations limited to the particular photographic token of each person that they had seen many times throughout the experiment? A similar issue arises in the case of word learning, in which listeners must recognize multiple tokens, spoken with different intonation or by different speakers, as being the same word. In the domain of word learning, this problem appears to be solved during the first year of life (Jusczyk, Pisoni, & Mullennix, 1992; Kuhl, 1979, 1983).

In Experiment 2, we investigated the abstractness of children's face-voice representations. Children saw photographs of the same women as in Experiment 1, but each time a particular woman appeared she had a different facial expression and different hand positions. We predicted that children would still succeed at forming stable face-voice pairings despite never seeing the same photograph twice.

3. Experiment 2

3.1. Methods

3.1.1. Participants

We anticipated that Experiment 2 would be more difficult than Experiment 1 due to the nonrepetition of the stimulus photographs. Therefore, we introduced a new training criterion (see below) and increased our sample size in order to examine any potential differences between children who met the training criterion (at least six trials correct out of the last eight training trials) versus those who did not. Thirty children (mean age: 4 years 9 months, range: 4 years 0 months to 5 years 5 months; 11 boys) were included in the final data analysis. Fourteen additional children were excluded for reasons potentially related to the method: Six for losing interest during the task, five for side bias (defined as choosing the face on either the right or left side of the screen on more than 75% of all trials), and three for persisting in a premature behavioral response unrelated to the actual trial (i.e., for choosing of the two stimulus faces prior to hearing the voice on more than 50% of all trials).

3.1.2. Stimuli

The same women posed for the stimulus photographs in Experiment 2 as in Experiment 1. But whereas Experiment 1 used only a single image of each person, Experiment 2 used 17 different images of each person (Fig. 1B). Children never saw any photograph more than once during the course of the experiment. The stimulus voices were identical to those in Experiment 1, and as in Experiment 1 children heard just a single token of each Known Voice throughout the experiment.

3.1.3. Procedure

Because children never saw the same token of a Known Face twice, we anticipated that Experiment 2 might be more difficult than Experiment 1. To test this prediction we presented a subset of children with additional training trials. These trials were identical to the initial training trials, and simply provided children with more opportunity to see varying tokens of the three Known Faces presented during the Training phase. We anticipated that this additional exposure would help children store a more robust representation of each of the three Known Faces presented during Training.

The number of Training trials each child received depended on their performance. All children were initially shown four Introduction trials and 12 Training Pair trials, just as in Experiment 1. As these trials progressed the experimenter tabulated the number of correct choices each child made. Based on their performance children were assigned to one of three groups. Group 1 consisted of children who met the criterion of choosing correctly on at least six of the last eight Training trials. The remaining children failed to meet this criterion, choosing correctly on five or fewer of the last eight Training trials. These children were randomly assigned to either Group 2 or Group 3. Children in Group 2 proceeded directly to the Test trials without further training practice. Children in Group 3 received additional Training trials until either they reached the training criterion or had received 24 total

Training trials (average number of additional Training trials for children in Group 3 = 9.2).[5] As in Experiment 1, all three groups proceeded directly from Training to Test trials without a break in between. Regardless of group, all children received 14 Test trials (six known-target-with-known-distractor trials, four known-target-with-novel-distractor trials, and four novel-target-with-known-distractor trials). For a complete list of trials see Table 2 in the Appendix.

Children's choices were coded from video. A secondary coder who was blind to the correct side coded the videos for 10 randomly selected children and agreed with the primary coder on 100% of all trials.

3.2. Results

3.2.1. All Children

3.2.1.1. Training phase: Children's performance on the first 12 Training Pair trials in Experiment 2 (63.7%) was significantly better than chance, $t(29) = 3.50$, $p = .002$, Cohens' $d = 0.74$ (Fig. 5). Performance on the extra Training Pair trials, received only by children in Group 3, is discussed below.

Overall, children performed better during the Training phase of Experiment 1 (when faced with just one photographic token of each person) than in the Training phase of Experiment 2 (when faced with multiple photographic tokens of each person). Across the first 12 Training Pair trials (with no extra training trials included), this difference was significant (77.1% correct across Experiment 1 Training Pair trials vs. 63.7% correct across Experiment 2 Training Pair trials, $t(44) = 2.42$, $p = 0.02$, Cohen's $d = 0.73$, two-tailed two-sample t test.

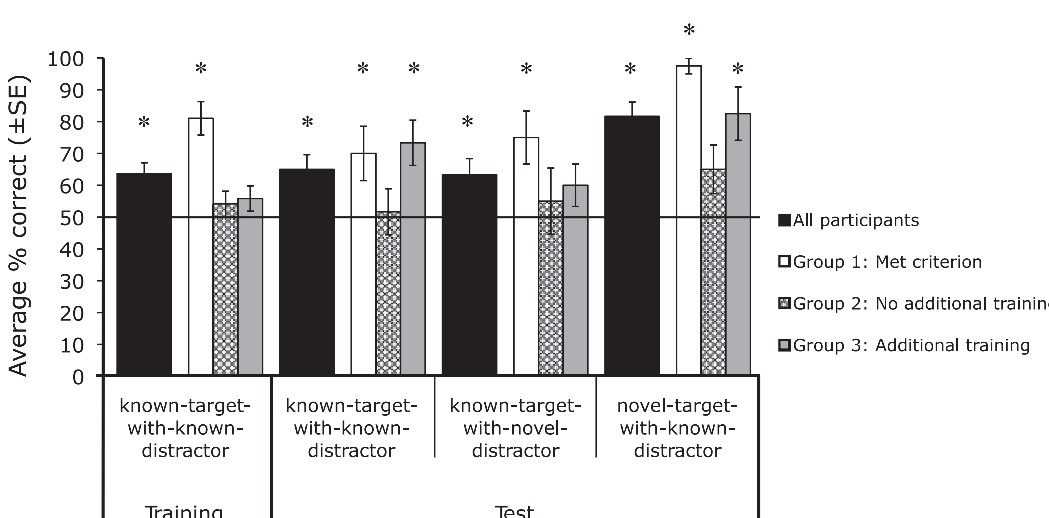

Fig. 5. Children's performance in Experiment 2. Solid line denotes chance performance (50%). Asterisks denote $p < .05$.

3.2.1.2. Test phase: We first analyzed Test trial data from all 30 children as a single group, regardless of their performance during Training. A 3 (Trial Type: known-target-with-known-distractor, known-target-with-novel-distractor, novel-target-with-known-distractor) × 2 (Gender) ANOVA revealed a significant main effect of Trial Type, $F(2,56) = 8.24$, $p < .001$, $\eta_p = 0.227$ (Fig. 5). There was also a significant main effect of Gender due to girls outperforming boys, $F(1,28) = 6.10$, $p = .02$, $\eta_p = 0.179$. The interaction between Trial Type and Gender was not significant, $F(2,56) = -0.86$, $p = .43$, $\eta_p = 0.030$.

We further explored the main effect of Trial Type with planned *t* tests. Children performed better on novel-target-with-known-distractor trials than on known-target-with-known-distractor trials—81.7% vs. 65.0%, $t(29) = 3.20$, $p = .003$, Cohen's $d = 1.19$—and on novel-target-with-known-distractor trials than on known-target-with-novel-distractor trials—81.7% vs. 63.3%, $t(29) = 3.00$, $p = .005$, Cohen's $d = 1.11$. There was no difference in performance for known-target-with-known-distractor trials and known-target-with-novel-distractor trials—65.0% vs. 63.3%, $t(29) = 0.41$, $p = .69$.

Finally, children were above chance for all three Trial Types: known-target-with-known-distractor: 65.0% correct, $t(29) = 3.25$, $p = .003$, Cohen's $d = 0.59$; known-target-with-novel-distractor: 63.3% correct, $t(29) = 2.64$, $p = .013$, Cohen's $d = 0.48$; novel-target-with-known-distractor: 81.7% correct, $t(29) = 7.08$, $p < .0001$, Cohen's $d = 1.29$ (Fig. 5).

3.2.2. Group analyses

3.2.2.1. Training phase: Of the 30 children included in the final analysis, 10 met the training criterion during the initial 12 Training Pair trials (Group 1: average age: 4 years 11 months, range: 4 years 7 months to 5 years 4 months; four boys) and 20 did not (average age: 5 years 0 months, range: 4 years 0 months to 5 years 6 months; six boys). The children who did not meet the criterion were randomly assigned to Group 2 (no additional training) or Group 3 (additional training) (Group 2: average age: 4 years 9 months, range: 4 years 4 months to 5 years 6 months; three boys; Group 3: average age: 4 years 7 months, range: 4 years 0 months to 5 years 3 months; three boys). There were no significant differences between Groups 1, 2, and 3 in gender, $\chi^2(2) = 0.30$, $p = .86$, or in age—Kruskal-Wallis one-way ANOVA, adjusted $H(2) = 5.097$, $p = .08$. For the subsequent analyses, performance on all Training Pair trials (12 trials for children in Groups 1 and 2, up to 20 trials for children in Group 3) was included.

A 3 (Group) × 2 (Gender) ANOVA revealed a significant effect of Group, $F(2,24) = 7.31$, $p = .003$, $\eta_p = 0.379$. Neither the main effect of Gender nor the interaction was significant. The main effect of group was further explored using planned *t* tests. Children in Group 1 (i.e., met criterion) performed better than children in Group 2 (i.e., no additional training), $t(18) = 4.08$, $p = .0007$, Cohen's $d = 1.87$, and children in Group 3 (i.e., additional training), $t(18) = 3.83$, $p = .001$, Cohen's $d = 1.76$. Children in Groups 2 (no additional training) and 3 (additional training) did not differ in their performance during the Training phase, $t(18) = 0.30$, $p = .77$.

Children in Group 1 (met criterion) were correct on 81.0% of Training Pair trials, performing significantly above chance—$t(9) = 5.91$, $p = .0002$, Cohen's $d = 1.87$, two-tailed t test against 50% (Fig. 5). These children demonstrated marginal improvement from the first (71.7%) to the second half (86.7%) of Training Pair trials—$t(9) = 1.96$, $p = .08$, Cohen's $d = 1.31$, two-tailed paired t test. Children in Group 2 (no additional training) were correct on 54.2% of Training Pair trials, performing at chance, $t(9) = 1.05$, $p = .32$. They showed no significant improvement from the first half (50.0%) to the second half (58.3%) of the Training Pair trials, $t(9) = 0.83$, $p = .43$. Children in Group 3 (additional training) were correct on 55.8% of Training Pair trials. Although this performance was not above chance, $t(9) = 1.47$, $p = .17$, these children did show improvement over the training session, performing significantly better on the second half of Training Pair trials (58.3%) than on the first half of Training Pair trials (51.5%)—$t(9) = 2.50$, $p = .03$, Cohen's $d = 1.67$, two-tailed paired t test.

3.2.2.2. Test phase: Next we examined children's performance on the critical test trials. A 3 (Group: Group 1, Group 2, Group 3) × 3 (Trial Type: known-target-with-known-distractor, known-target-with-novel-distractor, novel-target-with-known-distractor) × 2 (Gender) ANOVA revealed all main effects to be significant (Group: $F(2,24) = 5.07$, $p < .0001$, $\eta_p = 0.297$; Trial Type: $F(2,48) = 10.69$, $p < .0001$, $\eta_p = 0.308$; Gender: $F(1,24) = 9.90$, $p = .004$, $\eta_p = 0.292$, with girls outperforming boys overall). There was a three-way interaction between Group, Trial Type, and Gender, $F(4,48) = 3.46$, $p = .02$, $\eta_p = 0.224$, resulting from boys, but not girls, in Group 2 performing better on novel-target-with-known-distractor trials than on other trial types.

We further explored the main effects of Group and Trial Type using post-hoc contrasts. Pair-wise comparisons using Tukey's Honestly Significant Difference (Tukey's HSD) revealed that children in Group 1 (met criterion) performed significantly better across all Test trials than those in Group 2 (no additional training), $q_s(3, 24) = 23.61$, $p = .007$; no other group comparisons were significant (Fig. 5). For Trial Type, children's performance averaged across the three groups did not differ on known-target-with-known-distractor versus known-target-with-novel-distractor trials, but it was significantly better on novel-target-with-known-distractor trials than on the other two trial types (within-subject contrasts: novel-target-with-known-distractor vs. known-target-with-known distractor: $F(1,24) = 15.97$, $p = .001$, $\eta_p = 0.399$; novel-target-with-known-distractor vs. known-target-with-novel-distractor: $F(1,24) = 12.54$, $p = .002$, $\eta_p = 0.343$).

We next analyzed children's performance on the three different Test trial types according to Group. Children in Group 1 (met criterion) were correct on 70.0% of known-target-with-known-distractor trials, 75.0% of known-target-with-novel-distractor trials, and 97.5% of novel-target-with-known-distractor trials—all above chance with two-tailed t test: $t(9) = 2.33$, $p = .044$, Cohen's $d = 0.74$; $t(9) = 3.00$, $p = .015$, Cohen's $d = 0.95$; $t(9) = 19.00$, $p < .0001$, Cohen's $d = 6.01$, respectively. Despite the increased difficulty of the task introduced by having additional tokens of each face, children in Group 1 (met criterion) successfully learned the three face-voice pairings (as demonstrated by their success on known-target-with-known-distractor and known-target-with-novel-distractor trials) and used this knowledge to motivate mappings of Novel Faces to Novel Voices (as demonstrated by their success on novel-target-with-known-distractor trials).

Children in Group 2 (no additional training) performed at chance on all three Test trial types (known-target-with-known-distractor: 51.7%, $t(9) = 0.23$, $p = .82$; known-target-with-novel-distractor: 55.0%, $t(9) = 0.48$, $p = .64$; novel-target-with-known-distractor: 65.0%, $t(9) = 1.96$, $p = .08$. Combined with their poor performance during the training phase, this confirms that Group 2 children did not successfully learn the three Known Face-Voice pairings. It is therefore not surprising that they were unable to use these pairings in order to infer new mappings when presented with Novel Faces and Voices.

Children in Group 3 (additional training) were correct on 73.3% of known-target-with-known-distractor trials, $t(9) = 3.28$, $p = .01$, Cohen's $d = 1.04$, 60.0% of known-target-with-novel-distractor trials, $t(9) = 1.50$, $p = .17$, and 82.5% of novel-target-with-known-distractor trials, $t(9) = 3.88$, $p = .004$, Cohen's $d = 1.23$. The experience of additional Training trials led Group 3 children to succeed on most of the Test trials. The performance of these children was identical to that of Group 2 children (no additional training) at the end of the first 16 Training trials, and children were randomly assigned to either Group 2 (no additional training) or Group 3 (additional training). Thus, the success of Group 3 children (additional training) suggests that children are capable of learning these new face-voice pairings even when presented with unique tokens of each face, but that some children may require further experience in order to do so. Furthermore, children's success on novel-target-with-known-distractor trials suggests that they used a one-to-one mapping bias to learn new face-voice pairings.

3.3. Discussion

The results of Experiment 2 replicate and extend those of Experiment 1.[6] Children in Groups 1 (met criterion) and 3 (additional training), who had successfully formed face-voice mappings during the Training phase, were able to use these mappings in order to make inferences about new faces and voices during the Test phase. As in Experiment 1, these results are consistent with the suggestion that children employed a one-to-one mapping bias to constrain their hypotheses about mappings between faces and voices. The success of these children also expands upon the ability demonstrated in Experiment 1, in that children in Experiment 2 never saw the same photograph more than once, and therefore had to rely on more abstract person representations in order to make their choices.

The variability in the photographs shown in Experiment 2 exerted a measurable cost on children's performance. Children in Group 2 (no additional training) performed poorly during the initial 12 Training Pair trials, did not receive additional Training Pair trials, and also performed poorly during the Test trials. Children in Group 3 (additional training) also performed poorly during the initial 12 Training Pair trials, and so received an average of 9.2 additional Training Pair trials before beginning the Test trials. These additional Training Pair trials improved the performance of Group 3 children relative to Group 2 children. This difference in the speed of children's learning motivates a further question concerning the parallels between word learning and social cognition.

In word learning, it has been demonstrated that children can learn a mapping between a novel word and its referent from a single brief exposure. For example, when shown a blue tray and an olive-green tray and asked to ''Find the chromium tray, not the blue one, the chromium one,'' preschoolers correctly chose the olive-green tray in response to the novel word ''chromium'' and retained information about this mapping (Carey & Bartlett, 1978).[7] Called ''fast mapping,'' this ability to rapidly form a new mapping between a word and its referent and to retain the mapping in memory has been discussed as one of the mechanisms that supports preschoolers' rapid vocabulary growth (for further discussion of the distinction between selecting the referent of a novel word and retaining it in memory, see Horst & Samuelson, 2008). Some data suggest that fast mapping may be a domain general ability. Markson and Bloom (1997) found that preschoolers taught a novel fact about an object (''This is the one my uncle gave me'') remembered this fast-mapped information when tested 1 week and 1 month later.

Our finding that some children in Experiment 2 required further training in order to form stable mappings between three faces and three voices motivates the question of whether the type of fast mapping ability that has been documented for word learning can also apply to the learning of face-voice mappings. Can children retain in memory a novel face-voice mapping from a single brief exposure and use it as the basis for learning about other new faces? The poor performance of some children in Experiment 2 during training may seem to suggest that children cannot fast map a face to a voice in a single trial (since they apparently were unable to map a face to a voice over many repeated trials). However, in Experiment 2 these trials presented multiple unique tokens of each face throughout both training and test, making learning more difficult. Multiple tokens are rarely used in the word-learning literature. It is therefore an open question whether children can perform single-trial fast mapping when not faced with this high level of stimulus variability. We addressed this question in Experiment 3.

4. Experiment 3

In Experiment 3, we asked (a) whether children retain in memory the face-voice mapping they formed from a single exposure during each novel-target-known-distractor trial, and (b) whether a just-learned face-voice mapping can serve as the basis for a further inference about other faces and voices. We based our method on that from a previous study of fast mapping in word learning. In an experiment by Golinkoff et al. (1992), 2-year-old children were first shown three familiar objects (e.g., brush, fork, crayon) and one unfamiliar object (e.g., broom handle tip), and were asked to ''Find the dax.'' Children successfully chose the unfamiliar object 78% of the time, consistent with having a one-to-one mapping bias between words and objects. Several trials later children were shown two familiar objects and two unfamiliar objects, one of which was the unfamiliar object from the first trial (i.e., the broom handle tip). They were then asked to ''Find the wug.'' The question was whether children who had previously chosen the broom handle tip as the referent of the word ''dax'' would now use this new knowledge to reject the broom handle tip as the referent of the new word ''wug,'' and thereby infer that ''wug'' must refer to the new unfamiliar object. Golinkoff and her colleagues found that 80% of children correctly chose the new unfamiliar object as the referent of ''wug,'' suggesting that they had spontaneously used the information learned in the first trial (i.e., that a broom handle tip is called a ''dax'') to eliminate one of the two objects as a potential referent for ''wug.'' This behavior demonstrates the powerful utility of the one-to-one mapping bias combined with fast mapping; children can use newly acquired knowledge gained via inference to bootstrap even further inferences. A similar computation in the social domain would also allow children to acquire new information about social entities quickly and efficiently. In Experiment 3, we asked whether similar learning might help guide children's learning of novel face-voice mappings.

4.1. Methods

4.1.1. Participants

Sixteen children (mean age: 4 years 9 months; range: 4 years 2 months to 5 years 5 months; nine boys) were included in the final data analysis. Seven additional children were excluded for reasons potentially related to the method: two for losing interest during the experimental task, and five for side bias (defined as choosing the face on either the right or left side of the screen on more than 75% of all trials). Four additional children were excluded for parent or sibling interference during the task.

4.1.2. Stimuli

The stimuli for Experiment 3 were identical to those in Experiment 1, with the addition of four more Novel Faces and two more Novel Voices (Fig. 1). As in Experiment 1, children saw just one photographic token of each face and heard just one auditory token of each voice.

4.1.3. Procedure

The experimental set up and structure of each trial were identical to those in Experiments 1 and 2. In contrast to Experiment 2, children were given no opportunity for extra Training Pair trials.[8]

After the Training phase (comprised of four Introduction trials and 12 Training Pair trials, just as in Experiment 1) children proceeded directly to the Test phase (18 trials, Fig. 6). During the Test phase children were presented with the same three trial types as in Experiments 1 and 2 (known-target-with-known-distractor, known-target-with-novel-distractor, and novel-target-with-known-distractor), along with two new trial types. These two new trial types (called ''just-learned-target-with-novel-distractor'' and ''novel-target-with-just-learned-distractor'' trials) always immediately followed novel-target-with-known-distractor trials (in which children saw a Known Face that had been learned during the Training phase presented next to a Novel Face, and heard a Novel Voice). Regardless of whether children correctly chose the Novel Face, the very next trial presented children with the same Novel Face just seen in the previous trial (now called the Just-Learned Face), alongside another, never before seen Novel Face. On just-learned-target-with-novel-distractor trials, the Just-Learned Voice from the previous trial was played, and the Just-Learned Face was the correct choice. These trials tested whether children could reidentify the Just-Learned Face-Voice mapping they had made during the previous trial (i.e., whether children had retained the

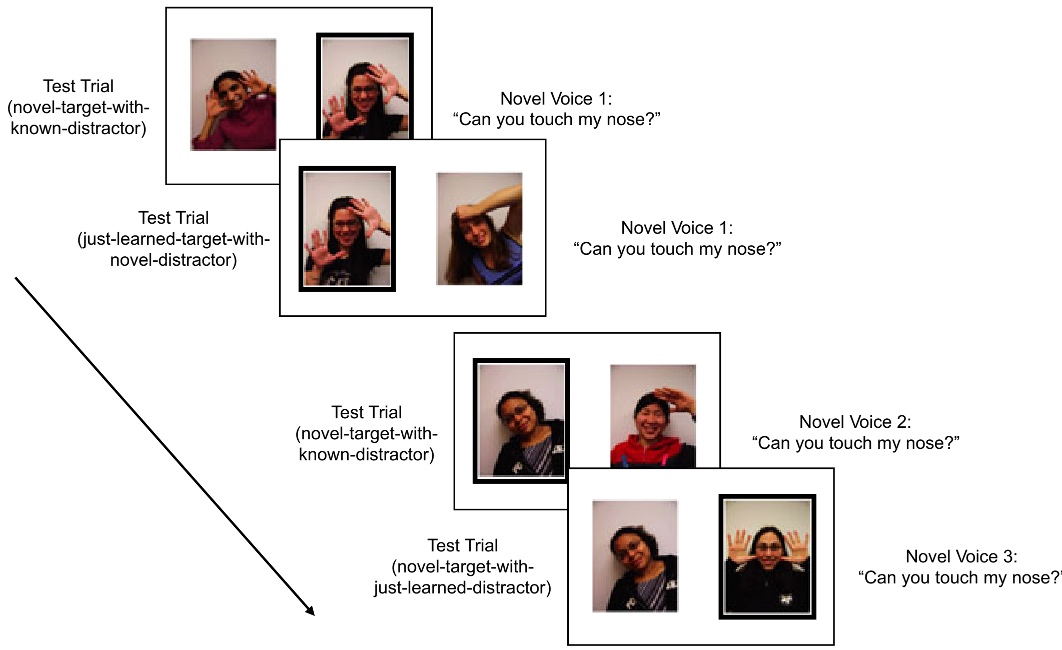

Fig. 6. Example of the two additional Test trial types used in Experiment 3, with the target face outlined in bold.

new face-voice mapping in memory). On novel-target-with-just-learned-distractor trials, a completely Novel Voice that the child had never heard before was played, and the Novel Face was the correct choice. These trials examined whether children could spontaneously use the Just-Learned Face-Voice mapping as the basis for making yet another inference about a novel face and a novel voice. These trials were parallel to the fast mapping trials in the word learning experiment by Golinkoff et al. (1992).

As in Experiments 1 and 2, children received no explicit feedback during the Test phase. There were 18 Test trials (six known-target-with-known-distractor trials, four known-target-with-novel-distractor trials, four novel-target-with-known-distractor trials, two just-learned-target-with-novel-distractor trials, and two novel-target-with-just-learned-distractor trials) presented in pseudorandom order. For a complete list of trial orders, see Table 3 in the Appendix.

Children's choices were coded from video. A secondary coder who was blind to the correct side coded the videos for six randomly selected children and agreed with the primary coder on 100% of all trials.

4.2. Results

4.2.1. Training phase

Children were correct on 63.54% of known-target-with-known-distractor trials, $t(15) = 2.69$, $p = .017$, Cohen's $d = 0.67$, two-tailed t test against 50% (Fig. 7). There was no

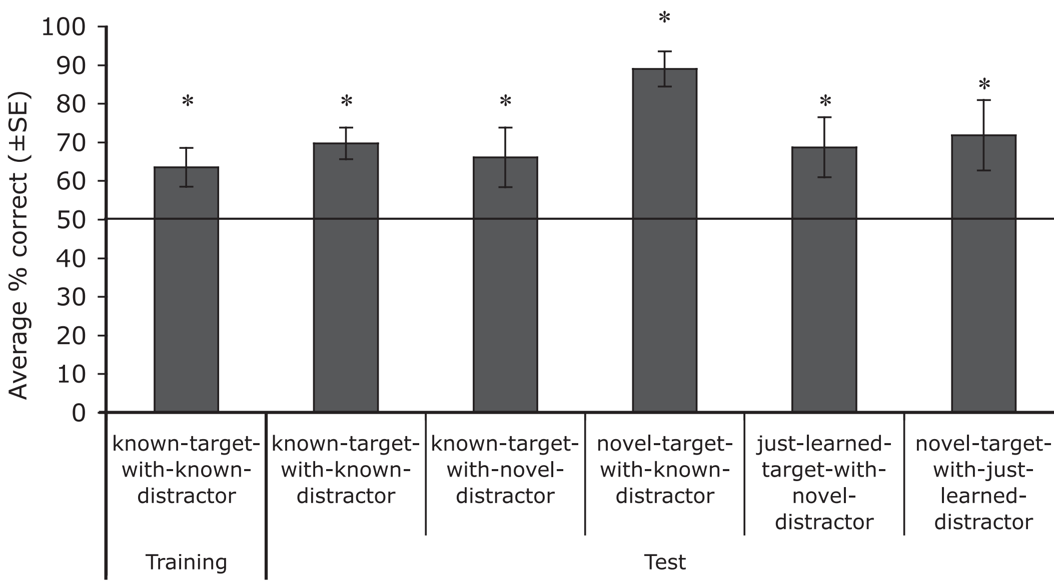

Fig. 7. Children's performance in Experiment 3. Solid line denotes chance performance (50%). Asterisks denote $p < .05$.

improvement from the first half of trials (61.5% correct) to the second half (65.6%), $t(15) = -0.60$, $p = .55$, two-tailed paired sample *t* test. Boys and girls did not perform differently—$t(14) = 0.28$, $p = .79$, two-tailed two-sample *t* test—and there was no difference in the performance of younger (mean age = 53 months) versus older children (mean age = 61 months), as revealed by a split-median two-sample *t* test, $t(14) = 1.18$, $p = .25$.

4.2.2. Test phase

We averaged children's performance for each trial type and entered these into a 5 (Trial Type: known-target-with-known-distractor, known-target-with-novel-distractor, novel-target-with-known-distractor, novel-target-with-just-learned-distractor, just-learned-target-with-novel-distractor) × 2 (Gender) ANOVA. There was no difference in children's performance on the five Test trial types—main effect of Trial Type: $F(4,56) = 1.73$, $p = .16$—nor any difference between boys and girls—main effect of Gender: $F(1,14) = 0.003$, $p = .96$. There was also no interaction between Trial Type and Gender, $F(4,56) = 0.155$, $p = .96$ (Fig. 7).

Planned 2-tailed *t* tests against chance (50%) revealed that children tended to choose the correct face at above chance rates on the three trial types that also appeared in Experiments 1 and 2—known-target-with-known-distractor trials: 69.8% correct, $t(15) = 4.84$, $p = .0002$, Cohen's $d = 1.21$; known-target-with-novel-distractor trials: 66.2% correct, $t(15) = 2.10$, $p = .053$, Cohen's $d = 0.52$; novel-target-with-known-distractor trials: 89.0% correct, $t(15) = 8.60$, $p < .0001$, Cohen's $d = 2.15$ (Fig. 7).

Importantly, children also succeeded on the two new trial types. Children chose correctly on 68.8% of just-learned-target-with-novel-distractor trials, on which they had to reaccess a face-voice mapping that had been made via inference on the previous trial $t(15) = 2.42$, $p = .029$, Cohen's $d = 0.61$. Children also chose correctly on 71.9% of novel-target-with-just-learned-distractor trials, on which they had to use a face-voice mapping that had been made via inference on the previous trial as input to a new inference about a novel face and a novel voice,[9] $t(15) = 2.41$, $p = .030$, Cohen's $d = 0.69$.

4.3. Discussion

Experiment 3 tested the strength of children's newly formed face-voice mappings. Despite the one-time ambiguous exposure to a new face and a new voice on each novel-target-with-known-distractor trial, and despite the absence of any explicit feedback, children were able to both reidentify these just-learned mappings and use them to make yet another productive inference about other novel faces and voices. Indeed, armed with the limited knowledge of just three face-voice mappings, children readily identified four additional mappings on novel-target-with-known-distractor trials, with no explicit feedback. This performance parallels similar successes observed in word learning (Golinkoff et al., 1992) and further extends claims that fast mapping is not specific to word learning but is observed in other domains as well (Markson & Bloom, 1997).

5. General discussion

In three experiments, we explored the hypothesis that children's learning of the mappings between individual faces and voices is supported by biases and constraints similar to those that support children's learning about mappings between objects and words. We tested 4- and 5-year-old children's ability to map novel faces to novel voices, and then assessed the abstractness and robustness of the resulting representations. In Experiment 1, we found that children successfully learned three face-voice mappings over the course of a 3-min training period. This rapid learning was robust enough to support children's explicit choices, as demonstrated by their successful touching of the face that correctly matched a just learned voice. Furthermore, children were able to use this newly acquired knowledge productively to spontaneously generate mappings between new faces and new voices during the Test trials. This success is consistent with use of a one-to-one mapping bias; children assume that each face must be paired with a single voice. Such a one-to-one mapping bias has been shown to help guide children's learning about words and their referents (Carey & Bartlett, 1978; Clark, 1990; Diesendruck, 2005; Golinkoff et al., 1992; Halberda, 2006; Markman & Wachtel, 1988; Markman, Wasow, & Hansen, 2003; Mervis & Bertrand, 1994; Wilkinson & Mazzitelli, 2003). Here, we show that a similar bias may also guide children's learning about social entities.

In Experiment 2, we tested the abstractness of children's face-voice mappings by presenting children with multiple unique tokens of each face. Using multiple unique tokens reduced the likelihood that children would form associations between voices and nonstable elements of particular photographs, such as facial expression or hand position. Although seeing multiple tokens of each face apparently increased the difficulty of the task relative to

Experiment 1, many children successfully learned the face-voice mappings presented during the Training phase. Furthermore, as in Experiment 1 children used these learned mappings to infer new mappings between completely novel faces and voices. This success suggests that children formed intermodal representations that were abstract enough to support recognition across novel perceptual input. Here, too, it is possible to draw a comparison with early word learning, which supports word recognition across multiple speakers (Jusczyk et al., 1992), and creates word-object mappings that are resilient to changes in a referent object's perceptual properties (Smith, Jones, & Landau, 1992).

Finally, in Experiment 3 we addressed two additional questions regarding the robustness of children's newly formed face-voice mappings. First, we asked whether children's performance in Experiments 1 and 2 revealed fast mapping—that is, whether children's rapidly formed face-voice mappings were stored in memory, or whether instead children merely had performed reflexive pointing at a novel face when hearing a novel voice (which would not have been indicative of forming a new representation). To test this we measured children's success at reidentifying face-voice mappings that they had previously made via inference on a preceding trial. We found that children successfully reidentified the just-learned pairings on a subsequent trial, despite having never received any explicit feedback about which face was paired with which voice. Second, we asked whether children's newly formed face-voice pairings supported still further inference. We found that when presented with a just-learned face and a completely novel face, and hearing a completely novel voice, children successfully paired the completely novel face with the completely novel voice. This suggests that the fast mapped representations children formed were stable enough to themselves serve as input for further inference. Combined with the one-to-one mapping bias demonstrated in Experiments 1 and 2, these newly formed representations allowed children to infer mappings between completely novel faces and voices. This is analogous to the way in which children's newly formed object-word mappings support inferences about completely novel objects and words (Golinkoff et al., 1992).

One question raised by this work is the extent to which children's success at forming face-voice mappings reflects characteristics specific to social cognition or person representation, versus domain general abilities. For example, if presented with a mapping task involving arbitrary pairings (e.g., mapping three different shapes to three different sounds), it is possible that children's performance would have been similar to the performance we observed in the present experiments. That some of the qualities demonstrated here are also seen in other domains (e.g., word-action mappings; object-word mappings) (Childers & Tomasello, 2003; Markson & Bloom, 1997; but see Waxman & Booth, 2000) supports this view. We suggest that rather than implicating any domain-specific processes, the importance of the present results for social cognition is in demonstrating that fast mapping and a one-to-one mapping bias can help learning in sparse or ambiguous situations, much as they do in other domains. Whereas some previous work suggests that children require long periods of exposure in order to learn face-voice mappings (Bahrick et al., 2005), or that children can not learn new individual voices in a laboratory task until middle childhood (Mann et al., 1979), here we show that children can quickly learn to pair faces and voices under conditions similar to those used to test early word learning.

However, we encourage caution when considering whether these parallels stem from computational resources that are shared between face-voice learning, word-object learning, or learning in other domains. If reasoning in these varied domains derives from domain general pragmatic reasoning (Bloom & Markson, 2001; Diesendruck, 2005; Markson & Bloom, 1997), then it is possible that the success seen here relies on the same mechanisms that empower rapid word learning and rapid mapping between objects and facts (Markson & Bloom, 1997). Alternatively, social cognition, word learning, and fact learning might instead benefit from computations that function similarly, but separately, to solve their respective mapping problems.

Indeed, several different mechanisms have been posited to underlie the one-to-one mapping strategy: Mutual Exclusivity (Markman & Wachtel, 1988), the principle of Contrast (Clark, 1990), a Pragmatic Account (Diesendruck & Markson, 2001), and the Novel-Name Nameless-Category Principle (Golinkoff et al., 1992). In order to determine which one of these mechanisms underlies the specific performance observed in the present experiments, additional fine-grained measures (such as eye movements; Halberda, 2006) would be necessary. In previous demonstrations of fast mapping in domains other than word learning (Childers & Tomasello, 2003; Markson & Bloom, 1997), researchers have remained largely agnostic as to the specific mechanism underlying children's performance in these tasks. We remain similarly agnostic but see the present results as opening the door to further exploration of mechanism. Furthermore, any search for mechanisms specific to social cognition would benefit from investigating learning in situations with greater ecological validity than the word-learning-inspired methods we used here as an initial test of fast mapping and of the one-to-one mapping bias in person representation.

Many questions remain regarding the degree to which face-voice learning is similar to word-object learning. One question is to what extent children's newly formed face-voice mappings endure over time. In word learning and some other domains, children's fast mapped representations have been shown to last for a week or more (Markson & Bloom, 1997; but see Horst & Samuelson, 2008). Future work on children's face-voice mappings will be needed to determine whether these mappings endure over longer time scales, as do children's word-object or object-fact mappings.

It is also unclear whether and how children's knowledge about the relationship between faces and voices affects children's fast mapping. For example, even in infancy children appear to match male faces to male voices and female faces to female voices (Patterson & Werker, 2002), and to match adult faces to adult voices and children's faces to children's voices (Bahrick, Netto, & Hernandez-Reif, 1998). Such assumptions on the part of the learner concerning the likely mappings between social stimuli must be integrated with the one-to-one mapping bias between faces and voices, much as children's knowledge about the syntactic context in which a new word is heard must be integrated with the one-to-one mapping bias between objects and words (Diesendruck, Hall, & Graham, 2006). How this integration occurs remains an area for further inquiry.

A third question concerns the development of both the one-to-one mapping bias and fast mapping for faces and voices. Here, we tested older preschoolers because we wanted children to be able to learn three new face-voice pairings and to use these as ''known'' mappings throughout the experiments. This was challenging even for 4- and 5-year-old children, as demonstrated by the fact that, despite their above-chance performance, children performed far from ceiling. The developmental trajectory of these biases remains to be explored. What is the youngest age at which the one-to-one mapping bias and fast mapping of faces and voices can be observed? In word learning, these abilities can be observed as early as 17 months (Halberda, 2003). Might the one-to-one bias and fast mapping of faces and voices also be active in these younger children? Are these biases the product of early learning about faces and voices, or are they initial assumptions that help get social cognition off the ground? Answering these questions will require testing younger children.

In summary, the experiments presented here offer some intriguing parallels between children's learning across seemingly disparate domains. Although different challenges are faced when learning about social agents and learning new words, here we show that the processes supporting learning in these domains share some characteristics. Using previous findings from the word learning literature as inspiration, the present studies reveal that children are adept at rapidly forming new representations of social entities, arguably one of the most important tasks they face.

Notes

1. This work is inspired by two lines of research pioneered by Susan Carey: the development of face processing abilities and early word learning.
2. We do not imply that use of such a bias need be explicit or under children's conscious control.
3. This high exclusion rate suggests that caution should be used in generalizing from the sample in Experiment 1. However, the lower exclusion rates in Experiments 2 and 3, which were more challenging tasks for the children, assuage potential concerns about exclusion.
4. Known Face 1 appeared in isolation on two of the Introduction trials in order to highlight for children at the start of the experiment that the face that matched the voice could appear on either the right or left side of the screen.
5. Many factors may have contributed to these individual differences in children's ability to meet our Training criterion. These include individual differences in executive function, auditory or visual processing, ability to form multimodal mappings, or attentiveness during the task. Although the source of individual differences in this task is beyond the scope of the present paper, this topic remains ripe for future investigation.
6. Although no children in Experiment 1 received extra Training trials, for the purpose of comparison we performed the same additional performance-based analyses of children in Experiment 1 as in Experiment 2. When children in Experiment 1 were separated into two groups based on their performance during the Training phase, we found that the two children who did not meet Experiment 2's training criterion of at least six correct on the last eight training trials performed significantly worse in the Test phase (53.6% correct) than the 14 children who met the criterion (87.3% correct) (Wilcoxon

Two Sample Test, $W = 3.5$, $p = .03$, two-tailed). Not surprisingly, this suggests that children's success in the test phase was influenced by their learning during the training phase.

7. Later research expanding on this result suggests that learning some classes of words (e.g., nouns) via fast mapping is easier than learning others (e.g., adjectives: Au & Markman, 1987; for recent discussion of this issue, see Sandhofer & Smith, 2007).
8. Because Experiment 3, like Experiment 1, presented children with just one token of each face, and because we observed high levels of performance in Experiment 1, we anticipated that as a group children would not require any extra trials in order to learn the Known Faces and Voices during Training.
9. In Experiment 2, we found it informative to separate children into groups based on their performance during Training, and to give additional training to the children who did not meet the criterion of correct pointing on six of the last eight Training trials. In Experiment 3, we did not provide any such additional training. However, for the purpose of comparison we conducted the same performance-based analyses of children in Experiment 3 as in Experiment 2. We found that the seven children who did not meet the training criterion of at least six correct on the last eight training trials performed worse across all test trial types than the nine children who met the criterion. However, this trend did not reach significance (did not meet criterion: 68.4% correct; met criterion: 78.8% correct), $t(14) = 1.52$, $p = .15$, two-tailed two-sample t test. This suggests that, as in Experiment 2, some of the children may have benefited from further trials in order to strengthen the representations of the face-voice pairs learned during Training.

Acknowledgments

We thank families who participated in this study; Rachel Austin, Mark Butler, Caitlin Duffy, Andi Fu, Catherine Juravich, Andrea Stevenson, and Allison Wessell for their help with testing the participants; the women who volunteered to have their pictures taken and to have their voices recorded for stimuli; Michael Lee, Barbara Sarnecka, and one anonymous reviewer for their generous and helpful comments on previous manuscripts. This work was supported by a Scholar Award from the James S. McDonnell Foundation to Lisa Feigenson and a National Science Foundation Graduate Research Fellowship to Mariko Moher (as Mariko Yamaguchi).

References

Au, T. K., & Markman, E. M. (1987). Acquiring word meaning via linguistic contrast. *Cognitive Development*, *2*, 217–236.

Bahrick, L. E., Hernandez-Reif, M., & Flom, R. (2005). The development of infant learning about specific face-voice relations. *Developmental Psychology*, *41*(3), 541–552.

Bahrick, L. E., Lickliter, R., Shuman, M. A., Batista, L. C., & Grandez, C. (2003). Infant discrimination of voices: Predictions from the intersensory redundancy hypothesis. Poster presented at the biennial meeting of the Society for Research in Child Development, Minneapolis, MN.

Bahrick, L. E., Lickliter, R., Vaillant, M., Shuman, M., & Castellanos, I. (2004 May). Infant discrimination of faces in the context of dynamic, multimodal events: Predictions from the intersensory redundancy hypothesis. Poster presented at the International Conference on Infant Studies, Chicago, IL.

Bahrick, L. E., Netto, D., & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development*, *69*(5), 1263–1275.

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*(2), B45–B53.

Barrera, M. E., & Maurer, D. (1981). Discrimination of strangers by the three-month-old. *Child Development*, *52*, 558–563.

Bartholomeus, B. (1973). Voice identification by nursery-school children. *Canadian Journal of Psychology*, *27*(4), 464–472.

Bloom, P., & Markson, L. (2001). Are there principles that apply only to the acquisition of words? A reply to Waxman and Booth. *Cognition*, *78*, 89–90.

Brookes, H., Slater, A., Quinn, P. C., Lewkowicz, D. L., Hayes, R., & Brown, E. (2001). Three-month-old infants learn arbitrary auditory-visual pairings between voices and faces. *Infant and Child Development*, *10*, 75–82.

Brooks, R. M., & Goldstein, A. G. (1963). Recognition by children of inverted photographs of faces. *Child Development*, *34*, 1033–1040.

Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant and Child Development*, *10*, 67–74.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, *11*(12), 535–543.

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

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference*, *15*, 17–29.

Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition—A maturational component? *Developmental Psychology*, *16*(4), 257–269.

Childers, J. B., & Tomasello, M. (2003). Children extend both words and non-verbal actions to novel exemplars. *Developmental Science*, *6*(2), 185–190.

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, *17*, 417–431.

Cleary, M., Pisoni, D. B., & Kirk, K. I. (2005). Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, *48*, 204–223.

Cohen, L. B., & Strauss, M. S. (1979). Concept acquisition in the human infant. *Child Development*, *50*, 419–424.

Cornell, E. H. (1974). Infants' discrimination of photographs of faces following redundant presentations. *Journal of Experimental Child Psychology*, *18*, 98–106.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176.

DeCasper, A. J., & Prescott, P. A. (1984). Human newborns' perception of male voices: Preference, discrimination, and reinforcing value. *Developmental Psychobiology*, *17*, 481–491.

Diamond, R., & Carey, S. (1977). Developmental changes in the representation of faces. *Journal of Experimental Child Psychology*, *23*, 1–22.

Diesendruck, G. (2005). The principles of conventionality and contrast in word learning: An empirical examination. *Developmental Psychology*, *41*, 451–463.

Diesendruck, G., Hall, D. G., & Graham, S. (2006). Children's use of syntactic and pragmatic knowledge in the interpretation of novel adjectives. *Child Development*, *77*, 16–30.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*, 630–644.

Easterbrook, M., Kisilevsky, B., Hains, S., & Muir, D. (1999). Faceness or complexity: Evidence from newborn visual tracking of facelike stimuli. *Infant Behavior and Development*, *22*, 17–35.

Fagan, J. F. (1976). Infants' recognition of invariant features of faces. *Child Development*, *47*(3), 627–638.

Floccia, C., Nazzi, T., & Bertoncini, J. (2000). Unfamiliar voice discrimination for short stimuli in newborns. *Developmental Science*, *3*(3), 333–343.

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Journal of Developmental Psychology*, *28*(1), 99–108.

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

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*, B23–B34.

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, *53*, 310–344.

Heibeck, T. H., & Markman, E. M. (1987). Word learning in children—An examination of fast mapping. *Child Development*, *58*, 1021–1034.

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

Jaswal, V. K., & Markman, E. M. (2003). The relative strengths of indirect and direct word learning. *Developmental Psychology*, *39*(4), 745–760.

Jusczyk, P. W., Pisoni, D. B., & Mullennix, J. (1992). Effects of talker variability on speech perception by 2-month-old infants. *Cognition*, *43*, 252–291.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy. *Psychological Science*, *18*(12), 1084–1089.

Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H., Zhang, K., & Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*(3), 220–224.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, *66*(6), 1668–1679.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, *6*, 263–285.

Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology*, *27*, 153–165.

Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins and domain specificity. In M. R. Grunnar & M. P. Maratsos (Eds.), *Modularity and constraints in language and cognition: The Minnesota Symposium on Child Psychology* (pp. 59–101). Hillsdale, NJ: Erlbaum.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic vs. thematic relations. *Cognitive Psychology*, *16*, 1–27.

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

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

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, *385*, 813–815.

Maurer, D., & Barrera, M. (1981). Infants' perception of natural and distorted arrangements of a schematic face. *Child Development*, *52*, 196–202.

Mehler, J., Bertoncini, J., Barrière, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, *7*(5), 491–497.

Merriman, W. E., & Bowman, L. L. (1989). Mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, *54* (3–4, Serial number 220).

Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 201–203). Cambridge, England: Cambridge University Press.

Mervis, C. B., & Bertrand, J. (1994). Acquisition of the novel name-nameless category (N3C) principle. *Child Development*, *65*, 1646–1662.

Mills, M., & Melhuish, E. (1974). Recognition of mother's voice in early infancy. *Nature*, *252*, 123–124.

Pascalis, O., & de Schonen, S. (1994). Recognition in 3- to 4-day-old human neonates. *NeuroReport*, *5*, 1721–1724.

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and extension. *Infant Behavior and Development*, *18*, 79–85.

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences*, *102*(14), 5297–5300.

Patterson, M. L., & Werker, J. F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*, *81*(1), 93–115.

Proops, L., McComb, K., & Reby, D. (2009). Cross-modal individual recognition in domestic horses (Equus caballus). *Proceedings of the National Academy of Sciences*, *106*(3), 947–951.

Sandhofer, C., & Smith, L. B. (2007). Learning adjectives in the real world: How learning nouns impedes learning adjectives. *Language Learning and Development*, *3*(3), 233–267.

Smith, L. B., Jones, S. S., & Landau, B. (1992). Count nouns, adjectives, and perceptual properties in children's novel word interpretations. *Developmental Psychology*, *28*(2), 273–286.

Spelke, E. S., & Owsley, C. J. (1979). Intermodal exploration and knowledge in infancy. *Infant Behavior and Development*, *2*, 13–27.

Walton, G. E., Bower, N. J. A., & Bower, T. G. R. (1992). Recognition of familiar faces by newborns. *Infant Behavior and Development*, *15*, 265–269.

Waxman, S. R., & Booth, A. E. (2000). Principles that are invoked in the acquisition of words, but not facts. *Cognition*, *77*(2), B33–B43.

Wilkinson, K. M., & Mazzitelli, K. (2003). The effect of 'missing' information on children's retention of fast-mapped labels. *Journal of Child Language*, *30*, 47–73.

Appendix

Table 1
Trial order for Experiment 1

Trial Number	Target	Distractor	Left Face	Right Face	Voice
Training					
1			K1		K1
2				K1	K1
3	Known	Known	K1	K2	K1
4			K2		K2
5	Known	Known	K3	K2	K2
6				K3	K3
7	Known	Known	K2	K3	K3
8	Known	Known	K3	K2	K3
9	Known	Known	K1	K3	K3

Table 1 (Continued)

Trial Number	Target	Distractor	Left Face	Right Face	Voice
10	Known	Known	K3	K2	K2
11	Known	Known	K1	K2	K1
12	Known	Known	K1	K3	K1
13	Known	Known	K1	K2	K2
14	Known	Known	K3	K2	K3
15	Known	Known	K2	K1	K2
16	Known	Known	K2	K1	K1
Test					
17	Known	Novel	**N1**	K3	K3
18	Known	Known	K1	K3	K1
19	Novel	Known	K2	**N2**	**N1**
20	Known	Known	K1	K3	K3
21	Novel	Known	**N3**	K3	**N2**
22	Known	Novel	K1	**N4**	K1
23	Known	Known	K3	K1	K1
24	Known	Novel	K1	**N5**	K1
25	Known	Known	K1	K2	K2
26	Novel	Known	**N6**	K1	**N3**
27	Known	Known	K2	K3	K3
28	Novel	Known	K2	**N7**	**N4**
29	Known	Known	K2	K3	K2
30	Known	Novel	**N8**	K2	K2

Note. K denotes Known Faces and Voices and N denotes Novel Faces and Voices. Novel Faces and Voices appear in bold. The correct choice for each trial is highlighted in bold.

Table 2
Trial order for Experiment 2

Trial Number	Target	Distractor	Left Face	Right Face	Voice
Training					
1			K1		K1
2				K1	K1
3	Known	Known	K1	K2	K1
4			K2		K2
5	Known	Known	K3	K2	K2
6				K3	K3
7	Known	Known	K2	K3	K3
8	Known	Known	K3	K2	K3
9	Known	Known	K1	K3	K3
10	Known	Known	K3	K2	K2
11	Known	Known	K1	K2	K1
12	Known	Known	K1	K3	K1
13	Known	Known	K1	K2	K2
14	Known	Known	K3	K2	K3
15	Known	Known	K2	K1	K2
16	Known	Known	K2	K1	K1

Table 2 (Continued)

Trial Number	Target	Distractor	Left Face	Right Face	Voice
Additional training (Group 3 only)					
17	Known	Known	K2	K3	K2
18	Known	Known	K1	K2	K1
19	Known	Known	K1	K3	K3
20	Known	Known	K2	K1	K1
21	Known	Known	K2	K1	K2
22	Known	Known	K3	K2	K2
23	Known	Known	K3	K1	K3
24	Known	Known	K1	K2	K1
Test					
29	Known	Novel	**N1**	K3	K3
30	Known	Known	K1	K3	K1
31	Novel	Known	K2	**N2**	**N1**
32	Known	Known	K1	K3	K3
33	Novel	Known	**N3**	K3	**N2**
34	Known	Novel	K1	**N4**	K1
35	Known	Known	K3	K1	K1
36	Known	Novel	K1	**N5**	K1
37	Known	Known	K1	K2	K2
38	Novel	Known	**N6**	K1	**N3**
39	Known	Known	K2	K3	K3
40	Novel	Known	K2	**N7**	**N4**
41	Known	Known	K2	K3	K2
42	Known	Novel	**N8**	K2	K2

Note. K denotes Known Faces and Voices, and N denotes Novel Faces and Voices. Novel Faces and Voices appear in bold. The correct choice for each trial is highlighted in bold. Only the children in Group 3 received the additional training in Trials 17–24.

Table 3
Trial order for Experiment 3

Trial Number	Target	Distractor	Left Face	Right Face	Voice
Training					
1			K1		K1
2				K1	K1
3	Known	Known	K1	K2	K1
4				K2	K2
5	Known	Known	K3	K2	K2
6			K3		K3
7	Known	Known	K2	K3	K3
8	Known	Known	K3	K1	K3
9	Known	Known	K2	K1	K1
10	Known	Known	K2	K3	K2
11	Known	Known	K1	K3	K3
12	Known	Known	K2	K1	K2
13	Known	Known	K1	K2	K2

Table 3 (Continued)

Trial Number	Target	Distractor	Left Face	Right Face	Voice
14	Known	Known	K3	K1	K1
15	Known	Known	K3	K2	K3
16	Known	Known	K1	K3	K1
Test					
17	Known	Novel	N1	K3	K3
18	Known	Known	K1	K3	K1
19	Novel	Known	K2	**N2**	**N1**
20	Novel	Just-learned	**N2**	**N3**	**N1**
21	Known	Known	K2	K3	K3
22	Known	Known	K2	K1	K2
23	Novel	Known	**N4**	K3	**N2**
24	Just-learned	Novel	**N4**	**N5**	**N3**
25	Known	Novel	K1	**N6**	K1
26	Known	Known	K3	K1	K1
27	Novel	Known	**N7**	K1	**N4**
28	Novel	Just-learned	**N8**	**N7**	**N5**
29	Known	Known	K1	K2	K2
30	Known	Novel	K1	**N9**	K1
31	Novel	Known	K2	**N10**	**N6**
32	Just-learned	Novel	**N11**	**N10**	**N6**
33	Known	Known	K3	K2	K3
34	Known	Novel	**N12**	K2	K2

Note. K denotes Known Faces and Voices, and N denotes Novel Faces and Voices. Novel Faces and Voices appear in bold. The correct choice for each trial is highlighted in bold.