PAPER

WILEY | **Developmental Science**

Natural reference: A phylo- and ontogenetic perspective on the comprehension of iconic gestures and vocalizations

Manuel Bohn[1,2] | Josep Call[3,4] | Michael Tomasello[4,5]

[1]Department of Psychology, Stanford University, Stanford, California

[2]Leipzig Research Center, for Early Child Development, Leipzig University, Leipzig, Germany

[3]School of Psychology and Neuroscience, University of St. Andrews, St. Andrews, UK

[4]Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[5]Department of Psychology and Neuroscience, Duke University, Durham, North Carolina

Correspondence
Manuel Bohn, Department of Psychology, Stanford University, Stanford, CA.
Email: bohn@stanford.edu

Funding information
Marie Sklodowska-Curie Actions grant no 749229; Synergy Grant no 609819

Abstract

The recognition of iconic correspondence between signal and referent has been argued to bootstrap the acquisition and emergence of language. Here, we study the ontogeny, and to some extent the phylogeny, of the ability to spontaneously relate iconic signals, gestures, and/or vocalizations, to previous experience. Children at 18, 24, and 36 months of age (N = 216) and great apes (N = 13) interacted with two apparatuses, each comprising a distinct action and sound. Subsequently, an experimenter mimicked either the action, the sound, or both in combination to refer to one of the apparatuses. Experiments 1 and 2 found no spontaneous comprehension in great apes and in 18-month-old children. At 24 months of age, children were successful with a composite vocalization-gesture signal but not with either vocalization or gesture alone. At 36 months, children succeeded both with a composite vocalization-gesture signal and with gesture alone, but not with vocalization alone. In general, gestures were understood better compared to vocalizations. Experiment 4 showed that gestures were understood irrespective of how children learned about the corresponding action (through observation or self-experience). This pattern of results demonstrates that iconic signals can be a powerful way to establish reference in the absence of language, but they are not trivial for children to comprehend and not all iconic signals are created equal.

KEYWORDS
evolution, gesture, iconicity, language development, onomatopoeia, sound-symbolism

1 | INTRODUCTION

Language refers to things through the use of conventional signs or symbols. Perhaps more basic are "naturally referential" signals, that is, signals that denote their referent not by convention but by being somehow naturally connected to it, either as indices (e.g., pointing) or as icons (Peirce, 1932). In contrast to symbols, iconic signals resemble their referents, and thereby directly relate to interlocutors' experience of it. They can be created and understood on the spot to communicate a wide variety of meanings in the absence of pre-established conventions. This has led scholars to assume a substantial contribution of iconic signals to the development of language, both in ontogeny and phylogeny (Brentari & Goldin-Meadow, 2017; Cartmill, Beilock, & Goldin-Meadow, 2012; Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Donald, 1991; Fay, Ellison, & Garrod, 2014; Perniss & Vigliocco, 2014; Sterelny, 2017; Tomasello, 2008; Werner & Kaplan, 1963).

The role of iconic properties of signals for comprehension and learning has been studied in two domains, visual and auditory. Work on visual gesture comprehension with young children has mostly focused on whether children are better at learning gestures with iconic properties compared to arbitrary gestures as labels for objects. Children accept gestures as labels for objects at around 18 months (Namy, 2001; Namy & Waxman, 1998). Given communicative training, children identify iconic gestures as labels for familiar objects at 18 months (Tomasello, Striano, & Rochat, 1999). When learning

labels for novel objects, reliable comprehension of iconic gestures emerges around 26 months of age (Namy, 2008). However, there seems to be no advantage for iconic over arbitrary gestures early in development (Namy, Campbell, & Tomasello, 2004). Children appear to accept any kind of label if it is explicitly taught. Nevertheless, iconic, but not arbitrary gestures can potentially be created and understood on the spot, enabling successful communication in the absence of explicit prior instruction. For example, from 30 months onwards, children use iconic gestures as input to learn novel verbs (Goodrich & Hudson Kam, 2009; see also Mumford & Kita, 2014). In these studies, the gesture grounded the meaning of the word in prior experience by representing an action that happened earlier. In the studies presented here, we investigate the early development (starting at 18 months) of spontaneous comprehension of flexibly created, novel iconic gestures. Furthermore, we directly contrast iconic gestures with iconic vocalizations.

Iconicity in vocal signals has been studied in the form of sound-symbolism and onomatopoeia. The equivalents to visual iconic gestures in vocal signals are onomatopoeia, words, or vocalizations that mimic some property of their referent. A classic example would be saying "woof woof" to refer to a dog. Recent work found that onomatopoetic words are more frequent across languages than previously thought (Monaghan, Shillcock, Christiansen, & Kirby, 2014) and are acquired earlier in development (Perry, Perlman, & Lupyan, 2015; Perry, Perlman, Winter, Massaro, & Lupyan, 2017). However, onomatopoetic properties co-vary with other variables known to support word learning, for example mean pitch, word duration, and repetition (Laing, Vihman, & Keren-Portnoy, 2017). Furthermore, the advantage of onomatopoetic over conventional words in infants' ability to match words to their referents is mediated by infants' familiarity with the form (Laing, 2017). Thus, whether or not onomatopoetic words are acquired earlier because children spontaneously recognize the iconic correspondence between word and referent is somewhat unclear.

Studies on sound-symbolism exploit the idea of cross-modal iconic correspondence between visual and auditory stimuli. For example, a round shape (a circle) bears resemblance to a vowel rich word ("bouba"). A number of studies report evidence for cross-modal matching already in very young infants (e.g., Asano et al., 2015; Ozturk, Krehm, & Vouloumanos, 2013). Furthermore, sound-symbolic properties have been shown to facilitate word learning in older children (e.g., Imai, Kita, Nagumo, & Okada, 2008; Maurer, Pathman, & Mondloch, 2006; Nygaard, Cook, & Namy, 2009). As a consequence, Imai and Kita (2014) suggested that the ability to recognize cross-modal matching is biologically endowed, allowing sound-symbolism to bootstrap language acquisition. A recent meta-analysis (Fort et al., 2018) paints a slightly less enthusiastic picture, finding early cross-modal matching only for round shape-sound pairings and increasing effect sizes with age. Furthermore, sound-symbolism has been studied in terms of online cross-modal matching, relating shapes to sounds in the here and now. Part of the importance of iconic signals lies in the fact that they allow flexible communication about absent or transient aspects of experience.

RESEARCH HIGHLIGHTS

- Iconic signals ground reference in interlocutors' experience, thereby supporting language development in ontogeny and potentially also in phylogeny.
- 24- and 36-month-olds, but not 18-month-olds and not great apes, spontaneously identified the referent of an iconic signal.
- Iconic gestures are understood better compared to iconic vocalizations.
- Not all iconic signals are created equal and their comprehension develops substantially in the third year of live.

Investigating the evolutionary origins of the ability to match iconic signals to aspects of previous experience complements developmental lines of research. Studying humans' closest living relatives, the great apes, allows us to make inferences about the last common ancestor and thereby approximates whether the cognitive processes enabling the ability in question likely emerged before or after the lineages leading to humans and the other great apes separated. Reconstructing the evolutionary history of iconic signal comprehension is particularly relevant because a number of theoretical accounts on language evolution suggest some sort of intermediate iconic proto-language, vocal, or signed (see e.g., Donald, 1991; Fitch, 2010; Tomasello, 2008; Zlatev, Persson, & Gärdenfors, 2005). Finding that great apes understand iconic gestures and vocalizations would make it plausible that communication based on iconic signals might have scaffolded the emergence of the cognitive architecture underlying human language.

Research on iconicity in great ape communication is fairly sparse compared to research with children. With respect to visual gesture, larger observational studies do not find evidence for a widespread use of such iconic gestures in great apes in the wild or in captivity (Call & Tomasello, 2007; Genty, Breuer, Hobaiter, & Byrne, 2009; Graham, Furuichi, & Byrne, 2017; Hobaiter & Byrne, 2011). However, there are occasional reports of great apes using seemingly iconic gestures (Douglas & Moscovice, 2015; Genty & Zuberbühler, 2014; Russon & Andrews, 2010). Based on observations alone, it remains unclear whether the iconicity in these gestures lies in the eye of the (human) beholder or plays a role in them being used and understood. The only experimental study so far found no evidence for spontaneous comprehension in chimpanzees (Bohn, Call, & Tomasello, 2016). In this study, participants first learned how to retrieve a reward from two apparatuses together with an experimenter. Later, the experimenter used a gesture mimicking the action performed at one of the apparatuses to inform the participant that this would be the one yielding a reward next. Results showed no signs of spontaneous (within first 24 trials) gesture comprehension in chimpanzees. However, the same study found that iconic gestures are learned faster compared to arbitrary gestures, suggesting some

recognition of the iconicity involved. To our knowledge, there have been no reports or experimental studies on iconic vocalizations in any of the great apes.

Here, we follow up on the results found by Bohn et al. (2016), extending the type of signals studied to vocalizations as well as gestures. More specifically, we seek to integrate previous work into a unified design that allows us to directly study and compare different types of iconic signals. Our current study focused on spontaneous comprehension, that is, whether participants were able to identify a referent by relating a novel iconic signal to aspects of a previously shared episode. The setup varied slightly between groups but had the same overall structure. Participants interacted with an experimenter around two apparatuses, each involving a distinct action and producing a distinct sound upon operation. These properties were not highlighted during the initial interaction. Later on, the experimenter referred to one of the apparatuses by mimicking its action, its sound or both. Participants responded by approaching or indicating one of the apparatuses. In all groups, we first looked at combined signals, comprising gesture and vocalization, and, in case of comprehension, we studied gesture and vocalization separately. That is, we assumed an additive effect of the iconic information provided in gesture and vocalization. In experiment 1, we studied the great apes. Experiments 2 and 3 traced the development of the ability in question in children between 18 and 36 months of age. Experiment 4 focused on two alternative ways in which gestures can be understood.

2 | EXPERIMENT 1

A previous study found no spontaneous comprehension of iconic gestures by chimpanzees (Bohn et al., 2016). In experiment 1, we extended this earlier work in four ways. First, we introduced a communicative training in which the participants learned to use a hand gesture (pointing) from a human to decide between two alternatives. This training was designed to ensure that participants understood the structure of the task at hand and knew that the experimenter provided them with useful information. Second, we enriched the iconic signal by adding a vocalization. Participants could rely on the similarity between gesture and action performed at the apparatus as well as between a vocalization and the sound emitted by the apparatus. Third, we modified the setup in line with suggestions to improve

apes' performance in object choice tasks more generally (Mulcahy & Hedge, 2012). Finally, we tested two more great ape species (bonobos and orangutans) in addition to chimpanzees, thereby diversifying the sample.

2.1 | Methods

2.1.1 | Participants

Thirteen great apes (mean = 22.00 years, range = 7.00–42.00) participated in the study, including five bonobos (*Pan paniscus*, four females), four chimpanzees (*Pan troglodytes*, three females) and four orangutans (*Pongo abelii*, three females). One of the chimpanzees had participated in a previous study (Bohn et al., 2016). Two additional chimpanzees were initially part of the sample but had to be excluded because they lost interest in the study or were unable to operate the apparatuses. All apes were housed at the Wolfgang Köhler Primate Research Center at Zoo Leipzig, Germany. The sample size for apes was determined by the number of apes that were available for testing. Research was noninvasive and strictly adhered to the legal requirements of Germany. Animal husbandry and research complied with the European Association of Zoos and Aquaria (EAZA) Minimum Standards for the Accommodation and Care of Animals in Zoos and Aquaria and the World Association of Zoos and Aquariums (WAZA) Ethical Guidelines for the Conduct of Research on Animals by Zoos and Aquarium. Participation was voluntary, all food was given in addition to the daily diet, and water was available ad libitum throughout the study. Data were collected between November 2015 and June 2016.

2.1.2 | Setup

The left panel in Figure 1 shows a schematic overview of the setup. Apes were tested in a subsection of their familiar sleeping rooms comprising two adjacent cages with a small booth between them. The walls of the booth either consisted of large windows or transparent panels. The door between the cages remained open throughout a session. Plexiglas panels (69 × 48 cm) with an opening at the bottom (8.5 × 2.5 cm) were installed left and right to the booth. The test apparatuses were attached to these windows. The distance between apparatuses was 130 cm. In the beginning of a trial, the experimenter
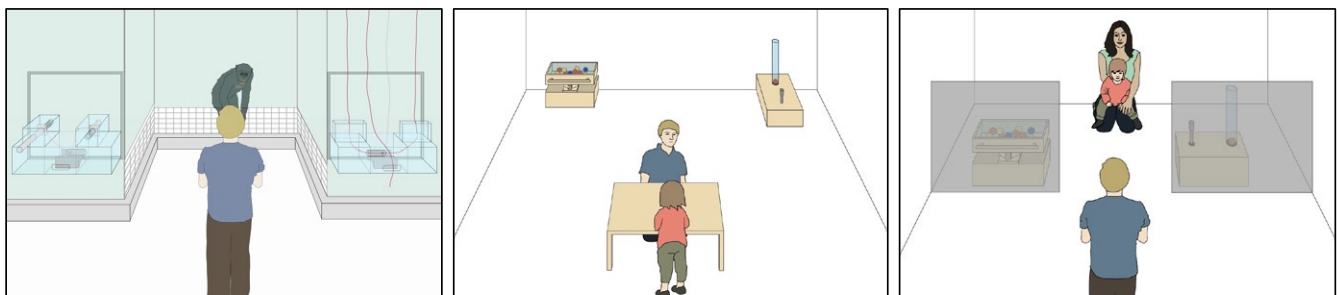

FIGURE 1 Schematic overview of the setup in experiment 1 (left), experiment 2 (middle), and experiments 3 and 4 (right). In each drawing, the position of the apparatuses, participant, and the experimenter(s) corresponds to the configuration at test

stood between the apparatuses, ~150 cm away from the participant. Apparatuses were the same as in Bohn et al. (2016). Each consisted of a rectangular box (50 × 25 × 17.5 cm) and two releasing mechanisms attached to the top. A reward was released to the participant through a corresponding hole in the bottom of the box when participant and experimenter operated the mechanism simultaneously. Retrieving the reward alone was not possible. The actions required to operate the mechanisms were: pulling down a rope, pushing in a lever, turning a crank, and moving a lever from right to left. These actions were turned into gestures by performing the same bodily movements detached from the apparatus. Each box also contained a speaker and a MP3-player. Contingent on operating the releasing mechanism, the speaker played a distinct sound (either a high-pitched bell or two alternating low-pitched notes), otherwise they remained silent. During the test, the experimenter imitated these sounds vocally. Drawings of the releasing mechanisms and sound files can be found in the supplementary material. A juice dispenser was used to release small amounts of diluted grape juice to the participant when located in the middle of the booth. Thereby, the experimenter could center the participant in the beginning of a trial. All trials were videotaped using a wide-angle camera installed above the experimenter, providing a full view of the setup. Participants were tested individually and received banana pellets as rewards.

2.1.3 | Design and procedure

For each participant, we randomly assigned a releasing mechanism and a sound to the left and the right box. The only constraint was that the chimpanzee who participated in the Bohn et al. (2016) study received different releasing mechanisms compared to that earlier study. We created matched pairs by age and species and assigned members of a pair randomly to one of two groups. One group (seven individuals) started directly with the training phase and received the test phase afterwards. The other group (six individuals) first received the test, then the training and then again, the test. Starting with the test right away in one group allowed us to re-assess spontaneous comprehension without training as in Bohn et al. (2016). The training had two phases (details below).

Training and test sessions each comprised of 12 trials. The order in which apparatuses were indicated from trial to trial was randomly determined with the constraints that both sides were indicated equally often, and the same side was never indicated more than two times in a row. Participants received two sessions (24 trials) in test and a maximum of 16 sessions in each training phase. The learning criterion for each training phase was 18/24 trials correct within two sessions. Participants who did not reach the training criterion during the second training phase nevertheless proceeded to the test. As described above, apes started with the combined condition. The general procedure was the same for training and test. The experimenter initiated the trial by centering the participant in the middle of the booth using the juice dispenser. Then, the experimenter attracted the participant's attention by calling her name and started signaling. While signaling, the experimenter faced the participant,

trying to make eye contact. The signal varied depending on the condition. In the first training phase, the experimenter pointed with the index finger by moving his arm contralateral across his chest while also turning the head and looking at the apparatus. In the second phase, pointing remained the same while head-turning and gazing were omitted. In the test phase, the experimenter mimicked the action and the sound of the corresponding apparatus. Gestures and vocalizations were produced simultaneously in blocks of four repetitions and continued until the participant approached one of the apparatuses or 2 min had elapsed. If the participant started moving before the first signal was emitted, the experimenter centered her again. Regardless of the condition, if the choice was correct, the experimenter followed suit and together they operated the apparatus, dispensing a reward to the participant. If the choice was incorrect, the experimenter approached the correct apparatus during training (no reward was dispensed) but followed the participant during the test. That is, reinforcement was differential in training and non-differential during test.

2.1.4 | Coding and analysis

We coded the correct choice, that is, whether or not the participant approached the apparatus indicated by the experimenter's signal (gesture and vocalization). For apes and all subsequent studies with children, we used generalized linear mixed models (GLMM) with binomial error structure to analyze the data. All models included a maximal random effect structure with random intercepts for participant and random slopes for trial. To assess whether the inclusion of predictors improved model fit relative to a null model, we used likelihood ratio tests (Dobson & Barnett, 2008). Performance within a given group was compared against chance (50% correct) by fitting a GLMM with centered predictors and testing whether the intercept differed from zero. This approach allowed us to account for unequal numbers of trials for participants (more relevant for children than for apes). All models were fitted in R (R Core Team, 2017) using the function *glmer* of the R-package lme4 (Bates, Mächler, Bolker, & Walker, 2015). Code and data for all models can be found in the supplementary material. The model for apes included predictors for group (start with test or training), training (reached criterion in second training phase), and trial. Given the small number of individuals, we did not analyze species separately. On an individual level, performance in test was considered above chance if 18 or more trials (out of 24) were correct (binomial test, $p < 0.05$). A second coder, blind to the purpose of the experiment coded 25% of randomly selected test trials. Agreement between coders was 100%.

2.2 | Results

Looking at spontaneous comprehension in the group that started with the test, we found that performance did not significantly differ from chance on an individual or group level (mean proportion correct = 0.47, β = −0.11, p = 0.505, 95% CI = [−0.44: 0.20]). Next, we looked at the effect of communicative training. All but one

chimpanzee reached learning criterion in the first training phase. Six apes (4 bonobos, 1 chimpanzee, and 1 orangutan) also reached the criterion in the second training phase. Apes who succeeded in both the training phases needed an average of 143.83 trials in total (range: 48–204) to do so. Figure 2 shows the performance across all participants in the test phase following training. Again, neither on the group (mean correct: 0.50, β = 0, p = 1, 95% CI = [−0.24: 0.24]) nor on an individual level did performance differ from chance. Group and training success had no influence on performance (χ^2(3) = 0.43, p = 0.930), suggesting that training success did not improve the performance.

2.3 | Discussion

Experiment 1 replicates the findings from Bohn et al. (2016) in a largely independent sample. None of the measures (training, enriched signal, setup, diverse sample) introduced to improve apes' performance turned out to be fruitful. All except one participant showed a general understanding of the task by passing one or both training stages. Nevertheless, switching from one hand gesture to another (enriched by vocalization) led to a breakdown in performance by all participants. Together with Bohn et al. (2016), 23 different apes have now been tested with none of them showing signs for spontaneous comprehension of iconic signals. However, all these apes were housed in a zoo setting. It is unclear if apes with different housing/rearing conditions would perform differently. Another caveat of these studies is that the gesturer was always a

human instead of a conspecific. Future studies should definitely try to alleviate these shortcomings. However, apes have been shown to be able to comprehend iconic gestures performed by a human after a longer learning period (Bohn et al., 2016; see also Buttelmann, Carpenter, Call, & Tomasello, 2013) and do not generally perform better when tested with a conspecific model (Boesch, 2007). Taken together, this research suggests that great apes have difficulties with spontaneously inferring the referent of a novel, representational communicative act.

3 | EXPERIMENT 2

In the following experiments, we explored the developmental origins of the ability to spontaneously comprehend iconic signals based on shared experience in human children. Whenever reasonable, we also contrasted the modalities in which the signal was presented. In experiment 2, we tested 18-month-olds. This is the youngest age at which children have been shown to comprehend iconic gestures (Tomasello et al., 1999). While the general structure of the experiment was the same in all the remaining experiments with children, there were some differences in setup and procedure for 18-month-olds compared to 24- and 36-month-olds. The data for 18-month-olds were collected first and, after finding negative results in the combined condition, the procedure was refined to make the choice situation and overall structure more plausible for older children. This prevents a direct comparison between
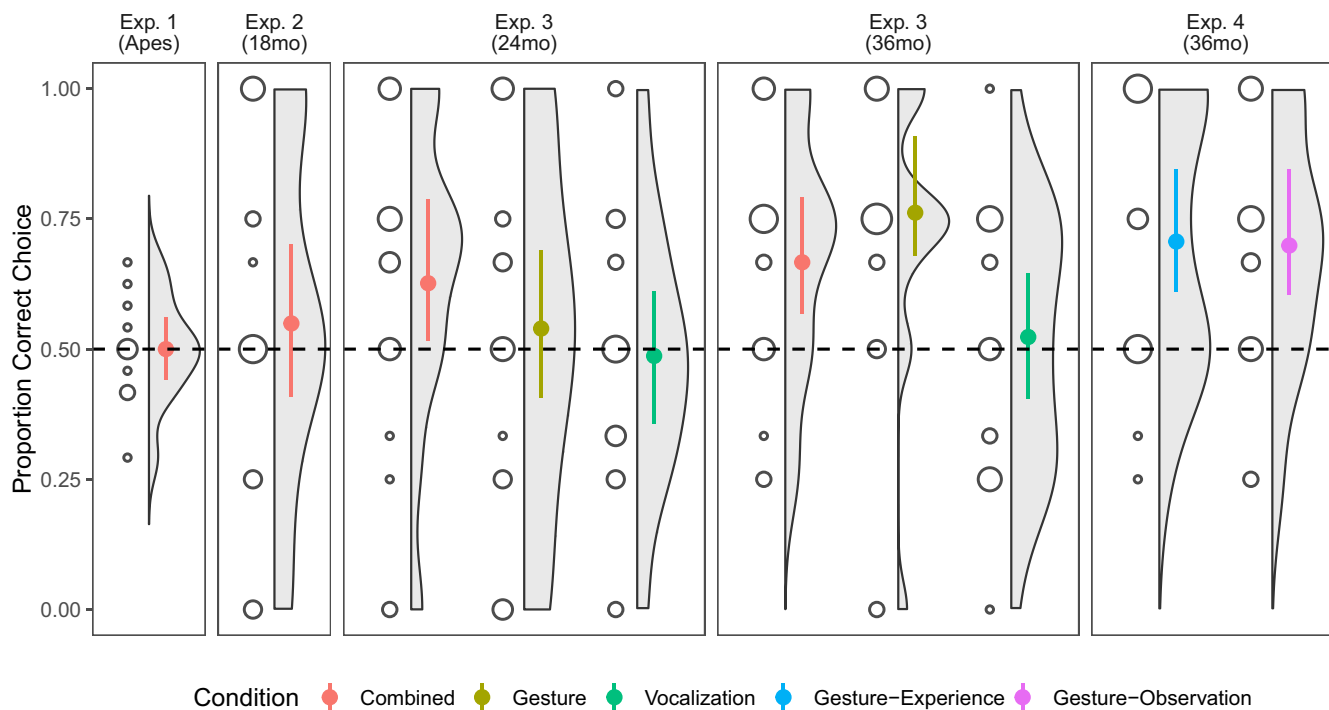

FIGURE 2 Proportion of correct choices by participants for experiments 1–4. The dashed line indicates performance expected by chance. Error bars represent 95% confidence intervals (see also Table 1). Differently sized circles represent the number of participants with a certain level of performance. Smoothed density distributions are shown in grey

Age group	Condition	β	CI 2.5%	CI 97.5%	p	T1[a]	T1: p[b]
18 months	Combined	0.19	−0.37	0.86	0.49	15	0.31
24 months	Combined	0.56	0.07	1.31	0.04	17	0.06
	Gesture	0.16	−0.37	0.80	0.50	13	0.84
	Sound	−0.05	−0.59	0.45	0.82	16	0.15
36 months	Combined	0.69	0.27	1.33	<0.01	17	0.06
	Gesture	1.16	0.75	2.29	<0.01	20	<0.01
	Sound	0.10	−0.39	0.60	0.65	12	1.00
	Experience	0.92	0.45	1.70	<0.01	21	<0.01
	Observation	0.89	0.42	1.71	<0.01	20	<0.01

Notes. All estimates based on GLMMs with the following structure: correct ~1 + (trial|id). Estimates different from 0 indicate performance different from chance (50% correct). CI 2.5%/97.5% = Lower and upper bound of 95% confidence intervals.
[a]T1 = Number of children choosing correct (out of 24) in trial 1. [b]T1: p = p-values for trial 1 based on two-tailed binomial tests.

age groups and so we present and analyze the data as part of a separate experiment.

3.1 | Methods

3.1.1 | Participants

The sample sizes for the following experiments with children were based on a simulation study assuming a medium, additive effect of gesture and vocalization, and a small effect of age with no interaction between age and condition. We simulated 1,000 datasets with different sample sizes, different levels of within sample variation, and different numbers of trials per participant. Power was assessed by running a GLMM for each simulation and aggregating over the model outputs. For more details and access to the corresponding code, please contact the first author. Based on these simulations, sufficient power to detect the assumed effects would be achieved with a sample size of 24 children per cell. In experiment 2, we therefore tested 24 children (12 boys) with a mean age of 18.39 months (range = 17.82–18.97). Five additional children started participating but were excluded because they became uncomfortable with the testing situation (3), their parents interfered (1), or one apparatus broke (1). Children lived in an ethnically homogeneous, mid-sized German city with approximately 550,000 inhabitants and median household income of €1,767 per month in 2017 (Stadt Leipzig, 2018). Participants were mostly mono-lingual and had mixed socioeconomic background. They were recruited from a database of children whose parents volunteered to take part in studies on child development. Data were collected between October and December 2015.

3.1.2 | Setup

The middle panel of Figure 1 shows a schematic overview of the setup. Children were tested in a large, rectangular room within a child laboratory. The two apparatuses were located at the short wall on the far end of the room 180 cm apart from one another.

A small table stood in the middle of the room, 370 cm away from the apparatuses. One apparatus consisted of a cuboid wooden box (52 × 25 × 30 cm) with a Plexiglas tube (height 53 cm, diameter 5 cm) and a lever (height 22 cm) sticking out on top. Inside the tube was a red ping pong ball. Pushing down the lever pushed the ball upwards, making it jump when pushed with sufficient force. When pushing down, the lever also hit a bell located inside the box, producing a single ring. The experimenter always pushed down the lever with her right hand, palm facing down. The gesture derived from this apparatus involved the same movement without the object. The vocalization involved vocally imitating the ring of the bell. The second apparatus was composed of a lower wooden box (38 × 33 × 37 cm) with a vertical stick on top. Mounted on top of this stick was a second wooden box (50 × 37 × 11 cm). Like a seesaw, the upper box could be tilted left and right. It had a Plexiglas window on top, granting view to 21 colored balls inside it. Tilting the apparatus made the balls roll from one side to the other. A bone-shaped squeaky toy was attached to the bottom of the box. When tilting the box left or right, the toy produced a squeaking sound. A vertical handle was attached to the long side of the box to facilitate tilting it back and forth. The gesture corresponding to this apparatus was pretending to hold the handle with both hands and tilting the box from left to right. The vocalization was a vocal imitation of the squeaking sound. In between trials, the experimenter used a stacking rings toy to center the child at the table. All trials were videotaped.

3.1.3 | Design and procedure

Children received a single session comprising four trials with each apparatus being indicated twice. Across participants, we counterbalanced the order in which the apparatuses were indicated, the location they were positioned in (left or right) and which apparatus was introduced first. Like all other groups, 18-month-olds also started with the combined condition. Children arrived in the child laboratory in a play room where they met the experimenter. After

a short period of free play, the experimenter, child, and parent entered the test room. Parents were asked to take a seat in a corner and to stay passive throughout the experiment. The experimenter first introduced the child to the stacking ring toys at the central table. After playing there for a while, the experimenter introduced the child to the first apparatus. The experimenter first demonstrated to the child how the apparatus functioned by playing with it and then encouraged the child to play herself. After playing with the first apparatus, the child and the experimenter shortly returned to the central table for a round of stacking rings and then approached the second apparatus. The introduction to each apparatus lasted approximately 20 s with each child performing the action at least three times. Importantly, the experimenter never commented on the way the apparatus moved or sounded and never used iconic gestures to instruct the child. Following the second introduction, the child and the experimenter again returned to the table. The experimenter knelt down with his back toward the apparatuses and encouraged the child to go to the opposite end of the table. After another round of stacking rings, the first test trial started. The experimenter removed the toys from the table, caught the child's attention and said: "Let's play with the [iconic signal] again" (German: "Lass uns nochmal mit dem [iconic signal] spielen"). Within the utterance, the iconic signal was repeated four times. If the child did not point to or approach one of the apparatuses, this utterance was repeated. Some children answered with an equivalent of "Ok" but did not approach or indicate one of the apparatuses. In those cases, the experimenter said, "Let's go, you first" (German: "Los geht's, Du zuerst"). If the child did not approach or indicate one of the apparatuses within 2 min, the experimenter turned around and approached the correct one and encouraged the child to follow. Whenever the child approached or otherwise indicated an apparatus, the experimenter approached it (non-differential reinforcement). Experimenter and child briefly played with the indicated apparatus, then played with the other apparatus and finally returned to the table where the next trial began. Because they did not perform above chance in the combined condition, 18-month-olds were not tested in the gesture or vocalization condition.

3.1.4 | Coding and analysis

We coded whether or not the child pointed to or approached the apparatus indicated by the experimenter's signal. Only trials in which the participant made a choice were considered for the analysis. Reliability coding of 25% of randomly selected trials yielded an agreement of 100% between coders. The model for 18-month-olds had the same general random effect structure (see above) and no additional predictors.

3.2 | Results

Table 1 shows detailed results of the model as well as performance in the first trial. Children in this experiment did not choose the correct apparatus above chance (see also Figure 2).

3.3 | Discussion

In contrast to Tomasello et al. (1999), we found no spontaneous comprehension of iconic signals. In this earlier study, children were given a communicative training before the test and were tested on gestures for familiar objects (e.g., a hammer). The latter makes the iconic interpretation of the gestures somehow questionable as children might have learned the corresponding gestures as conventional ones. It is, however, unclear whether a communicative training, like the one provided for apes, would boost 18-month-olds' performance.

4 | EXPERIMENT 3

In experiment 3, we followed the developmental pathway of the ability in question by studying slightly older children: 24- and 36-month-olds. The procedure was slightly different as well. The reasons for the adjustments were the following: We wanted to get closer to the idea of making spontaneous reference to an earlier shared episode by introducing a longer delay between exposure and test. Furthermore, to make the choice situation clearer, we asked children a question instead of prompting them to approach one of the apparatuses.

4.1 | Methods

4.1.1 | Participants

For each combination of condition and age group, we tested 24 children. In total, 144 children participated in this study, 72 two-year-olds (36 boys, mean age 24.02 months (range = 23.50–24.52) and 72 three-year-olds (36 boys, mean age 35.68 months (range = 32.91–39.25). In addition, 20 two-year-olds and four three-year-olds started with the study but had to be excluded because they were uncomfortable with the test situation (20), parents interfered (2), or one of the apparatuses malfunctioned (2). Children lived in the same city described above. Two-year-olds were again recruited from a database, while three-year-olds were recruited from local kindergartens. Data for two-year-olds were collected between July 2016 and March 2017, and for three-year-olds between March and May 2017.

4.1.2 | Setup

The right panel of Figure 1 shows a schematic overview of the setup in the test situation. Two-year-olds were tested in a room within a child laboratory and three-year-olds in a separate room in their kindergarten. For two-year-olds, parents were present in the room but were asked to remain passive. The apparatuses were the same as in experiment 2. Two wooden occluders (85 × 85 cm) were placed 145 cm apart from one another. In the exposure phase, the apparatuses were placed next to one another, in between the two occluders. During the test phase, they were put behind them. A pillow was

placed in the back of the room, 110 cm away from the apparatuses, marking the child's position during test (L2). From this position, the child had full view of the apparatuses, even when placed behind the occluders. A second pillow was placed in the front of the room, 180 cm away from the apparatuses, marking the experimenter's position at test (L1). From this position, the experimenter could not see the apparatuses when placed behind the occluders. The iconic signals were the same as in experiment 2.

4.1.3 | Design and procedure

Number of trials and counterbalancing were identical to experiment 2. Children entered the room together with two experimenters (E1 and E2). The apparatuses were initially placed next to the occluders. After a short play period at L1, E1 introduced the child to the two apparatuses in the same way as in experiment 2. After the introduction, E1 pretended to have forgotten something outside and left the room. Next, E2 took out a ball and asked the child if she wanted to play. In order to make room for playing with the ball, E2 moved the apparatuses out of the space between the occluders behind them. Then, E2 and the child tossed the ball back and forth. After 90 s, E1 knocked on the door, announcing his return. Thereupon, E2 led the child to L2, and asked E1 to enter. E1 entered the room, positioned himself at L1, looked puzzled from left to right, and asked the child, "Uhm…[child's name], where is the [iconic signal]"? (German: "Ähm…[child's name], wo ist denn das [iconic signal]"?). Repetitions of this utterance and the choice phase were the same as in experiment 2. If the child did not respond within 2 min, E2 pointed out the correct apparatus. Whenever the child indicated one of the apparatuses, E1 moved forward and looked behind the occluder. In case it was the one corresponding to the signal, he cheered, moved it from behind the occluder and encouraged the child to play with him. In case it was the wrong apparatus, he looked over to the other side and said, "Oh no, this is the one I was looking for", again followed by moving the apparatus and a play period. That is, in contrast to experiment 2, children were differentially reinforced. After also moving and playing with the second apparatus, E1 again left the room and the next trial began. After the second trial, E2 switched the position of the two apparatuses. The reason for doing so was to keep E1's asking and searching plausible. Two- and three-year-olds performed above chance in the combined condition and were therefore also tested in the gesture and vocalization only condition.

4.1.4 | Coding and analysis

Coding was identical to experiment 2. Agreement between reliability coders (25% of trials) was 100% for two- and three-year-olds. First, we compared performance in each condition and age group combination to chance using models akin to the one used in experiment 2. Next, we compared performance across conditions and age groups in a model comprising age group, condition, and their interaction as fixed effects.

4.2 | Results

Table 1 shows detailed results for the models comparing performance against chance as well as first trial performance (see also Figure 2). Two-year-olds performed above chance in the combined condition while three-year-olds did so in the combined and in the gesture condition. In each of these cases, overall performance was also reflected in the first trial (all p near or below 0.05). When comparing performance across age groups and conditions, we found that including these predictors improved model fit ($\chi^2(5) = 18.88$, $p < 0.001$). There was no significant interaction between age and condition ($\chi^2(2) = 4.29$, $p = 0.120$). We had no hypothesis about such an interaction (see power simulation) and therefore removed it to evaluate the main effects. We found a main effect of condition ($\chi^2(2) = 9.80$, $p = 0.010$) and age ($\chi^2(1) = 5.05$, $p = 0.020$). Children performed better in the combined and the gesture condition compared to the vocalization condition (vocalization as reference level—combined: $\beta = 0.60$, $p = 0.008$, 95% CI = [0.15: 1.10]; gesture: $\beta = 0.63$, $p = 0.006$, 95% CI = [0.19: 1.13]). Furthermore, three-year-olds outperformed two-year-olds ($\beta = 0.42$, 95% CI = [0.08: 0.82]).

4.3 | Discussion

We found evidence for spontaneous comprehension of iconic signals at 24 months. Around the same age, children also start to produce iconic gestures (Behne, Carpenter, & Tomasello, 2014). Overall, 36-month-olds performed better compared to the younger age group. Furthermore, children performed better in the combined and the gesture condition compared to the vocalization condition. At odds with the spontaneous comprehension interpretation of the onomatopoetic/sound-symbolic advantage in early word learning (Imai & Kita, 2014), we did not find spontaneous comprehension of vocalizations alone in the two age groups. Vocalization might have enriched the gesture for 24-month-olds as they performed above chance in the combined but not in the gesture condition. The developmental pattern found here also mirrors children's understanding of representations in gestures and pictures or scale models more generally (DeLoache, 1987, 2000; Novack, Goldin-Meadow, & Woodward, 2015; Tolar, Lederberg, Gokhale, & Tomasello, 2007). For example, when pictures are used to inform the child of the location of a hidden toy, 30- but not 24-month-olds were able to retrieve it (DeLoache & Burns, 1994). Interestingly, and reminiscent of our finding for combined signals in 24-month-olds, the more iconic (i.e., more realistic) pictures are, the earlier children are able to transfer labels from pictures to real objects (Ganea, Allen, Butler, Carey, & DeLoache, 2009).

Why was gesture easier? On a closer look, gestures could be understood in two non-exclusive ways. On the one hand, the gesture visually resembled the action performed by the experimenter at the apparatus. Seeing the gesture therefore activates an episodic memory trace of the play episode at the indicated apparatus and singled

it out as the intended referent. On the other hand, seeing the gesture activates the child's motor representation of the action she performed at this apparatus (mediated by the mirror neuron system, see e.g., Andric et al., 2013; Cook, Bird, Catmur, Press, & Heyes, 2014; Montgomery, Isenberg, & Haxby, 2007; Villarreal et al., 2008), with similar consequences. Vocalization, on the other hand, only shares a perceptual similarity with the earlier episode but does not resonate in motor experience. To test the idea that gesture is easier because it relates to multiple aspects of experience, we conducted a fourth experiment.

5 | EXPERIMENT 4

In experiment 4, we followed up on the idea that gestures are easier understood because they provide more referential information. We isolated the two ways in which gestures could be understood by changing the way that children learned about the apparatuses. In one group, children only saw the experimenter act on the apparatus (observation), in the other condition, children never saw anybody act on the apparatus but only did so themselves (experience). The following predictions can be made: If gestures in experiment 3 contained more referential information, performance in both the conditions in experiment 4 should resemble the vocalization condition of experiment 3. This is because each condition, like vocalizations, only relates to one aspect of the previous experience. If gesture is primarily understood because it resonates in motor experience, performance should be better in the experience condition while performance in the observation condition should be at chance. If gesture in general constitutes something like a privileged modality compared to vocalization, the two conditions in experiment 4 should differ from the vocalization but not the gesture condition in experiment 3.

5.1 | Methods

5.1.1 | Participants

A total of 48 three-year-olds (24 boys, mean age 36.29 months, range = 32.77–38.92) participated in the study. Children were again recruited from local kindergartens. Thirteen additional children started participating but had to be excluded because they were uncomfortable with the test situation (10), experimenter error (2), or apparatus malfunctioning (1). Data were collected between September and November 2017.

5.1.2 | Setup

The same setup as in experiment 3 was used.

5.1.3 | Design and procedure

Design and procedure were identical to the gesture condition of experiment 3, except for changes during the exposure phase. In the observation condition, only E1 played with the apparatuses while the child remained in L1 with E2. Children were encouraged to watch E1 play, however the actions performed by E1 were never imitated or labeled. In the experience condition, only the child played with the apparatuses. E's encouraged the child to discover the functionality of the apparatuses on her own, occasionally directing her attention to the relevant parts using pointing gestures. Importantly, neither E performed nor labeled the actions required to play with the apparatus. During test, only gestures were used as iconic signals.

5.1.4 | Coding and analysis

Coding was identical to experiments 1, 2 and 3. Reliability coding of 25% of trials yielded an agreement of 95.74% (κ = 0.91). In addition to models comparing performance to chance, we ran a model comparing the gesture and vocalization conditions of experiment 3 to the two conditions of experiment 4 (vocalization as reference level) and a model comparing all gesture only conditions.

5.2 | Results

Children selected the correct apparatus above chance in both gesture conditions (see Table 1 and Figure 2). Performance in all gesture conditions was better compared to the vocalization condition (main effect of condition: χ^2(3) = 12.22, p = 0.007; gesture: β = 1.07, p = 0.001, 95% CI = [0.47: 1.86]; observation: β = 0.75, p = 0.017, 95% CI = [0.16: 1.46]; experience: β = 0.79, p = 0.013, 95% CI = [0.16: 1.48]). There was no effect of condition in a model excluding the vocalization condition (χ^2(2) = 1.04, p = 0.600), suggesting that removing one sort referential information from gestures does not impair comprehension.

5.3 | Discussion

Children understood iconic gestures in this experiment regardless of how they were introduced to the corresponding action. This finding suggests that iconic information presented in the gestural modality seems to be more accessible to children early in development compared to vocalization. There are, however, some limitations to this interpretation, which we will discuss below.

6 | GENERAL DISCUSSION

We investigated the phylo- and ontogenetic origins of spontaneous comprehension of iconic gestures and vocalizations. We found no signs for such comprehension in great apes and in the youngest human age group, 18-month-olds. At 24 months of age, children showed comprehension of iconic signals mimicking both the action and the sound associated with an apparatus. Thirty-six-month-olds showed an overall higher performance and robust comprehension of combined signals as well as gestures, but not vocalization. Overall,

gestures were more readily understood compared to sounds, suggesting that information presented in this modality is more readily available to children early in development. A follow-up experiment showed that this was not due to gesture relating to multiple aspects of experience. Taken together, this research suggests that while recognizing iconic correspondence between signal and referent may boost communicative development from 2 years onwards, it is unlikely to contribute to the early emergence of referential communication.

Previous work with apes suggested that they have problems with spontaneously inferring the referent of a novel iconic gesture (Bohn et al., 2016). Here, we find similar results, despite the fact that we trained participants to use a different informative hand gesture provided by the same human experimenter beforehand and also enriched the signal and the setup in various ways compared to the earlier study. As mentioned in the discussion of study 1, there are several limitations to our approach that should be addressed in future research. Importantly, however, the work by Bohn et al. (2016) also showed that great apes learn (over a longer period of time) iconic gestures faster compared to arbitrary gestures, suggesting that the iconicity of the gestures used here is accessible to apes in principle. Our results therefore suggest that great apes, at least the ones tested in these studies, have problems with *spontaneously* comprehending iconic signals.

Iconicity in the signal can contribute to children's language learning in three ways. First, children can directly identify the intended referent of the word by recognizing the similarity between the iconic signal and the corresponding aspect of their previous experience. Second, overlap in perceptual features makes referents more salient when referred to by a signal comprising iconic properties, creating more opportunities for direct word to meaning mapping. Finally, iconic properties can systematically co-vary with other signal properties that facilitate learning. Here, we were primarily interested in the first type. Our results show that children spontaneously recognize correspondence at around 24 months of age if the signal comprises both vocalization and gesture. At 36 months of age, this ability becomes more robust, with solid recognition of combined signals and gestures. In all cases, performance in the first trial mirrored the overall performance, suggesting that children indeed spontaneously recognized the iconic correspondence. While vocalizations contributed to comprehension when combined with gestures, at least for 24-month-olds, we did not find any evidence that vocalization alone allowed children to spontaneously identify the referent. Together with the findings by Laing et al. (2017) and Fort et al. (2018), our results therefore question the interpretation that sound-symbolic or onomatopoetic words offer children with "a nascent representation of the word meaning without effort" (Imai & Kita, 2014).

The difference between gesture and vocalization conditions further suggests that children differentiate between different forms of iconicity. Experiment 4 showed that gesture comprehension was immune to the way that children learned about the action (through observation or experience), suggesting that the difference between conditions is not due to gestures being more informative because they relate to multiple aspects of experience simultaneously. This pattern resonates with other research, finding that action-based gestures, like the ones we used here, are especially suited to establish reference in the absence of language (Cartmill, Rissman, Novack, & Goldin-Meadow, 2017; Fay, Lister, Ellison, & Goldin-Meadow, 2014; Ortega, Sümer, & Özyürek, 2017). Ortega et al. (2017) proposed that the advantage of action-based gestures stems from their direct connection to motor experience. Our results from experiment 4 show that motor and visual experience are interchangeable as the basis for gesture comprehension, suggesting that gesture relates to action representations in a more general way. Vocalization, on the other hand, does not directly relate to one's own or others' action and might therefore lack an easily accessible representational basis.

There are a number of limitations to the studies presented here that need to be addressed in the future. First, it will be important to show that the difference between gesture and vocalization comprehension holds for a wider variety of gestures and vocalizations. Furthermore, even though both aspects mimicked parts of the same experience, gestures were derived from intentional human actions whereas vocalizations mimicked a by-product of these actions. Children might encode and therefore remember these two aspects differently. In future studies, it would be interesting to see if the difference between vocalization and gesture persists if the intended outcome of the action is a sound, for example when playing a musical instrument. Gesture and vocalization also differ in the number of overlapping features compared to their source. Iconic gestures involve the same bodily movements (by the same person) as the corresponding action. Vocalizations might be considered to be further removed because they are produced by a different entity than the corresponding sound. This is, however, an inherent feature of this mode of communication. Equalizing this difference (e.g., by using a playback instead of a vocalization) would have resulted in a rather unusual type of signal and we therefore refrained from doing so. Finally, cross cultural work will be necessary to evaluate whether the developmental transition in representational abilities between two and three is something that is characteristic of western societies or more universal. Recent comparative work within western societies finds differences in iconic gesture production in this age range (Marentette, Pettenati, Bello, & Volterra, 2016), suggesting that comprehension might also vary across cultures.

To summarize, our findings indicate that iconic signals can be a powerful way to spontaneously ground reference in experience, thereby providing rich information about a signal's meaning in context. Their comprehension, however, is not trivial and develops substantially in the third year of life. Finally, not all iconic signals seem to be created equal.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Manuel Bohn http://orcid.org/0000-0001-6006-1348

REFERENCES

Andric, M., Solodkin, A., Buccino, G., Goldin-Meadow, S., Rizzolatti, G., & Small, S. L. (2013). Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia*, 51(8), 1619–1629. https://doi.org/10.1016/j.neuropsychologia.2013.03.022

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in pre-verbal infants. *Cortex*, 63, 196–205. https://doi.org/10.1016/j.cortex.2014.08.025

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Behne, T., Carpenter, M., & Tomasello, M. (2014). Young children create iconic gestures to inform others. *Developmental Psychology*, 50(8), 2049. https://doi.org/10.1037/a0037224

Boesch, C. (2007). What makes us human (homo sapiens)? The challenge of cognitive cross-species comparison. *Journal of Comparative Psychology*, 121(3), 227. https://doi.org/10.1037/0735-7036.121.3.227

Bohn, M., Call, J., & Tomasello, M. (2016). Comprehension of iconic gestures by chimpanzees and human children. *Journal of Experimental Child Psychology*, 142, 1–17. https://doi.org/10.1016/j.jecp.2015.09.001

Brentari, D., & Goldin-Meadow, S. (2017). Language emergence. *Annual Review of Linguistics*, 3, 363–388. https://doi.org/10.1146/annurev-linguistics-011415-040743

Buttelmann, D., Carpenter, M., Call, J., & Tomasello, M. (2013). Chimpanzees, *Pan troglodytes*, recognize successful actions, but fail to imitate them. *Animal Behaviour*, 86(4), 755–761. https://doi.org/10.1016/j.anbehav.2013.07.015

Call, J., & Tomasello, M. (2007). *The gestural communication of apes and monkeys*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Cartmill, E. A., Beilock, S., & Goldin-Meadow, S. (2012). A word in the hand: Action, gesture and mental representation in humans and non-human primates. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1585), 129–143. https://doi.org/10.1098/rstb.2011.0162

Cartmill, E. A., Rissman, L., Novack, M. A., & Goldin-Meadow, S. (2017). The development of iconicity in children's co-speech gesture and homesign. *Language, Interaction and Acquisition*, 8(1), 42–68. https://doi.org/10.1075/lia.8.1.03car

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: From origin to function. *Behavioral and Brain Sciences*, 37(2), 177–192. https://doi.org/10.1017/S0140525X13000903

DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science*, 238(4833), 1556–1557. https://doi.org/10.1126/science.2446392

DeLoache, J. S. (2000). Dual representation and young children's use of scale models. *Child Development*, 71(2), 329–338. https://doi.org/10.1111/1467-8624.00148

DeLoache, J. S., & Burns, N. M. (1994). Early understanding of the representational function of pictures. *Cognition*, 52(2), 83–110. https://doi.org/10.1016/0010-0277(94)90063-9

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19(10), 603–615. https://doi.org/10.1016/j.tics.2015.07.013

Dobson, A. J., & Barnett, A. (2008). *An introduction to generalized linear models*. Boca Raton, FL: CRC Press.

Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Cambridge, MA: Harvard University Press.

Douglas, P. H., & Moscovice, L. R. (2015). Pointing and pantomime in wild apes? Female bonobos use referential and iconic gestures to request genito-genital rubbing. *Scientific Reports*, 5, 13999. https://doi.org/10.1038/srep13999

Fay, N., Ellison, M., & Garrod, S. (2014). Iconicity: From sign to system in human communication and language. *Pragmatics & Cognition*, 22(2), 244–263. https://doi.org/10.1075/pc.22.2.05fay

Fay, N., Lister, C. J., Ellison, T. M., & Goldin-Meadow, S. (2014). Creating a communication system from scratch: Gesture beats vocalization hands down. *Frontiers in Psychology*, 5, 354. https://doi.org/10.3389/fpsyg.2014.00354

Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511817779

Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: A meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental Science*, 21, e12659. https://doi.org/10.1111/desc.12659

Ganea, P. A., Allen, M. L., Butler, L., Carey, S., & DeLoache, J. S. (2009). Toddlers' referential understanding of pictures. *Journal of Experimental Child Psychology*, 104(3), 283–295. https://doi.org/10.1016/j.jecp.2009.05.008

Genty, E., Breuer, T., Hobaiter, C., & Byrne, R. W. (2009). Gestural communication of the gorilla (gorilla gorilla): Repertoire, intentionality and possible origins. *Animal Cognition*, 12(3), 527–546. https://doi.org/10.1007/s10071-009-0213-4

Genty, E., & Zuberbühler, K. (2014). Spatial reference in a bonobo gesture. *Current Biology*, 24(14), 1601–1605. https://doi.org/10.1016/j.cub.2014.05.065

Goodrich, W., & Hudson Kam, C. L. (2009). Co-speech gesture as input in verb learning. *Developmental Science*, 12(1), 81–87. https://doi.org/10.1111/j.1467-7687.2008.00735.x

Graham, K. E., Furuichi, T., & Byrne, R. W. (2017). The gestural repertoire of the wild bonobo (*Pan paniscus*): A mutually understood communication system. *Animal Cognition*, 20(2), 171–177. https://doi.org/10.1007/s10071-016-1035-9

Hobaiter, C., & Byrne, R. W. (2011). The gestural repertoire of the wild chimpanzee. *Animal Cognition*, 14(5), 745–767. https://doi.org/10.1007/s10071-011-0409-2

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1651), 20130298. https://doi.org/10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109(1), 54–65. https://doi.org/10.1016/j.cognition.2008.07.015

Laing, C. E. (2017). A perceptual advantage for onomatopoeia in early word learning: Evidence from eye-tracking. *Journal of Experimental Child Psychology*, 161, 32–45. https://doi.org/10.1016/j.jecp.2017.03.017

Laing, C. E., Vihman, M., & Keren-Portnoy, T. (2017). How salient are onomatopoeia in the early input? A prosodic analysis of infant-directed

speech. *Journal of Child Language*, 44(5), 1117–1139. https://doi.org/10.1017/S0305000916000428

Marentette, P., Pettenati, P., Bello, A., & Volterra, V. (2016). Gesture and symbolic representation in Italian and English-speaking Canadian 2-year-olds. *Child Development*, 87(3), 944–961. https://doi.org/10.1111/cdev.12523

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound–shape correspondences in toddlers and adults. *Developmental Science*, 9(3), 316–322. https://doi.org/10.1111/j.1467-7687.2006.00495.x

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1651), 20130299. https://doi.org/10.1098/rstb.2013.0299

Montgomery, K. J., Isenberg, N., & Haxby, J. V. (2007). Communicative hand gestures and object-directed hand movements activated the mirror neuron system. *Social Cognitive and Affective Neuroscience*, 2(2), 114–122. https://doi.org/10.1093/scan/nsm004

Mulcahy, N. J., & Hedge, V. (2012). Are great apes tested with an abject object-choice task? *Animal Behaviour*, 83(2), 313–321. https://doi.org/10.1016/j.anbehav.2011.11.019

Mumford, K. H., & Kita, S. (2014). Children use gesture to interpret novel verb meanings. *Child Development*, 85(3), 1181–1189. https://doi.org/10.1111/cdev.12188

Namy, L. L. (2001). What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories. *Infancy*, 2(1), 73–86. https://doi.org/10.1207/S15327078IN0201_5

Namy, L. L. (2008). Recognition of iconicity doesn't come for free. *Developmental Science*, 11(6), 841–846. https://doi.org/10.1111/j.1467-7687.2008.00732.x

Namy, L. L., Campbell, A. L., & Tomasello, M. (2004). The changing role of iconicity in non-verbal symbol learning: A u-shaped trajectory in the acquisition of arbitrary gestures. *Journal of Cognition and Development*, 5(1), 37–57. https://doi.org/10.1207/s15327647jcd0501_3

Namy, L. L., & Waxman, S. R. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference. *Child Development*, 69(2), 295–308. https://doi.org/10.1111/j.1467-8624.1998.tb06189.x

Novack, M. A., Goldin-Meadow, S., & Woodward, A. L. (2015). Learning from gesture: How early does it happen? *Cognition*, 142, 138–147. https://doi.org/10.1016/j.cognition.2015.05.018

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, 112(1), 181–186. https://doi.org/10.1016/j.cognition.2009.04.001

Ortega, G., Sümer, B., & Özyürek, A. (2017). Type of iconicity matters in the vocabulary development of signing children. *Developmental Psychology*, 53(1), 89. https://doi.org/10.1037/dev0000161

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, 114(2), 173–186. https://doi.org/10.1016/j.jecp.2012.05.004

Peirce, C. S. (1932). *Collected papers of Charles sanders Peirce, volumes I and II: Principles of philosophy and elements of logic*. Cambridge, MA: Harvard University Press.

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1651), 20130300. https://doi.org/10.1098/rstb.2013.0300

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE*, 10(9), e0137147. https://doi.org/10.1371/journal.pone.0137147

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2017). Iconicity in the speech of children and adults. *Developmental Science*, 21, e12572. https://doi.org/10.1111/desc.12572

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Russon, A., & Andrews, K. (2010). Orangutan pantomime: Elaborating the message. *Biology Letters*, 7(4), 627-30. https://doi.org/10.1098/rsbl.2010.0564

Stadt Leipzig. (2018). Persönliches und Haushaltsnettoeinkommen. Retrieved from https://statistik.leipzig.de/statcity/table.aspx?cat=9&rub=2

Sterelny, K. (2017). From code to speaker meaning. *Biology & Philosophy*, 32(6), 819–838. https://doi.org/10.1007/s10539-017-9597-8

Tolar, T. D., Lederberg, A. R., Gokhale, S., & Tomasello, M. (2007). The development of the ability to recognize the meaning of iconic signs. *Journal of Deaf Studies and Deaf Education*, 13(2), 225–240. https://doi.org/10.1093/deafed/enm045

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tomasello, M., Striano, T., & Rochat, P. (1999). Do young children use objects as symbols? *British Journal of Developmental Psychology*, 17(4), 563–584. https://doi.org/10.1348/026151099165483

Villarreal, M., Fridman, E. A., Amengual, A., Falasco, G., Gerscovich, E. R., Ulloa, E. R., & Leiguarda, R. C. (2008). The neural substrate of gesture recognition. *Neuropsychologia*, 46(9), 2371–2382. https://doi.org/10.1016/j.neuropsychologia.2008.03.004

Werner, H., & Kaplan, B. (1963). *Symbol formation: An organismic-developmental approach to language and the expression of thought*. Hoboken, NJ: John Wiley & Sons.

Zlatev, J., Persson, T., & Gärdenfors, P. (2005). Bodily mimesis as "the missing link" in human cognitive evolution. *Lund University Cognitive Studies*, 121, 1–45.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.