Fifteen-month-olds accept arbitrary shapes as symbols of familiar kind tokens

Barbu Revencu (revencu_barbu@phd.ceu.edu)
Central European University, Austria
Neurospin Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris- Saclay

Barbara Pomiechowska
University of Birmingham, UK

Gabor Brody
Brown University, USA

Gergely Csibra
Central European University, Austria
Birkbeck, University of London, UK

Abstract

Across three experiments, we show that 15-month-old infants understand that arbitrary objects can be used as symbols. Experiment 1 shows that infants map geometric shapes (e.g., a triangle) onto familiar discourse referents (e.g., a duck) based on labeling (e.g., "Look, a duck!"). Experiment 2 shows that infants do not generalize these mappings to a new speaker. This rules out the alternative hypothesis that infants interpret the labeling events literally. Experiment 3 shows that infants are sensitive to the conceptual identity of the discourse referent. After being told that one shape represents an agent (e.g., a duck) and another shape represents a patient (e.g., a cup), infants attend differentially when the agent symbol moves towards the patient symbol than the opposite. This rules out the alternative hypothesis that infants interpret the labeling events as referential pacts. The findings jointly indicate that symbolic relations are easily activated and available early in human development.

Keywords: symbols; discourse referents; communication; cognitive development; pretend play

Introduction

In communication, humans often set up STAND-FOR relations—local mappings between visual symbols and discourse referents, the entities they want to communicate about (e.g., Clark, 2016). The symbols may be objects that are immediately available in the environment and repurposed to depict a relevant scene (e.g., bottles on the table manipulated to convey the dynamics of a car accident). Alternatively, the symbols may be generated specifically for the occasion, as in graphs or maps (e.g., rectangles used to represent bridges).

Often, what a symbol represents cannot be retrieved from its visual features. In such cases, the identity of the discourse referent can be conveyed via linguistic stipulation, either verbally (e.g., "This bottle is a car, this pencil is a pedestrian") or in a legend appended to the representation (e.g., □ = bridge). Taken literally, these predicative expressions would cause confusion since bottles are not cars and nor are rectangles bridges. Yet the literal interpretation does not even seem to be even considered. In such cases, human adults intuitively infer that "is a" and the equality sign are shorthand for "stands for" without being confused about the literal falsity of the predication. Moreover, adults are also aware that these mappings are local. Outside of the current communicative context, the objects may not stand for the same referents (e.g., rectangles can stand for other things in a different graph).

In this study, we ask, first, whether human infants can set up STAND-FOR relations between arbitrary visual objects (e.g., a triangle) and discourse referents belonging to familiar kinds (e.g., a duck). Second, we test the locality of STAND-FOR relations by checking whether infants generalize these mappings to a different discourse. Third, we test whether infants, like adults, interpret events involving the symbols in terms of the identity of the discourse referents. That is, after establishing that the bottle stands for a car, and the pencil for a pedestrian, will they interpret the bottle moving towards the pencil as the car approaching the pedestrian?

Experiment 1: Arbitrary Symbols

Experiment 1 tested whether infants can represent STAND-FOR relations between geometric shapes and discourse referents in a looking-while-listening paradigm based on (Pomiechowska, Brody, Csibra, & Gliga, 2021). If they can, they should accept objects as symbols for discourse referents even when those objects do not belong to the kind to which the discourse referent belongs—just like they do in pretense. Infants were exposed to geometric shapes (e.g., an octagon and a triangle), one of which received a familiar label (e.g., "car" applied to an octagon). They were then asked a question containing the same word used at stipulation (e.g., "Where is the car?") or a different word, not heard previously in the trial (e.g., "Where is the spoon?"). We predicted that, upon hearing the same word at test, infants (i) would look above chance at the labeled object; and (ii) would look longer at it than when hearing a different word.

Methods

Transparency and Openness The hypotheses and methods were preregistered (`http://tinyurl.com/sjcct7yv`). The stimuli, sample trial videos, anonymized data, and analysis scripts are available on the project's online Open Science Framework repository, accessible at `http://tinyurl.com/vt9zbt3w`. The local ethical committee approved all experiments, and informed consent was obtained from the participants' caregivers before the testing session.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society*.

Participants The final sample consisted of 32 typically developing German-speaking 14–16-month-old infants (M_{age} = 15 months 9 days, SD_{age} = 22.3 days). An additional 4 infants were tested and excluded due to fussiness ($n = 2$) or failure to provide sufficient valid data ($n = 2$). The sample size was set based on a pilot experiment with 10 participants analyzed with a growth-curve model that we eventually discarded. However, the sample size is large enough to detect a mid-sized effect of trial type with 80% power and is above average for infant studies.

Apparatus Infants' gaze was recorded using a Tobii Pro Spectrum eye tracker with an integrated 23.8-inch-diagonal monitor (resolution: 1920 × 1080; refresh rate: 60 Hz). External speakers delivered the sound. A custom Python program built on PsychoPy 2021.1.3 (Peirce et al., 2019) was used to calibrate infants' gaze, present the stimuli, and collect the eye data.

Stimuli Two sets of visual stimuli were used: color photographs representing 12 objects that are familiar to German-speaking infants of this age (Grimm & Doil, 2019) and eight pairs of geometric shapes (see Figure 1 for examples). The objects' bounding boxes were matched in width and, whenever possible, in height; their display size was approximately 330 × 330 pixels. The stimuli also included an image of a pointing hand, displayed at 213 × 366 pixels. Audio stimuli were the 12 nouns corresponding to the familiar kinds depicted by the photographs, embedded in carrier phrases: "Hi baby! Look! An X! Here's an X! Wow, an X!", "Where is the X? X!". The stimuli were recorded by a female native speaker of Austrian German in infant-directed speech.

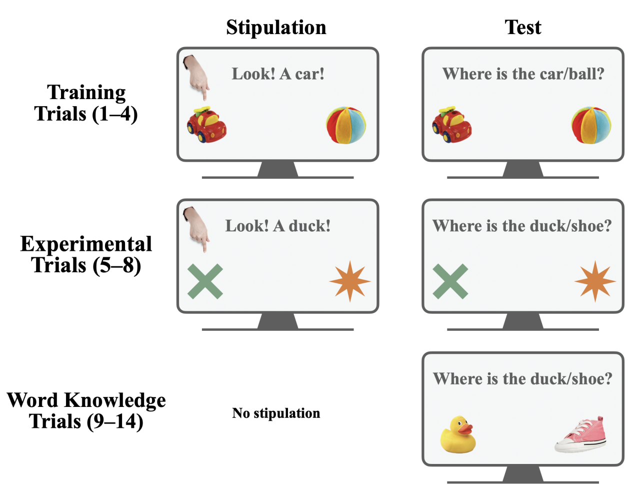

Figure 1: Overview of the three trial types in Experiment 1.

Procedure Infants were shown animated clips while seated on their caregivers' laps. Caregivers wore opaque glasses throughout the procedure. The experiment consisted of 14 trials, split into 4 Training, 4 Experimental, and 6 Word Knowledge trials (Figure 1).

Experimental trials (Trials 5–8) consisted of three parts: *baseline*, *stipulation*, and *test*. All trials started with a blue curtain covering the display. An attention-getter appeared in the center of the screen and rotated until the infant oriented to it for 500 ms. The curtain then went up to reveal a static display of two geometric shapes, one on the left and one on the right of the screen. During *baseline*, the static display was shown to infants for 2 seconds in silence. Then, the *stipulation* part started. An animated hand appeared above one of the two shapes, pointing to it. The hand moved up and down while the infant was greeted ("Hi baby! Look!") to draw their attention to the object. The hand stopped above the object, and infants heard a familiar word 3 times in different carrier phrases ("A duck! Here's a duck! Wow, a duck!"). This part lasted 10 seconds and was followed by a 750-ms break.

The *test* event started immediately afterward, with a colored rotating spiral appearing in the center of the screen to draw infants' attention to a neutral point on the display, equidistant from both objects. Once the infant oriented to it for 500 ms, the spiral started expanding and contracting cyclically while the test question was played. Depending on the trial type, the test question contained the same word used during the *stipulation* event (e.g., "Where is the duck?") or a different word, not heard before during the trial (e.g., "Where is the cup?"). The attention-getter disappeared at the offset of the test question, which coincided with the start of the measurement period. One second into the measurement period, infants heard the test word one more time (e.g., "Duck!"/"Cup!"). The measurement period ended after 3.5 seconds, when the blue curtain went down to cover the entire display. Each trial lasted approximately 22 seconds.

Training trials (Trials 1–4) were identical to Experimental trials except that the two shapes were replaced by familiar objects, which received their regular labels. These trials were meant to familiarize infants with the general procedure and speaker and to give them evidence that the disembodied voice is connected to what is happening on the screen. Word Knowledge trials (Trials 15–18) were identical to Training trials, except that the *stipulation* event was removed. The two-second silent period at the beginning of the trial was immediately followed by the test question.

Design The experiment had a within-subjects design with one independent variable, Trial type, which was two-leveled: Same Word and Different Word. For each infant, the Experimental trials alternated according to an ABBA pattern, with the type of the first trial (Same Word or Different Word) counterbalanced across subjects. The side of the object that was pointed to and labeled during the *stipulation* event followed an ABAB pattern, the first side (left or right) counterbalanced across subjects. For the six Word Knowledge trials, the side of the correct response followed an ABBABA structure, with first side (left or right) counterbalanced across subjects. The object pairings were randomly sampled for each subject. First, eight of the 12 object photographs were sampled and grouped into four pairs to create the Training trials.

In each pair, the two objects were barred from belonging to the same superordinate kind (e.g., animates: bird–duck), and the two words referring to them were not allowed to start with the same phoneme. Within each pair, one object was the target of labeling in the *stipulation* phase. Note that six words are needed to create four trials (one for each Same Word trial, two for each Different Word trial). For the Experimental trials, we sampled four shape pairings and assigned the remaining six words for the *stipulation* phase. There was thus no overlap between the words used at Training and those used in the Experimental block. The Word Knowledge trials tested whether infants knew the words used during the Experimental trials. In each trial, an object photograph denoted by a noun previously applied to the geometric shapes was paired with an object photograph from the Training block.

Data Processing and Exclusion At the end of each testing session, the Python script outputted a data file containing the infant's gaze information on each sample (every 16.67 ms). The gaze coordinates were averaged across eyes along both horizontal and vertical axes. The screen was divided into three regions, depending on whether the infant was looking at the left-object area of interest (AOI), the right-object AOI, or elsewhere on the screen. A gaze data point was considered valid if the eye-tracker registered the gaze for at least one eye. As preregistered, an Experimental trial was excluded from the analysis if infants provided less than 60% valid data during *baseline* ($n = 3$ trials) or *test* ($n = 6$ trials). Infants were excluded if they failed to provide at least one valid trial of each type in the Experimental block ($n = 2$). After this preprocessing step, we derived a new variable, Highlighted Object, which received a score of 1 if infants' gaze fell into the AOI of the highlighted object, 0 if infants looked to the AOI of the other object, and NA if infants looked elsewhere on the screen. The samples from the *test* period of each trial were grouped into 50-ms bins (70 bins total, corresponding to the 3.5-second *test* period). For each bin, we computed the variable PLH, representing the **p**roportion of **l**ooks to the region of the **h**ighlighted object out of the sample that fell into the regions of either object.

Missing Data There were several types of missing data, all of which were excluded from the analysis: (i) missing gaze data due to eye tracker signal loss (e.g., because infants looked away from the screen); (ii) missing AOI gaze data (e.g., because infants looked to other parts of the screen during a particular bin); and (iii) missing trials (e.g., because the infant did not provide enough valid data for that trial).

Data Analysis We initially preregistered a growth-curve analysis on infants' PLH in each test time bin. We decided against this analysis due to its high false-positive rate (Huang & Snedeker, 2020). Instead, we report two-tailed paired t-tests on PLH trial averages. Even though this analysis deviates from the preregistration, it is more conservative than the growth-curve one. This makes it more likely to falsify our predictions than to confirm them.

Results

Figure 2 plots infants' proportions of looking at the highlighted object at test in the Experimental block, averaged by trial type. Infants looked at the highlighted object above chance on Same Word trials, $t(31) = 3.78$, $p = .001$, Cohen's $d = 0.67$. On Different Word trials, infants' PLH scores did not differ from chance, $t(31) = 1.41$, $p = .169$, Cohen's $d = 0.25$. While they looked more at the highlighted object on Same Word than on Different Word trials, the difference does not exclude the null hypothesis, $t(31) = 1.55$, $p = .132$, Cohen's $d = 0.27$.

This was not because the words applied to the shapes were new to them. The results in the Word Knowledge block show that infants knew the labels used in the Experimental block, $t(31) = 4.28$, $p < .001$, Cohen's $d = 0.76$.

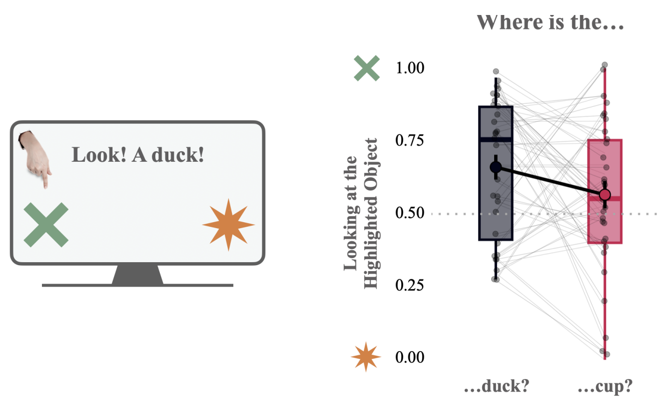

Figure 2: Experimental setup and results of Experiment 1. Infants look above-chance at a shape given a familiar noun (e.g., "duck") when asked about the same noun but not when asked about a different noun.

Discussion

The main finding of Experiment 1 is that infants reliably look to a geometrical shape labeled "duck" when asked where the duck is. By contrast, when asked about a different word, infants did not distinguish between the two objects[1]. But why would infants accept a mislabeling event with such ease? Our STAND-FOR account predicted this, but other accounts can accommodate this finding as well. Infants could have interpreted the statements literally and, for instance, recategorized the cross as a duck after it was labeled as such (e.g., Jaswal & Markman, 2007). A second possibility is that infants inferred that all the nouns used in this strange way are homonyms (e.g., Mazzocco, 1997). That is, on hearing "This is a duck!" applied to a green cross, perhaps they created a new lexical entry for "duck" that meant cross. These alternative hypotheses attribute a global representation to infants, since, in both cases, infants would learn generalizable information: either

[1]The lack of a statistical difference between the two trial types is likely due to the growth curve model on which the sample size was based, which probably overestimated the magnitude of the effect. Experiment 2 confirms this.

that the green cross is a duck or that the noun "duck" has a second meaning. This is unlike STAND-FOR relations, where the link between the cross and the duck should not be extended outside the communicative episode in which the link is created. To test this, we reasoned that if infants come to believe that the cross is a duck, or if they create a new lexical entry for "duck", then it should not matter who probes that piece of information. Experiment 2 therefore introduced a second speaker and varied the identity of the speaker who asks the test question.

Experiment 2: Different Speakers

Methods

Transparency and Openness The Open Science Framework preregistration can be accessed at `http://tinyurl.com/5tykfuh7`. Except where noted, the methods were identical to Experiment 1.

Participants The final sample consisted of 64 typically developing German-speaking 14–16-month-olds (M_{age} = 15 months 10 days, SD_{age} = 27.26 days)[2]. An additional 16 subjects were tested and excluded due to fussiness ($n = 9$), parental intervention ($n = 1$), or failure to provide enough valid data ($n = 6$).

Stimuli The audio stimuli recorded by the female speaker in Experiment 1 were doubled by a new set of stimuli recorded by a male speaker. The recordings from the two speakers were closely matched in duration ($r = .98$).

Design Training trials were identical to Experiment 1, except that two were delivered by the female speaker and two by the male speaker. This was done to accustom infants to both speakers. As infants were not supposed to infer that the speakers were part of the same scene, the same speaker delivered both the stipulation and the test question at Training.

Experiment 2 had two Experimental blocks—Same Speaker and Different Speaker—consisting of the same shape–word mappings. Each block was delivered either before (Trials 5–8) or after the Word Knowledge block (Trials 15–18), counterbalanced across subjects. In the Different Speaker block, one speaker delivered the stipulation, and the other asked the test question. In the Same Speaker block, the same speaker delivered the stipulation and test question (as in Experiment 1). For the Training and first Experimental block, the speaker delivering the stipulation alternated in an ABBA-ABAB pattern for half the subjects and in an ABAB-ABBA pattern for the other half (Male vs. Female on the first trial counterbalanced). In the Word Knowledge trials (Trials 9–14), half the subjects were tested by the male speaker and half by the female speaker.

[2]Experiment 2 was originally planned as a smaller experiment (n = 32) with the Experimental blocks presented in fixed order: Different Speaker followed by Same Speaker. We obtained the predicted result but wanted to ensure its robustness and therefore ran a second experiment with the Experimental blocks presented in reverse order. For brevity, we present the data collapsed across the two samples.

Data Analysis Using the same trial exclusion criteria as in Experiment 1, we excluded 17 trials in the first Experimental block (Trials 5–8) and 36 trials in the second Experimental block (Trials 15–18). Five infants fussed out before the second Experimental block but were included in the analysis because they passed the valid-data threshold in the first block.

Results

The Same Speaker block replicated Experiment 1 (Figure 3). Infants looked at the highlighted object above chance in Same Word trials, $t(59) = 3.70$, $p < .001$, Cohen's $d = 0.48$, but not in Different Word trials, $t(59) = 0.50$, $p = .617$, Cohen's $d = 0.065$. The effect of trial type was higher than in Experiment 1 and incompatible with the null hypothesis, $t(26) = 2.03$, $p = .046$, Cohen's $d = 0.27$.

Infants' behavior in the Different Speaker block was markedly different. They did not look at the highlighted object above chance in either Same Word trials, $t(61) = 1.67$, $p = .1$, Cohen's $d = 0.21$, or in Different Word trials, $t(31) = 0.90$, $p = .37$, Cohen's $d = 0.12$. In addition, infants looked equally at the highlighted object in Same Word and Different Word trials, $t(60) = 0.43$, $p = .666$, Cohen's $d = 0.06$.

An exploratory mixed ANOVA on the subset of infants who passed the inclusion criteria in both blocks, with Speaker Identity and Trial Type as within-subjects factors and Block Order (Same-Speaker first versus Different-Speaker first) as a between-subjects factor, revealed that Block Order may have interfered with the main independent variables (Trial Type × Speaker Identity × Block Order: $F(1, 54) = 3.748$, $p = .058$). When the Different-Speaker block was presented first, there was a significant interaction between Trial Type and Speaker Identity, $F(1, 26) = 5.628$, $p = .025$. When the Same-Speaker block was presented first, there was no such interaction, $F(1, 28) = 0.17$, $p = .684$. Nevertheless, even infants in this subset looked reliably at the highlighted object only on Same Speaker–Same Word trials, $t(31) = 2.77$, $p = .009$ (all other ps $> .08$).

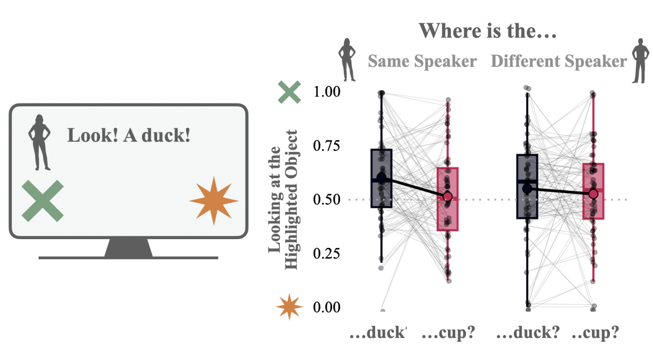

Figure 3: Experimental setup and results of Experiment 2. Infants look above-chance at a shape labeled with a familiar noun (e.g., "duck") only if the same speaker who stipulated the mapping asks the test question.

Discussion

Experiment 2 shows that infants did not generalize the shape–label mapping to a new speaker. Thus, they neither recategorized the objects based on the labeling event nor created new lexical entries for the known nouns. Therefore, the mappings infants set up are local to the context in which they are created, in line with the STAND-FOR account.

However, it remains possible that the locality has a different origin from the one we hypothesized. For instance, infants might have taken the stipulating speaker to be an unreliable source (e.g., Koenig & Woodward, 2010), who falsely believe the cross is a duck. In this case, the locality of the shape-label mapping would have arisen due to an idiosyncratic property of the speaker. Alternatively, infants may have understood the labeling events as introducing referential pacts (e.g., Matthews, Lieven, & Tomasello, 2010). In this case, infants would have interpreted "duck" like a proper name—an arbitrary label the speaker uses to refer to the cross. There is no reason under these accounts to interpret the cross in a duck-related way once the mapping is in place. In the unreliable-testimony case, infants know the cross is not a duck. In the referential-pact case, the noun "duck" is an empty label that picks out the cross. By contrast, in the STAND-FOR case, once a cross stands for a duck, the information conveyed via the cross will be interpreted as information predicated about the (imagined) duck. To test this, Experiment 3 asked whether infants distinguish events that are consistent with the identities ascribed to the shapes from those that are not, building on the early-developing distinction between animate agents and inert objects (e.g. Rakison, Cicchino, & Hahn, 2007).

Experiment 3: Moving Symbols

Methods

Transparency and Openness The experiment was preregistered at the Open Science Framework (`http://tinyurl.com/mr27wnba`). Except where noted, the methods are identical to Experiments 1–2.

Participants The final sample consisted of 32 typically developing Hungarian-speaking 14–16-month-olds (M_{age} = 15 months 23 days, SD_{age} = 24.3 days). Results from a pilot with eight participants indicated that 28 participants would be enough to detect an effect of looking time with 80% power. An additional eight infants were tested and excluded due to fussiness ($n = 6$), technical error ($n = 1$), or maxing out on seven out of the eight trials ($n = 1$).

Procedure The experiment consisted of eight trials, split into four Training and four Experimental trials. As in Experiments 1–2, Training trials (1–4) were meant to familiarize the infant with the general procedure and to give them evidence that the voice they hear is connected to what is happening on the screen. The structure of a trial was very close to that in Experiments 1–2. Infants were shown displays of the same shapes as in Experiments 1 and 2, but here both shapes were pointed to and labeled. One of the shapes was labeled with a noun denoting an animate kind (e.g., "duck"), the other with a noun denoting an inanimate kind (e.g., "shoe"). Afterward, one of the objects started moving toward the other object, on a straight path, at uniform speed. On Congruent trials, the moving object was the shape standing for the animate entity; on Incongruent trials, the moving object was the shape standing for the inanimate one. After the moving object reached the stationary object, a ding sound was played, and the moving object wiggled for 666 ms (by rotating left–right around its vertical axis for two cycles). After the wiggling stopped, looking times were measured until infants looked away from the screen for 2 consecutive seconds or until 30 seconds passed. Each trial lasted between 30 seconds (without the test period) and up to a minute (with the test period).

Design The experiment had a within-subjects design with one independent variable, Trial type, and two levels, Congruent and Incongruent. Each infant was administered the following trial alternation: ABBA-BAAB (Training–Experimental), with type of first trial (Congruent or Incongruent) counterbalanced across subjects. The side of the object labeled with an animate noun and the side of the object labeled first were counterbalanced across subjects. The animate–inanimate object pairings and the visual symbol pairings were fixed, but the symbol–label pairings within each pair of symbols were counterbalanced. The pair succession cycle was fixed across subjects (duck–shoe, bear–spoon, cat–banana, dog–sippy cup), but the identity of the pairing shown in the first trial was counterbalanced across subjects. Thus, across subjects, each pair appeared an equal number of times in each of the eight serial positions.

Coding Infants' looking times were coded online by the experimenter, who pressed a key whenever the infant looked away from the screen (if the key was pressed for 2 seconds uninterruptedly, the trial ended). The key presses were recorded in the data file obtained for each infant at the end of the test session. Looking times were also coded offline by the first author based on video recordings. Online–offline inter-rater reliability was substantial (total looking time: Spearman's $\rho = .96$; first looks: Spearman's $\rho = .89$). The analyses are based on the offline coding.

Data Exclusion As preregistered, experimental trials were excluded if infants attended the screen less than 40% of the time during the movement phase ($n = 2$ trials). Two additional Experimental trials were ended too early by the experimenter and had to be excluded as well.

Measures We preregistered two paired t-tests for both total looking times and first looks. Total looking times measure the amount infants spend looking anywhere on the screen once movement stopped until looking away from the screen for 2 seconds without looking back or until 30 seconds pass. First looks measure the amount infants spend looking on-screen before disengaging for the first time. Looking times were log-transformed before the analysis (Csibra, Hernik, Mascaro, Tatone, & Lengyel, 2016). If infants are sensitive to the nouns heard during the stipulation, they should distinguish the trials in which animate referents move toward inanimate referents from the trials in which the opposite occurs. We predicted that infants would look longer on Congruent than on Incongruent trials. This may be surprising, since infants tend to look longer at incongruent stimuli (but see Hernik, Fearon, & Csibra, 2014). However, the pilot revealed a preference for Congruent trials in both total and first looks. This may be because looking times do not index violation of expectation in this task but preference or anticipation[3].

Results

As predicted, infants looked longer at the screen on Congruent trials than on Incongruent trials (Figure 4). The effect was significant on first looks, $t(31) = 2.53$, $p = .021$, Cohen's $d = 0.43$. Total looking times exhibited a similar but weaker pattern, $t(31) = 1.45$, $p = .157$, Cohen's $d = 0.26$.

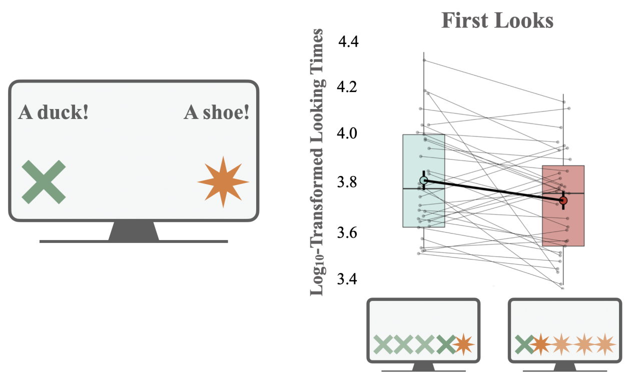

Figure 4: Experimental setup and results of Experiment 3. Infants look longer on trials in which the movement of the symbols is compatible with their conceptual identities.

Discussion

Experiment 3 shows that infants interpret the events involving the symbols based on the identity of the corresponding discourse referents. This is not due to the expectation that symbols will inherit the properties of the discourse referents they stand for (e.g., that duck symbols will move) but to the fact that the motion of the shape is interpreted symbolically as the motion of the discourse referent the shape stands for. Thus, after assigning the cross to a duck and the star to a shoe, the movement of the cross is more easily interpretable (i.e., as the duck moving) than the movement of the star, because ducks are animate and shoes are not.

Experiment 3 thus rules out that infants in Experiments 1–2 inferred that the speaker was unreliable. If they had, they would not have had a preference for the movement of the shapes depending on how they were labeled. Experiment 3 also rules out the referential-pact alternative. The nouns applied to the shapes were not interpreted as proper names: infants took the noun "duck" to mean duck[4].

General Discussion

The present study indicates that 15-month-olds represent STAND-FOR relations. Experiment 1 shows that 15-month-olds accept arbitrary shapes as symbols. Experiment 2 shows that infants know that STAND-FOR relations are local to the discourse and should thus not be generalized to a new speaker. Finally, Experiment 3 shows that infants interpret the events involving the symbols based on their knowledge about the discourse referents.

The results have implications for the investigation of several other cognitive phenomena in development. On the theoretical side, the results suggest that object substitution pretense is not a special cognitive capacity (Harris & Kavanaugh, 1993; Leslie, 1987). Instead, infants—at least from 15 months onward—can set up local assignments between arbitrary objects and discourse referents, and this capacity manifests itself under many guises, one of which is pretend play.

On the methodological side, these findings are relevant to labeling studies in general and to mislabeling studies in particular (e.g., Csink, Mareschal, & Gliga, 2021; Dautriche, Goupil, Smith, & Rabagliati, 2021; Koenig & Woodward, 2010). In mislabeling experiments, infants (or children) are exposed to an adult speaker who consistently mislabels everyday objects, and researchers measure infants' subsequent inferences about the speaker. In the present experiments, mislabeling occurred as well, yet infants embraced this without inferring anything about speaker reliability. Thus, mislabeling events may be taken as stipulating STAND-FOR relations, at least on some occasions. This would explain why infants are less likely to learn new words in such contexts: They know this is not a word-learning situation, so they will simply restrict the mappings to the local discourse.

At the same time, the results raise a challenging question. How do infants know when to interpret a labeling event as STAND-FOR stipulation versus IS-A predication? It could be that infants can access both interpretations and use additional information to choose between the two. In our study, infants may have opted for the STAND-FOR interpretation based on world knowledge: the cross is not a duck, so it must stand for one. Alternatively, infants may go for the STAND-FOR interpretation by default and extract the IS-A meaning over many encounters with repeated stipulations of the same kind. To explore these possibilities, a version of Experiment 2 in which the familiar nouns are replaced by nonce words would be a good place to start. If infants generalize these nouns to a different speaker (e.g., Buresh & Woodward, 2007), this will establish that they can access both interpretations and select the one that fits best in the context.

[3]In a follow-up experiment conducted after submitting this manuscript, we found that more extensive familiarization with the task leads to the standard incongruency effect on total looking times.

[4]This does not mean that STAND-FOR relations cannot be construed as a type of pact but that they involve relations between objects and referents, not between labels and objects.

Acknowledgments

We thank Dori Mészégető and Bori Széplaki-Köllőd for data collection, and Eva Wittenberg and Jesse Snedeker for their healthy skepticism toward this project.

This research has received partial funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement no. 742231 (PARTNERS).

References

Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition*, *104*(2), 287–314. doi: 10.1016/j.cognition.2006.07.001

Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, *123*(3), 324–347. doi: https://doi.org/10.1037/rev0000026

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology*, *52*(4), 521–536. doi: 10.1037/dev0000083

Csink, V., Mareschal, D., & Gliga, T. (2021). Does surprise enhance infant memory? Assessing the impact of the encoding context on subsequent object recognition. *Infancy*, *26*(2), 303–318. doi: 10.1111/infa.12383

Dautriche, I., Goupil, L., Smith, K., & Rabagliati, H. (2021). Knowing how you know: Toddlers reevaluate words learned from an unreliable speaker. *Open Mind: Discoveries in Cognitive Science*, *5*, 1–19. doi: 10.1162/opmi{_}a{_}00038

Grimm, H., & Doil, H. (2019). *ELFRA. Elternfragebögen für die Früherkennung von Risikokindern.* Unpublished doctoral dissertation, [Parent Questionnaire for Early Detection of Children at Risk]. Göttingen: Hogrefe.

Harris, P. L., & Kavanaugh, R. D. (1993). Young children's understanding of pretense. *Monographs of the Society for Research in Child Development*, *58*(1), v-92. doi: 10.2307/1166074

Hernik, M., Fearon, P., & Csibra, G. (2014). Action anticipation in human infants reveals assumptions about anteroposterior body-structure and action. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1781), 20133205. doi: 10.1098/rspb.2013.3205

Huang, Y., & Snedeker, J. (2020). Evidence from the visual world paradigm raises questions about unaccusativity and growth curve analyses. *Cognition*, *200*, 104251. doi: 10.1016/j.cognition.2020.104251

Jaswal, V. K., & Markman, E. M. (2007). Looks aren't everything: 24-month-olds' willingness to accept unexpected labels. *Journal of Cognition and Development*, *8*(1), 93–111. doi: 10.1080/15248370709336995

Koenig, M. A., & Woodward, A. L. (2010). Sensitivity of 24-month-olds to the prior inaccuracy of the source: Possible mechanisms. *Developmental Psychology*, *46*(4), 815–826. doi: 10.1037/a0019664

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind.". *Psychological Review*, *94*(4), 412–426. doi: 10.1037/0033-295x.94.4.412

Matthews, D., Lieven, E., & Tomasello, M. (2010). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Developmental Psychology*, *46*(4), 749–760. doi: 10.1037/a0019657

Mazzocco, M. M. (1997). Children's interpretations of homonyms: A developmental study. *Journal of Child Language*, *24*(2), 441–467. doi: 10.1017/s0305000997003103

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Hochenberger, R., Sogo, H., ... Lindelov, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. doi: 10.3758/s13428-018-01193-y

Pomiechowska, B., Brody, G., Csibra, G., & Gliga, T. (2021). Twelve-month-olds disambiguate new words using mutual-exclusivity inferences. *Cognition*, *213*, 104691. doi: 10.1016/j.cognition.2021.104691

Rakison, D. H., Cicchino, J. B., & Hahn, E. R. (2007). Infants' knowledge of the path that animals take to reach a goal. *British Journal of Developmental Psychology*, *25*(3), 461–470. doi: 10.1348/026151006x171893