



# 15-month-old infants fast map words but not representational gestures of multimodal labels

Daniel Puccini<sup>1\*</sup> and Ulf Liszkowski<sup>1,2</sup>

<sup>1</sup> Max Planck Research Group Communication before Language, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

<sup>2</sup> Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

## Edited by:

Jessica S. Horst, University of Sussex, UK

## Reviewed by:

Emily Mather, University of Hull, UK  
Susanne Grassmann, University of Groningen, Netherlands

## \*Correspondence:

Daniel Puccini, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, Netherlands.  
e-mail: daniel.puccini@mpi.nl

This study investigated whether 15-month-old infants fast map multimodal labels, and, when given the choice of two modalities, whether they preferentially fast map one better than the other. Sixty 15-month-old infants watched films where an actress repeatedly and ostensibly labeled two novel objects using a spoken word along with a representational gesture. In the test phase, infants were assigned to one of three conditions: Word, Word + Gesture, or Gesture. The objects appeared in a shelf next to the experimenter and, depending on the condition, infants were prompted with either a word, a gesture, or a multimodal word–gesture combination. Using an infant eye tracker, we determined whether infants made the correct mappings. Results revealed that only infants in the Word condition had learned the novel object labels. When the representational gesture was presented alone or when the verbal label was accompanied by a representational gesture, infants did not succeed in making the correct mappings. Results reveal that 15-month-old infants do not benefit from multimodal labeling and that they prefer words over representational gestures as object labels in multimodal utterances. Findings put into question the role of multimodal labeling in early language development.

**Keywords:** representational gesture, multimodal, baby signs, infant communication, word learning, fast mapping

## INTRODUCTION

Multimodal speech–gesture combinations are an integral part of language development. Caregivers, for example, often provide labels for infants in temporal synchrony with gestures such as pointing and showing (Ninio, 1980; Masur, 1997; Gogate et al., 2000, 2006). These multimodal speech–gesture combinations scaffold infants' referential understanding (Iverson et al., 1999), since deictic gestures help establish joint attentional episodes which are crucial to the process of word learning (Tomasello and Farrar, 1986; Baldwin, 1991). Further, infants combine their own deictic gestures with words, and these speech–gesture combinations are predictive of the two-word stage (Iverson and Goldin-Meadow, 2005). As a whole, these studies show that deictic gestures are an integral part of multimodal gesture–speech combinations and are intimately connected with language learning, laying the grounds for first language acquisition.

Much less is known about the role of representational gestures in infants' word learning. Whereas deictic gestures direct attention to objects in the immediate environment, representational gestures stand in for the entities to which they refer (Bates et al., 1979; McNeill, 1992; Iverson et al., 1994; Capirci et al., 1996). Representational gestures can be either iconic, which have some perceptual resemblance to their referent, or arbitrary, which have no perceptual resemblance to their referent. Interestingly, infants younger than 2 years learn arbitrary gestures as easily as iconic gestures, suggesting that they are not sensitive to gestural iconicity until later in development (Bates et al., 1979; Namy et al., 2004; Namy, 2008).

One interesting possibility is that, similar to deictic gestures, representational gestures, also, facilitate language development. A recent study suggests a correlation between parents' multimodal labeling with iconic gestures and infants' acquisition of the labels (Zammit and Schafer, 2011). However, in that study parents' multimodal labeling with deictic gestures as well as their labeling without gestures also correlated with infants' acquisition of the labels. Accordingly, the study does not reveal a direct relation between iconic gestures and word learning. It has also been suggested that infants' production of representational gestures may be related to early vocabulary development, as indicated by a concurrent correlation (Acredolo and Goodwyn, 1988). The idea of a relation between representational gestures and language has recently received a great deal of attention in the public. In particular, it has led to a mini-industry offering “baby-sign” courses to parents and their babies, guided by the claim that babies can be taught representational gestures to communicate before they can talk. In an experimentally controlled training study (Goodwyn et al., 2000), one group of parents was encouraged to provide their infants with multimodal labels for a number of words that are commonly learned around this age range. Infants who received multimodal training outperformed infants in another group who received no explicit training in nearly all of the receptive and productive language measures at nearly all ages investigated (from 15 to 36 months), suggesting that multimodal labels may facilitate language development. However, there are several criticisms of this study (see Johnston et al., 2005). The authors, for example, did not report on how subjects were recruited, nor did they report

on how infants were assigned to groups. Therefore, one cannot rule out the possibility that infants in the multimodal group outperformed their peers because their parents were more motivated to begin with. And, although the study included an additional control group exposed to increased verbal labeling, no comparisons between the multimodal group and the verbal group were reported. Accordingly, it remains unclear whether prelinguistic infants benefit from multimodal utterances involving representational gestures more so than from verbal utterances without representational gestures.

Other studies have used fast mapping paradigms to investigate the role of representational gestures in word learning. On the basis of the finding that young infants can associate labels with novel objects with very little exposure (Schafer and Plunkett, 1998; Houston-Price et al., 2005), Namy and Waxman (1998, 2000), and Namy (2001) investigated whether infants can also fast map representational gestures. They found that 17- and 18-month-olds fast map both spoken words and representational gestures, suggesting that infants' early symbolic capacity is not specific to a single modality. Other fast mapping studies show that infants as young as 13 months can also associate other types of stimuli, including beeps and tones as labels for novel objects (Namy and Waxman, 1998; Woodward and Hoyne, 1999; Namy, 2001). In contrast, infants as young as 6 months expect object labels in the form of spoken words (Fulkerson and Waxman, 2007). As Fulkerson and Waxman acknowledge, the conflicting findings in these studies might be due to the methodologies employed. Namely, Fulkerson and Waxman trained infants object labels using disembodied sound-object pairings which involved no joint attention, while Namy (2001) used an interactive paradigm that assured infants were jointly attending to the intended referent while hearing the object labels.

Few fast mapping studies have directly compared the learning of gestural labels versus spoken labels when presented simultaneously in multimodal utterances. One recent experimental study suggests that multimodal labeling facilitates 3-year-olds' comprehension of novel verbs (Goodrich and Hudson Kam, 2009). In that study, participants were first shown a distinct action for each of two objects. They were then exposed to two multimodal labels containing a spoken novel verb and an iconic action depiction corresponding to each of the objects. At test, participants heard the spoken novel verb without the iconic action depiction and had to choose to which object the verb referred. Results showed that 3-year-old children benefited from iconic gestural information when learning novel verbs. Another study (Wilbourn and Sims, *in press*) investigated whether 26-month-olds can learn multimodal labels including spoken words and arbitrary manual gestures. They reported that 26-month-olds can correctly identify the referent of a gesture-word combination when trained using a multimodal label, but not when trained with a gestural label alone. The authors, however did not include a group of infants who were exposed solely to spoken words, thus, it remains unknown whether the gestural label actually facilitated word learning. Further, it is still unclear whether younger infants at the cusp of acquiring language benefit from gestural labels in multimodal utterances. One question is thus whether preverbal infants fast map multimodal labels, and, when given the choice of two

modalities, whether infants preferentially fast map one better than the other.

In the current study, we investigated how infants interpret utterances containing multimodal object labels. Using an infant eye tracker, we taught infants multimodal labels for two novel objects: Each time a label was produced, infants heard a spoken label coupled with a representational gesture. At test, infants were presented either with the verbal label, a verbal label accompanied by a gestural label, or only a gestural label. Measuring their looks to the objects, we assessed whether infants had made the correct object-label associations. Following Schafer and Plunkett (1998), we used a two-label procedure which is a more ridged method of establishing word learning, and which has been shown to be more reliable for looking than reaching measures (Gurteen et al., 2011). We expected that if infants would be able to map a word to a referent, they would look above chance to the referent. If representational gestures reinforce the association between the label and its referent, then infants should fast map better when both the representational gesture and the word are available at test. However, if infants do not know how to interpret the representational gestures, then infants should perform better when the word is decoupled from the gesture.

## MATERIALS AND METHODS

### PARTICIPANTS

Sixty 15-month-old infants participated in this study (mean age: 15; 14; 30 males and 30 females). Infants were randomly assigned to one of three conditions: Word ( $n = 20$ ), Word + Gesture ( $n = 20$ ), and Gesture ( $n = 20$ ). An additional 15 infants were tested but excluded from analysis (three in the Word condition, five in the Word + Gesture condition, and seven in Gesture condition; see *test phase* section). Five infants were excluded for fussiness; seven because less than 50% valid gaze points were recorded during the training (see Results), two due to caregiver interference, and one because his mother reported that he was familiar with one of the novel objects. All infants were recruited through a database, and all received a small gift for their participation. Written consent was obtained from all of the legal caregivers of all participants.

### EXPERIMENTAL SETUP AND PROCEDURE

Infants were seated on the laps of their caregivers approximately 50 cm in front of a Tobii 1750 remote eye tracker, equipped with an infant add-on. The eye tracker records gaze data at 50 Hz and has an average accuracy of 0.5° visual angle and a spatial resolution of 0.25° visual angle. Stimuli were presented on a 17" flat screen monitor. The visual area was 1,280 × 1,024 pixels, and extended over the entire area of the screen. Infants' eye-gaze was calibrated using a nine-point calibration. If fewer than seven points were calibrated successfully, the calibration was repeated. Infants' behavior during the experiment was also recorded with a digital video camera mounted on a tripod below the monitor.

### STIMULI

Stimuli were recorded with a Canon HV 30 camera and edited with Adobe Premiere Pro CS4. The video clips consisted of a training phase and a test phase. During the training phase, an actress sat behind a table with two tall, narrow shelves standing on either side of the table: one to the left and one to the

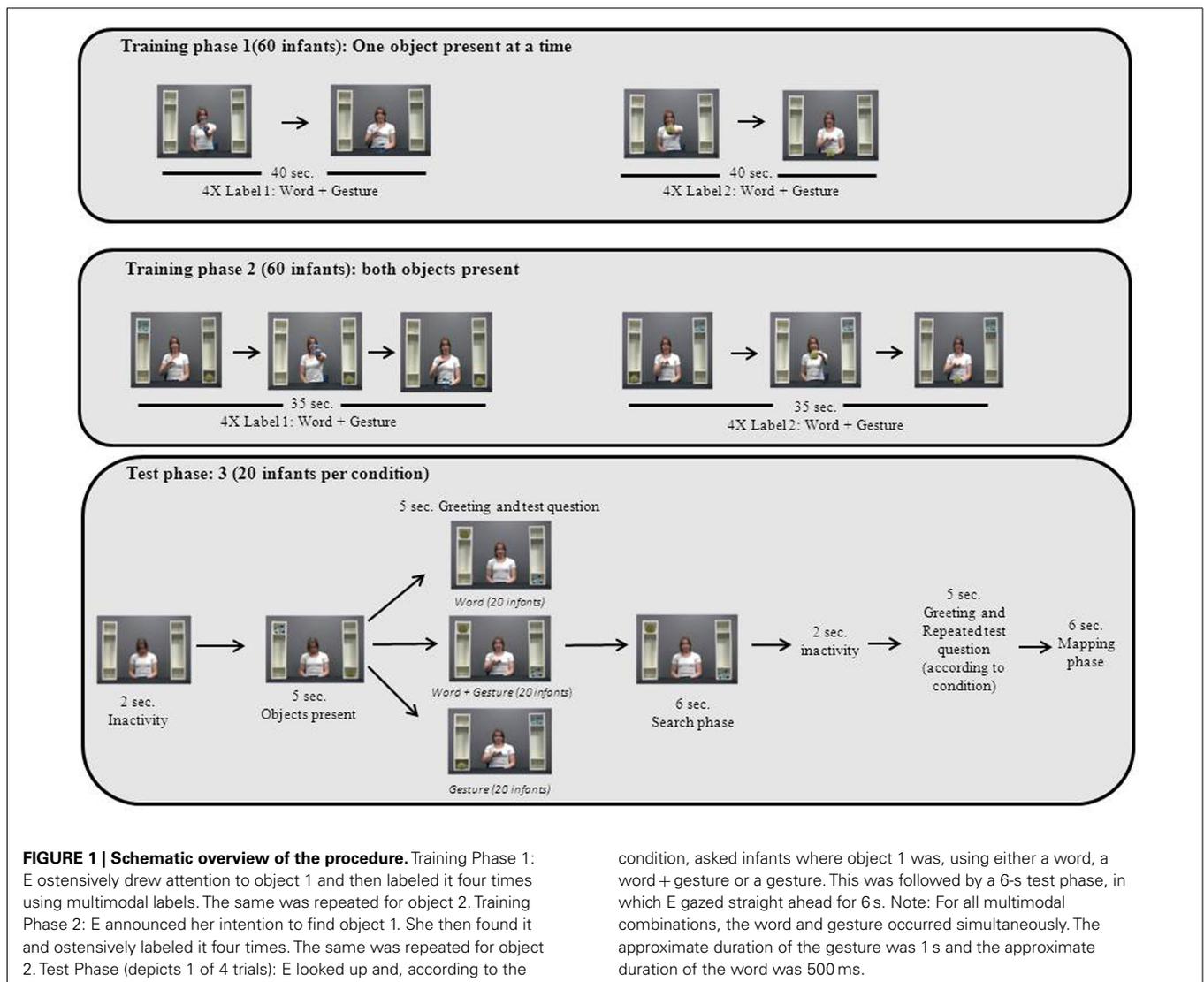
right. Each shelf contained two compartments, within which the objects could sit (see **Figure 1**). The experimenter gazed into the camera, and, after a brief greeting, she proceeded to teach two novel labels for two novel objects, as if the camera was an infant. Each time the actress labeled the objects, she produced both an arbitrary word and an arbitrary manual gesture, presented simultaneously. The words were both one syllable CVC words and had no phonemes in common: /fɪm/ and /ni:p/. The gestures were meant to be within the motor repertoire of a typically developing 15-month-old infant and were based on real contrastive signs in a natural signed language. They were thus easily discernable from one another (roughly YES and NO in American Sign Language). **Table 1** provides a summary of the spoken words, manual gestures, and novel objects used in the study. The experiment was conducted by adhering to the guidelines for good scientific practice by the Max-Planck-Society.

In the training phase, infants were exposed to a total of 16 word + gesture labels, eight for one object and eight for the other object. The training phase was divided into two parts, each of

**Table 1 | Spoken words, manual gestures, and novel objects used in this study.**

Novel words	Novel manual gestures	Novel objects
/fɪm/	 ASL YES	 A
/ni:p/	 ASL NO	 B

Novel objects were counterbalanced so that for half of the infants, /fɪm/ and ASL yes were paired with Object A, and for the other half, /fɪm/ and ASL yes were paired with Object B.



which presented infants with four word + gesture object labels for one object and four word + gesture labels for the other object. In the first part, only one novel object was visible to infants at a time. Using infant directed speech, the actress labeled the first object four times, using both the spoken label and the gesture simultaneously. The labels were embedded in familiar naming contexts in order to facilitate learning (Namy and Waxman, 2000). During the training, the actress ostensibly drew attention to the objects by shifting her gaze between the object and the camera. She said:

“Look!” (gaze to object and back), “A [word plus gesture]! Look here,” (gaze to object and back to camera), “A [word plus gesture]” (gaze to object and back to infant). “This is a [word plus gesture], Wow!” (gaze to object and back to infant), “A [word plus gesture].”

The actress then placed the first object out of sight and presented the second object. Using the same script, she proceeded to teach the label for the second object using the other word and gesture.

The second part of the training was meant to familiarize infants further with the objects' multimodal labels and to introduce them to the task that would later be used in the test phase. Each of the two novel objects sat in a separate compartment and the actress stated that she wanted to find one of the objects saying, “Hmm? Where is the [word plus gesture]?” She then leaned over and gazed into the compartments to search for the object. When she found the appropriate object, she emoted positively, and said: “There it is!” She then took the object out from the compartment, showed it to the camera, and again, using ostensive language and alternating gaze between the camera and the object, labeled it an additional three times with the word and accompanying gesture. She said: “The [word plus gesture]! Look! The [word plus gesture]. Wow, the [word plus gesture].” The same procedure was repeated for the second object, using the other word and sign. In total, infants heard and saw the multimodal object labels eight times for each object during the training phase.

Immediately following the training phase, infants viewed the test phase. Each trial began with the actress looking downward. After 1 s, the two objects each appeared in a separate compartment. Using film editing software, objects were superimposed so that they appeared as if they had actually been present. After 5 s, the actress raised her head, gazed into the camera as if addressing an infant and said: “Hello, where is the [target label]?” In the Word condition, the target label was the spoken word. In the Word + Gesture condition, the target label was the spoken word along with the gesture, and in the Gesture condition, the target label was only the gesture. Following the question, the actress continued to gaze directly into the camera for 6 s, which served as the search phase, in which infants were expected to locate the target object. After the 6-s had elapsed, the same question was repeated. After a further 6 s, an attention getter was displayed in the center of a black screen to bring infants' attention to the middle of the screen (approximately 2 s). Then the second trial started, within which the other label served as the target. In total, there were four trials, each consisting of two questions with the same label thus totaling eight questions, with the target label alternating between the trials, and the order being counterbalanced across participants.

We counterbalanced the positions of the objects based on the two following criteria: First, over the course of all trials, each object appeared in each compartment once. Second, the target object never appeared in the same box for two consecutive trials. All infants received a fixed trial order whereby the positions of the target- and non-target objects across trials were held constant for all infants. We also counterbalanced which object was last seen before the test phase as well as to which object each set of labels was paired. Infants received no feedback with regard to whether they had made the correct mappings.

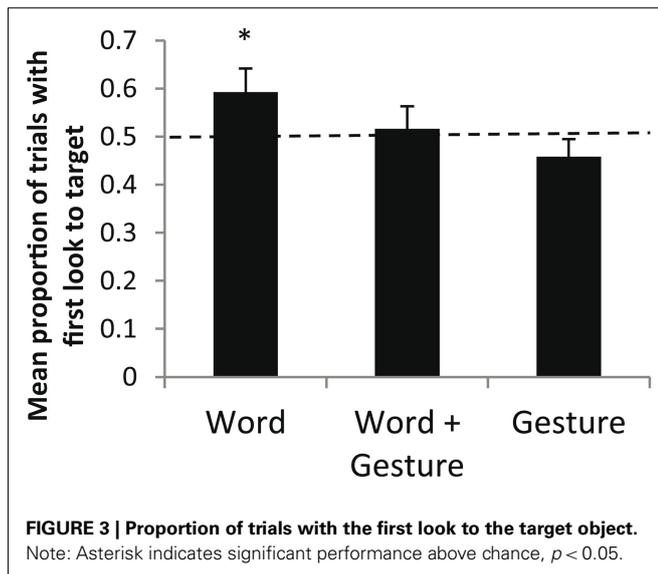
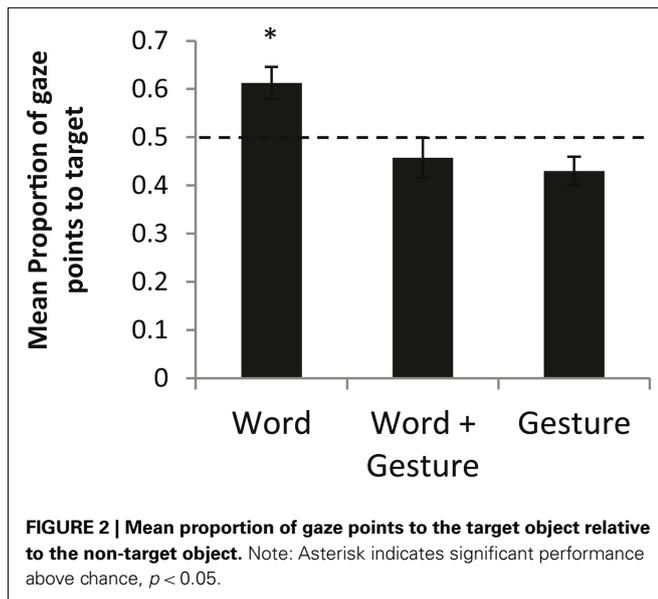
## ANALYSIS AND DATA REDUCTION

Infants who had less than 50% valid gaze points recorded during the training were excluded from analysis. In total, this resulted in the exclusion of seven infants (one in the Word condition, four in the Word + Gesture condition, and two in the Gesture condition). For several infants, visual inspection indicated that there was a shift in the eye tracking data, suggesting that the infant had shifted positions between the calibration and the start of the test. For these infants, an adjustment was made to estimate where the infant was actually looking throughout the duration of the videos (see also Frank et al., in press). We first identified these infants by visual inspection. In order to provide an objective criterion, we also required that their data differed significantly from the norm. To calculate this, we first calculated the mean  $X$  and  $Y$  coordinates of all infants during the attention getters between the test trials. Any infant whose individual mean  $X$  or  $Y$  coordinate differed by more than 2 SD from the overall mean in at least two out of three possible attention getters was selected for an adjustment. The adjustment was computed by calculating the difference between the individual's mean in the  $X$  and  $Y$  axes and the overall mean. This adjustment was then added to each valid recorded coordinate for that particular infant. In total, data from five infants were adjusted (two infants in the Word condition, one infant in the Word + Gesture condition, and two in the Gesture condition).

We created rectangular areas of interest of equal size surrounding both the target and non-target objects following the test questions. For further analyses, two additional areas of interest were created: one encompassing the gesture area surrounding the actress' torso, and another surrounding the actress in full. All AOIs began 200 ms after the relevant behavior, which is the approximate time it takes for infants to program eye movements (Canfield et al., 1997). Our primary dependent measures were (1) the proportion of gaze points to the target object relative to the non-target object, and (2) whether infants' first look was to the target or the non-target object. We used Analyses of Variance to compare performance between conditions, and one-tailed one-sample  $t$ -tests to test whether infants performed above chance.

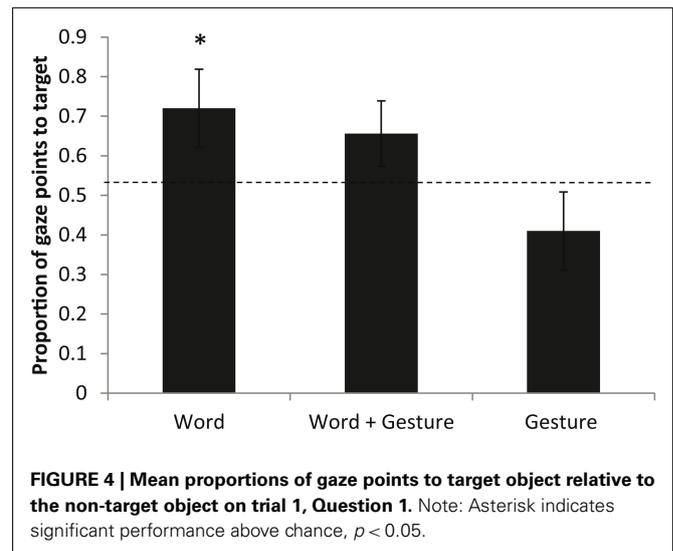
## RESULTS

Figure 2 shows the mean proportion of gaze points that infants looked to the target object, relative to the non-target object during the search period following the test questions. A one-way ANOVA on the mean proportion of gaze points to the target object relative to the non-target object revealed a significant difference between conditions,  $F(2, 57) = 7.83$ ,  $p = 0.001$ , partial  $\eta^2 = 0.22$ . *Post hoc*



tests (two-tailed) using Least Significant Difference comparisons showed that infants in the Word condition looked significantly longer to the target ( $M = 1.8$  ms;  $SD = 1.1$ ) object than infants in both the Word + Gesture condition (Mean difference =  $-0.16$ ,  $p = 0.003$  Cohen's  $d = 0.45$ ;  $M = 1.7$  ms,  $SD = 1.9$ ) and in the Gesture condition (Mean difference =  $-0.18$ ,  $p = 0.001$ , Cohen's  $d = 0.64$ ;  $M = 1.1$  ms,  $SD = 1.5$ ). Infants in the Word + Gesture and Gesture conditions did not differ significantly from one other (Mean difference =  $0.03$ ,  $p = 0.58$ , ns). One-sample  $t$ -tests revealed that only infants in the Word condition looked significantly longer to the target object than would be expected by chance [Word condition:  $t(19) = 3.37$ ,  $p = 0.002$ ; Word + Gesture condition:  $t(19) = -1.03$ ,  $p = 0.841$ ; Gesture condition:  $t(19) = -2.33$ ,  $p = 0.985$ ; all one-tailed probabilities for the upper tails].

**Figure 3** shows the mean proportion of infants' first looks to the target object relative to the non-target object across



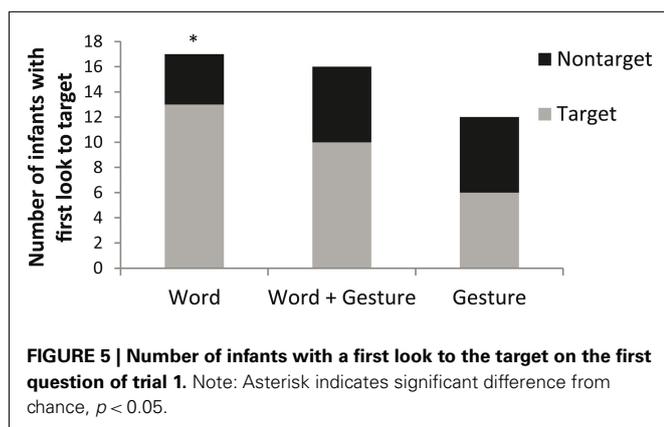
the eight test questions. A one-way ANOVA did not reveal significant differences between conditions [ $F(2, 56) = 2.098$ ,  $p = 0.13$ ], however the overall pattern followed that of the looking time measure. Infants in the Word condition looked significantly more often first to the target object than would be expected by chance [ $t(19) = 2.008$ ,  $p = 0.030$ ], while infants in the Word + Gesture condition and Gesture condition did not [respectively,  $t(19) = 0.333$ ,  $p = 0.372$ ;  $t(19) = 0.920$ ,  $p = 0.816$  one-tailed probabilities for the upper tails].

Finally, we analyzed infants' performance on the first question of the first trial since it avoids any possible biases stemming from switching labels, objects moving positions, repeated questioning, or fatigue. A one-way ANOVA on the mean proportion of gaze points to the target object relative to the non-target object on the first question of the first trial revealed a marginally significant difference between conditions,  $F(2, 42) = 2.73$   $p = 0.077$  (see **Figure 4**). One-tailed  $t$ -tests again confirmed that only infants in the Word condition performed above chance [Word:  $t(16) = 2.698$ ,  $p = 0.008$ ; Word + Gesture condition:  $t(15) = 1.582$ ,  $p = 0.068$ ; Gesture  $t(12) = -0.869$ ,  $p = 0.799$ ; all one-tailed probabilities for the upper tail].

To look at infants' individual performance, we computed binomial tests to compare the number of infants who first looked to the target with the number of infants who first looked to the non-target on the first question of the first trial. Only in the Word condition, was there a significant difference, indicating that only infants in the Word condition had made the correct mappings (13 of 17 infants;  $p = 0.049$ , two-tailed; see **Figure 5**).

#### ADDITIONAL ANALYSES

It is possible that infants did not make the correct mappings in the Gesture condition because they did not attend to the gestures, either in the training phase and/or during the test phase. Indeed, during the training phase, infants focused their attention on an average of 11.2 of 16 possible gestures (70%), with no significant difference between the gestures [5.55 for ASL YES and 5.65 for gesture ASL NO;  $F(1,$



57) = 1.56,  $p = 0.21$ ], and no significant difference between conditions,  $F(2, 57) = 1.56$ ,  $p = 0.21$ . There were, however, no significant correlations between how many signs infants attended to in the training phase and their performance in the test phase. When analyzing each condition separately, the correlations remained non-significant Word:  $r = 0.108$ ,  $p = 0.650$ ; Gesture:  $r = 0.167$ ,  $p = 0.482$ ; Word + Gesture  $r = -0.043$ ,  $p = 0.856$ .

During the test phase, infants in the Gesture condition looked to the gesture area for almost all of the questions ( $M = 7.67$  questions;  $SD = 0.594$ ). A significant one-way ANOVA on the mean number of test questions in which infants attended the sign space [ $F(2, 52) = 16.65$ ,  $p < 0.001$ , partial  $\eta^2 = 0.39$ ] revealed that infants in the Gesture condition looked to the gesture area in significantly more test questions than infants in the Word condition ( $M = 4.58$ ,  $SD = 2.17$ ; Mean difference = 3.09,  $p < 0.001$ ) and marginally more than those in the Word + Gesture condition ( $M = 6.61$ ,  $SD = 1.75$ ; Mean Difference = 1.06,  $p = 0.062$ ). Infants in the Word + Gesture condition also looked at the gesture area in significantly more trials than infants in the Word condition (Mean difference = 2.03,  $p < 0.001$ ). Thus, the gestures clearly elicited infants' attention in the test phase.

A further analysis also confirmed that infants, in the test phase, attended more to the gesture area in the conditions with gestures than in the Word condition. A significant one-way ANOVA on the mean proportion of gaze points per question to the gesture area [ $F(2, 52) = 17.83$ ,  $p < 0.001$ , partial  $\eta^2 = 0.406$ ] relative to the total amount of gaze points revealed that infants looked to the gesture area significantly more in the Gesture condition ( $M = 0.23$ ,  $SD = 0.12$ ), than in both the Word + Gesture condition (where the face and the gesture were competing for visual attention;  $M = 0.14$ ,  $SD = 0.11$ ; Mean difference = 0.09,  $p = 0.007$ ), and the Word condition (where there was no gesture;  $M = 0.04$ ,  $SD = 0.05$ ; Mean difference = 0.19,  $p < 0.001$ ), and significantly more in the Word + Gesture condition than in the Word condition (Mean difference = 0.10  $p = 0.003$ ). Taken together, the additional analyses indicate that (1) in the training phase, infants in the Gesture condition attended to fewer gestures, while they had presumably heard all the auditory labels, and (2) in the test phase, infants allocated more attention to the gesture area in the two conditions where a gesture was present.

## DISCUSSION

In the current study, infants were taught multimodal labels for objects. Results from the test phase showed that infants were able to map the verbal label to the referent. However, when the verbal label was accompanied by a representational gesture, or when the representational gesture was presented alone, infants did not succeed in identifying the correct referents. These findings suggest that when mapping labels to objects, infants initially rely more on the verbal reference of multimodal utterances than on representational gestures, and further, that accompanying representational gestures may interfere with making the correct mappings.

Why did infants in the Gesture condition fail to correctly identify the referents at test? It is unlikely that infants lack the general capacity to integrate information of two modalities. For example, 12-month-olds appropriately process multimodal speech–gesture utterances that include *deictic* gestures (Gliga and Csibra, 2009). Further, when deictic gestures and spoken words are put in competition with one another, infants sometimes even prefer gestures over spoken words as referential cues (Grassman and Tomasello, 2010). One possible explanation for why infants failed in the Gesture condition is that infants do not sufficiently attend to *representational* gestures in multimodal utterances. Indeed, results from our additional analyses showed that infants, in the training phase, attended to only 70% of the gestures. Representational gestures, by the very nature of the visual modality, require that infants divide their visual attention between the referent and the gesture. This, of course, is not the case for spoken words as infants can visually attend to a referent and hear its spoken label simultaneously. Neither is this the case for deictic gestures since they direct attention away from the gesture, and to the referent, thus facilitating the mapping of an auditory label to the attended object. In the current paradigm, infants' visual attention was explicitly directed at the referents with ostensive referential gaze and showing gestures. It is likely that these deictic cues directed infants' attention to the novel objects and thus away from the gestural labels. Since the labels were already provided in the auditory modality, the gestural labels were presumably irrelevant for establishing a mapping.

However, infants did not simply ignore the gestures altogether. First, if infants had ignored the gestures altogether, then they should have performed equally well in both conditions where a word was present at test. This was not the case. To the contrary, infants performed worse when a spoken word was paired with a gesture. Second, our additional analyses show that during the test phase, infants in the Gesture condition indeed attended to the gestures, yet still failed to identify the correct referents. These findings suggest that during the test phase, the gestures actually interfered with infants' ability to identify the referents. Because there were no deictic cues to guide infants' visual attention, it is likely that infants paid more attention to the gestures in the test phase, which ultimately impeded their ability to map the words to the objects. Typically, word learning entails a two-way association between a spoken word and an object. However, when representational gestures are paired with spoken words, as was the case in this study, there is a three-way association since both a spoken word and a representational gesture are mapped onto the object. The three-way association (word–gesture–object) proved to be more difficult for infants to map than the two-way association (word–object). It is possible

that older infants are able to form these associations since they have more experience with spoken language. This would explain why the 26-month olds in Wilbourn and Sims (in press) made the mappings but the 15-month-olds in the current study did not.

The claim that infants younger than 18 months have no preference for either modality (Namy and Waxman, 1998; Namy, 2001), or that they even prefer the gestural over the verbal modality (Goodwyn et al., 2000), was not substantiated by the current findings. Instead, results clearly support the notion that by 15 months, when confronted with multimodal utterances containing representational gestures, infants rely on the verbal instead of the gestural reference in word learning. The age at which infants form this preference is still a matter of debate. With regard to unimodal utterances, some studies have found a preference for spoken labels arising as early as 6 months of age (Fulkerson and Waxman, 2007) while others suggest that it does not develop until later, for example, around infants' second birthdays (Namy and Waxman, 1998; Woodward and Hoyne, 1999; Namy, 2001). Of course, by 15 months, infants learning spoken languages have already had a great deal of experience with spoken words and may even comprehend some commonly used spoken words as young as 6 months (Tincoff and Jusczyk, 1999; Bergelson and Swingley, 2012). However, it is important to emphasize that typically developing hearing infants – who receive spoken language input – can indeed fast map gestural labels that are unimodal (i.e., gestural labels that are not combined with spoken words; Namy and Waxman, 1998; Namy, 2001). Similarly, research on the acquisition of signed languages suggests that infants who are exposed to signs from birth (e.g., infants born to signing parents) acquire language at the same rate as infants acquiring spoken languages (Petito, 1987; Meier and Newport, 1990; Folven and Bonvillian, 1991; Capirci et al., 1998; Schick, 2005). These studies show that infants can map representational gestures to objects, and it might well be the case that they also have no *a priori* preference for either modality. However, the current study shows that when the label is presented in two modalities simultaneously, representational gestures do not facilitate early word learning. This is because the deictic cues necessary for establishing joint reference take visual attention away from the gestural – but not the auditory – label. Representational gestures may even hinder infants' mapping of a spoken word to its referent when deictic cues are lacking, since representational gestures take attention away from the referent.

A wide range of programs currently promote the use of multimodal speech with young infants. Although the proclaimed benefits of these programs are extensive, empirical evidence supporting the claims is lacking (see Johnston et al., 2005 for a meta-analysis). Results from the current study do not support the claim that representational gestures accompanying spoken words

facilitate early word learning. It is, of course, still possible that early, increased exposure to multimodal labels could influence infants' attentional skills and their processing of the more complex three-way associations of multimodal utterances. More research needs to be carried out which investigates how much exposure is actually necessary for infants to form expectations about word modality, and when the cognitive requirements emerge that enable infants to use spoken words in combination with representational gestures as object labels.

Another important consideration has to do with the nature of the signs themselves. In the current study, we used arbitrary spoken words and manual gestures, which bear no physical resemblance to the actual objects. Research shows that multimodal utterances containing iconic gestures do facilitate word learning in toddlers and adult second language learners (Kelly et al., 2009; Marentette and Nicoladis, 2011). However, infants younger than 26 months routinely fail to recognize gestural iconicity, and presumably would not benefit from multimodal utterances including iconic gestures (Namy et al., 2004; Namy, 2008; Tolar et al., 2008)

Findings from this study show that 15-month-old infants make use of verbal but not gestural references in multimodal utterances containing representational gestures. These findings question whether arbitrary multimodal labels facilitate early word learning at 15 months. It remains speculative whether a facilitative effect of representational gesture–speech combinations on word learning would occur with increased earlier exposure to multimodal representational labels (e.g., through “baby signing”). More likely, the multimodal combinatory use of representational labels places too much cognitive demand on young infants since it requires the formation of multiple associations. If so, representational gesture–speech combinations are likely not facilitative of early language acquisition. A host of research has shown that gestures play a leading role in the acquisition of language (e.g., Iverson and Goldin-Meadow, 2005), however, few studies have explicitly discussed the fact that these facilitative effects are initially attributed primarily to deictic gestures. Research shows that the iconicity of representational gestures is not realized until around 26 months (Namy, 2008), and suggests that infants' use of iconic gestures is mediated by their vocabulary size (Nicoladis et al., 1999). Such findings reinforce the argument that representational gestures become relevant only after representational language is acquired (see Liszkowski, 2010). Thus, while deictic gestures and deictic gesture–speech combinations pave the way for language (Iverson and Goldin-Meadow, 2005), representational gestures presumably do not. Notwithstanding the enormous impact of deictic gesture–speech combinations on language acquisition, current findings suggest that representational gestures play a minor role in early multimodal labeling and word learning.

## REFERENCES

- Acredolo, L., and Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Dev.* 59, 450–466.
- Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Dev.* 62, 875–890.
- Bates, E., Benigni, L., Bretherton, I., Camaioni, L., and Volterra, V. (1979). *The Emergence of Symbols: Cognition and Communication in Infancy*. New York: Academic Press.
- Bergelson, E., and Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proc. Nat. Acad. Sci. U.S.A.* 109, 3253–3258.
- Canfield, R., Smith, E., Breznsnyak, M., and Snow, K. (1997). Information processing through the first year of life: a longitudinal study using the visual expectation paradigm. *Monogr. Soc. Res. Child Dev.* 62, 1–145.
- Capirci, O., Iverson, J. M., Pizzuto, E., and Volterra, V. (1996). Gestures and words during the transition to two-word speech. *J. Child Lang.* 23, 645–673.
- Capirci, O., Montanari, S., and Volterra, V. (1998). “Gestures, signs, and words in early language development,” in *The Nature and Functions of Gesture in Children's Communication*, eds J. M. Iverson and S. E. Goldin-Meadow (Rome: Italian National Council of Research, Institute of Psychology), 45–59.

- Folven, R. J., and Bonvillian, J. D. (1991). The transition from non-referential to referential language in children acquiring American sign language. *Dev. Psychol.* 27, 806–816.
- Frank, M. C., Vul, E., and Saxe, R. (in press). Measuring the development of social attention using free-viewing. *Infancy*. doi: 10.1111/j.15327078.2011.00086.x
- Fulkerson, A., and Waxman, S. (2007). Words (but not tones) facilitate object categorization: evidence from 6- and 12-month-olds. *Cognition* 105, 218–228.
- Gliga, T., and Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychol. Sci.* 20, 347–353.
- Gogate, L. B., Bahrick, L. E., and Watson, J. D. (2000). A study of multimodal motherese: the role of temporal synchrony between verbal labels and gestures. *Child Dev.* 71, 876–892.
- Gogate, L. B., Bolzani, L. H., and Betancourt, E. (2006). Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. *Infancy* 9, 259–288.
- Goodrich, W., and Hudson Kam, C. L. (2009). Co-speech gesture as input in verb learning. *Dev. Sci.* 12, 81–87.
- Goodwyn, S. W., Acredolo, L. P., and Brown, C. (2000). Impact of symbolic gesturing on early language development. *J. Nonverbal Behav.* 24, 81–103.
- Grassman, S., and Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Dev. Sci.* 13, 252–263.
- Gurteen, P., Horne, J., and Erjavec, M. (2011). Rapid word learning in 13- and 17-month-olds in a naturalistic two-word procedure: looking versus reaching measures. *J. Exp. Child Psychol.* 109, 201–217.
- Houston-Price, C., Plunkett, K., and Harris, P. (2005). Word learning wizardry at 1;6. *J. Child Lang.* 32, 175–189.
- Iverson, J. M., Capirci, O., and Caselli, M. C. (1994). From communication to language in two modalities. *Cogn. Dev.* 9, 23–43.
- Iverson, J. M., Capirci, O., Longobardi, E., and Caselli, M. C. (1999). Gesturing in mother-child interactions. *Cogn. Dev.* 14, 57–75.
- Iverson, J. M., and Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychol. Sci.* 16, 367–371.
- Johnston, J. C., Durieux-Smith, A., and Bloom, K. (2005). Teaching gestural signs to infants to advance child development: a review of the evidence. *First Lang.* 25, 235–251.
- Kelly, S. D., McDevitt, T., and Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Lang. Cogn. Process* 24, 313–334.
- Liszkowski, U. (2010). Deictic and other gestures in infancy. *Acción psicológica* 7, 21–33.
- Marentette, P., and Nicoladis, E. (2011). Preschoolers' interpretations of gesture: label or action associate? *Cognition* 121, 386–399.
- Masur, E. F. (1997). Maternal labelling of novel and familiar objects: implications for children's development of lexical constraints. *J. Child Lang.* 24, 427–439.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.
- Meier, R. P., and Newport, E. L. (1990). Out of the hands of babes: on a possible sign advantage in language acquisition. *Language* 66, 1–23.
- 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, 73.
- Namy, L. L. (2008). Recognition of iconicity doesn't come for free. *Dev. Sci.* 11, 841–846.
- Namy, L. L., Campbell, A. L., and Tomasello, M. (2004). The changing role of iconicity in non-verbal symbol learning: a U-shaped trajectory in the acquisition of arbitrary gestures. *J. Cogn. Dev.* 5, 37–57.
- Namy, L. L., and Waxman, S. R. (1998). Words and gestures: infants' interpretations of different forms of symbolic reference. *Child Dev.* 69, 295–308.
- Namy, L. L., and Waxman, S. R. (2000). Naming and exclaiming: infants' sensitivity to naming contexts. *J. Cogn. Dev.* 1, 405–428.
- Nicoladis, E., Mayberry, R. I., and Genee, F. (1999). Gesture and early bilingual development. *Dev. Psychol.* 35, 514–526.
- Ninio, A. (1980). Ostensive definition in vocabulary teaching. *J. Child Lang.* 7, 565–573.
- Petito, L. A. (1987). On the autonomy of language and gesture: evidence from the acquisition of personal pronouns in American sign language. *Cognition* 27, 1–52.
- Schafer, G., and Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Dev.* 69, 309–320.
- Schick, B. (2005). *Advances in the Sign Language Development of Deaf Children*. Oxford: University Press.
- Tincoff, R., and Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychol. Sci.* 10, 172–175.
- Tolar, T. D., Lederberg, A. R., Gokhale, S., and Tomasello, M. (2008). The development of the ability to recognize the meaning of iconic signs. *J. Deaf Stud. Deaf Educ.* 13, 225–240.
- Tomasello, M., and Farrar, M. J. (1986). Joint attention and early language. *Child Dev.* 57, 1454–1463.
- Wilbourn, M. P., and Sims, J. P. (in press). Get by with a little help from a word: multimodal input facilitates 26-month-olds' ability to map and generalize arbitrary gestural labels. *J. Cogn. Dev.* doi: 10.1080/15248372.2012.658930
- Woodward, A., and Hoynes, K. (1999). Infants' learning about words and sounds in relation to objects. *Child Dev.* 70, 65–77.
- Zammit, M., and Schafer, G. (2011). Maternal label and gesture use affects acquisition of specific object names. *J. Child Lang.* 38, 201–221.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 February 2012; accepted: 19 March 2012; published online: 03 April 2012.

Citation: Puccini D and Liszkowski U (2012) 15-month-old infants fast map words but not representational gestures of multimodal labels. *Front. Psychology* 3:101. doi: 10.3389/fpsyg.2012.00101  
This article was submitted to *Frontiers in Developmental Psychology*, a specialty of *Frontiers in Psychology*.  
Copyright © 2012 Puccini and Liszkowski. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.