INFANCY

THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES

Infancy, ***24***(3), 455–476, 2019

ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/infa.12278

WILEY Blackwell

The Signal in the Noise: The Visual Ecology of Parents' Object Naming

Sumarga H. Suanda
Department of Psychological Sciences, the Connecticut Institute for the Brain and Cognitive Sciences
University of Connecticut

Meagan Barnhart, Linda B. Smith, and Chen Yu
Department of Psychological and Brain Sciences
Program of Cognitive Science
Indiana University

The uncertainty of reference has long been considered a key challenge for young word learners. Recent studies of head camera wearing toddlers and their parents during object play have revealed that from toddlers' views, the referents of parents' object naming are often visually quite clear. Although these studies have promising theoretical implications, they were all conducted in stripped-down laboratory contexts. The current study examines the visual referential clarity of parents' object naming during play in the home. Results revealed patterns of visual referential clarity that resembled previous laboratory studies. Furthermore, context analyses show that such clarity is largely a product of manual activity rather than the object naming context. Implications for the mechanisms of early word learning are discussed.

The uncertainty of reference is a central idea in early word learning research. The notion that toddlers must sift through many candidate word-to-world mappings whenever they encounter a new word is the theoretical backbone behind experimental studies on how toddlers constrain the mapping space (e.g., Golinkoff & Hirsh-Pasek, 2006), observational research on how parents reduce uncertainty (e.g., Masur, 1997), and computational analyses on just how much uncertainty different learning systems can handle (Blythe, Smith, & Smith, 2016). Recently, Smith, Yu, and their colleagues have argued that the problem of referential uncertainty may have been overestimated (Pereira, Smith, & Yu, 2014; Yu & Smith, 2012; Yurovsky, Smith, & Yu, 2013). Through a series of studies employing mini head cameras worn by toddlers, they demonstrated that when parent object naming is viewed from the toddler learner's

Correspondence should be sent to Sumarga H. Suanda, Department of Psychological Sciences, University of Connecticut, Storrs, CT 06269. E-mail: s.suanda@uconn.edu

perspective, many times the referent of parents' object naming is hardly ambiguous, suggesting that the starting assumptions for many of our accounts of word learning may be inaccurate. Although the free-flowing play observed in these recent head camera studies mimicked toddlers' everyday play, it took place in an unnatural laboratory context, raising legitimate concerns about whether the conclusions would generalize to messier real-life environments (e.g., de Barbaro, Johnson, Forster, & Deak, 2013; Trueswell et al., 2016). The current study tests the generalizability of these laboratory-based findings by examining the nature of toddlers' views of objects during play in their homes and by investigating the processes that undergird those views. Examining these issues sharpens our understanding of the nature of the input for toddlers' word learning, has implications for many key debates in word learning research (e.g., the role of top-down versus bottom-up processes; Hoff & Naigles, 2002; Masur, 1997; Yu & Smith, 2012), and raises new questions about the constellation of attentional, motor, and visual processes that shape the input.

Parent object naming: The toddler's view

In Smith, Yu, and colleagues' studies, parents and their toddlers were equipped with head cameras and were observed as they played with, and as parents talked about, a set of objects (e.g., Yu & Smith, 2012). The key finding most relevant to the present study is that when head camera images during moments of parent object naming were analyzed, named objects often dominated toddlers' fields of view (FOV) by occupying a larger portion of those views than non-named objects (see Figure 1a; Pereira et al., 2014; Yu & Smith, 2012). Smith, Yu, and colleagues argued that this clarity is a result of the small visuomotor workspace that comes with toddlers' shorter arms such that the objects that they pick up and play with are close to the body and the eyes (see Yu & Smith, 2012; Yu, Smith, Shen, Pereira, & Smith, 2009). In addition, social partners often bring objects close to infants to show and to give to them (see Brand, Baldwin, & Ashburn, 2002; Clark & Estigarribia, 2011; Gogate, Bahrick, & Watson, 2000; Rader & Zukow-Goldring, 2012). The proximity of these objects to the body has consequences for the composition of objects in toddlers' FOV: (1) Focal objects have image sizes that are much larger than non-focal objects; and (2) focal objects often occlude or partially occlude non-focal objects. In brief, toddlers' bodies and associated visuomotor processes create a field of view that is much less cluttered and thus a visual experience when parents name objects that may be much more referentially clear. The implication of this finding is that it paints a picture of the environment that is more conducive for acquisition than often assumed (see Pereira et al., 2014; Yu & Smith, 2012; Yurovsky et al., 2013).

Referential clarity in toddlers' views: Do the findings scale?

Although these findings have promising implications for the role, or lack thereof, of referential uncertainty in word learning, the promise is mitigated by the context in which the findings were observed. All of these studies took place in a stripped-down setting: Parents and their toddlers played with a few laboratory-constructed objects while sitting across from one another at a table in a bare laboratory room (see Figure 1b). Intuitively, everyday learning takes place in a context very different from this contrived setting (Medina, Snedeker, Trueswell, & Gleitman, 2011; Trueswell et al.,

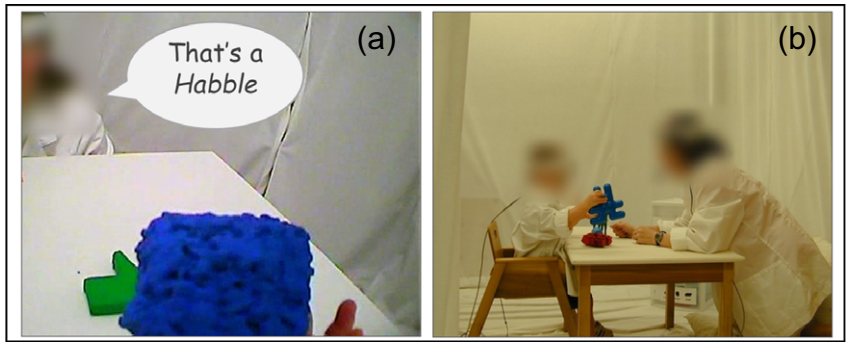

Figure 1 (a) Example of a visually clear referent in toddler's views during parent–toddler play in the laboratory. (b) Laboratory setting and play context in which referential visual clarity findings have been observed (Pereira et al., 2014; Yu & Smith, 2012).

2016). Previous studies that have explicitly compared parent–infant interactions in the home to those in the laboratory do indeed show key differences (Belsky, 1980; Stevenson, Leavitt, Roach, Chapman, & Miller, 1986). Compared to in the home, parents in the laboratory talked more (Stevenson et al., 1986; see also Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017), were more attentive (Belsky, 1980), and were more responsive to their children's behavior (Belsky, 1980). Additionally, although parents were not explicitly instructed to teach their toddlers the object names in these studies, parents were provided with and taught the set of novel names (e.g., "dodi") ahead of time. It is possible that the novel object–novel name context contributed to how and when parents named objects. In fact, research has documented that when parents introduce novel object names to their language learning toddlers, they rely on a suite of referential, semantic, and syntactic strategies (see Bird & Cleave, 2016; Clark, 2010; Cleave & Bird, 2006; Henderson & Sabbagh, 2010; Masur, 1997). Altogether, existing studies documenting laboratory and novel object effects on parent behavior underscore the need to test whether previous findings of visual referential clarity can generalize beyond these contexts and to contexts that more closely reflect toddlers' everyday experiences.

Current study

The goal of the current study was to test the generalizability of visual referential clarity in toddlers' experiences and to better understand the processes that underlie it. To test generalizability, we asked whether toddlers would experience visual referential clarity in free-flowing object play in the home with a set of common toy objects. If visual referential clarity is largely an artifact of the stripped-down laboratory context, then we should observe it less readily in the current study. In contrast, if visual referential clarity is due to toddlers' unique visuomotor experiences, as originally suggested (see Smith, Yu, & Pereira, 2011; Yu et al., 2009), then we should expect to observe such clarity even in the current study. To delve deeper into the processes that underlie visual referential clarity, we investigated the specific contexts in which visual clarity was observed. We first asked whether visual referential clarity was contingent on the context of object naming. To the extent that visual referential clarity is largely driven by

parents isolating optimally clear moments for object naming, we should expect different degrees of visual clarity during versus outside moments of object naming. In contrast, if visual referential clarity is in large part a product of toddlers' small visuomotor workspace, and how toddlers' and their social partners' actively shape that workspace, then we should expect visual clarity to not be contingent on parents' naming. Instead, we should expect visual clarity to be much more contingent on toddlers' and parents' manual actions. Thus, in our final analysis, we investigated how toddlers' views of objects were related to toddlers' and their parents' manual actions.

The approach we took in this study was corpus-based. That is, we collected video recordings of a small number of toddlers ($N = 5$), extracted video frames from those recordings at a relatively high resolution (1 frame/sec; 1 Hz), and manually annotated properties of all frames collected across all toddlers ($N = 3{,}866$ frames). Thus, the current approach mirrors efforts in language, motor, and social development research that employ small sample sizes but involve analyzing high-density data (Demuth & McCullough, 2009; Franchak, Kretch, Soska, & Adolph, 2011; Ninio & Bruner, 1978; Roy, Frank, DeCamp, Miller, & Roy, 2015; Thelen, 1986; Yoshida & Smith, 2008; Yu & Smith, 2012). A limitation of this approach is its potentially limited generalizability and that it does not speak to issues of inter-individual variability (see also Roy et al., 2015). We return to these issues in the *General Discussion*. To provide a glimpse of how each toddler's data conformed to the corpus-level patterns, we highlight subject-level means in all figures and present subject-level analyses in Appendix A.[1]

METHODS

Corpus

The corpus included audio and video recordings of five mother–toddler dyads as they played with a set of common toy objects in their homes. Table 1 describes the participants, their play session details, and the amount of data that they contributed to the corpus. Two additional dyads agreed to participate in the original study but did not contribute data due to toddlers' unwillingness to wear the head camera equipment. This research was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before data collection. All procedures in this study were approved by the Human Subjects and Institutional Review Boards at Indiana University.

Equipment

During play, all toddlers wore headgear (a padded protective helmet) that was fixed with a small lightweight head camera (see Figure 2a). The head camera was from Positive Science and had a 100^0 diagonal field of view (see Franchak et al., 2011). The head camera was wired to a small camcorder (Sony Handycam DCR-HC62) that recorded the video footage from the camera. The headgear and cap weighed approximately 50 g. The camcorder and a battery pack powering the head camera was placed in a toddler-worn backpack ($9.5'' \times 7'' \times 4''$). The backpack and its content were light

[1]Although subject-level data do not speak to the issue of generalizability of the data, they do speak to the robustness of the results observed in this sample.

TABLE 1
Participant Demographics and Play Session Details

	Toddler		*Number of objects*		*Play time analyzed (min)*			*Naming utterances*		
	Age	*Gender*	*Set 1*	*Set 2*	*Set 1*	*Set 2*	*Total*	*Set 1*	*Set 2*	*Total*
1	25 months	M	9	8	7.9	10.3	18.2	33	40	73
2	21 months	F	10	8	6.8	9.0	15.8	63	36	99
3	19 months	M	9	8	7.3	8.5	15.8	53	51	104
4	18 months	M	10	8	4.5	4.9	9.4	18	22	40
5	20 months	F	9	–	5.2	–	5.2	19		19
		Corpus totals					64.4			335

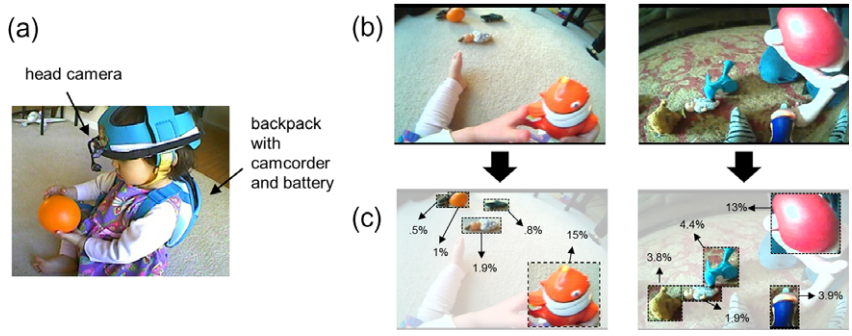

Figure 2 (a) Toddler wearing head-mounted camera, camcorder, and battery pack. (b) Sample frames obtained from toddler head camera. (c) Sample object image size coding; percentages reflect the percentage of toddlers' field of view taken up by the bounding box drawn around each object.

enough (approximately 275 g) to allow toddlers to move around the room if they chose to. Parents also wore headgear consisting of a head camera (allowing for an additional angle to code behavior) and a hands-free professional-quality microphone (ATM75 Cardioid Condenser Microphone from Audio-Technica).

Stimuli

There were two sets of commercially-available toy objects (see Appendix B).[2] All toy objects were small enough for toddlers to pick up and grasp.

Procedure

After toddlers were fitted with the head camera, they played with their mothers in their living rooms. Mothers were instructed to play with their toddlers as they naturally

[2]The two sets of objects differed in familiarity as determined jointly by normed vocabulary data and by parent report. Differences in the set sizes across participants were because some toddlers were familiar with objects that other toddlers were not. Because all analyses did not differ between the two object sets, the current analyses were based on data combined across sets.

would. The play session was divided into two periods of play, lasting up to 10 min. If toddlers became restless and uninterested with our toy set before the 10 min were up (e.g., leaving their living room, starting to play with their own toys), we cut the play period short. Brief interruptions of play (e.g., moments when a sibling or pet came through the play space, head camera adjustments) were marked in the video and excluded from any analysis. At the beginning of each period, an experimenter gave mothers a small box containing one of the toy sets. One toddler became fussy after one set and refused to continue to wear the head camera. This dyad contributed only one set of data. On average, each toddler contributed 12.9 min of data to analyze (see Table 1). During play, the experimenters were out of view in a hallway or in an adjacent room.

Data processing and coding

Speech transcription

A primary coder transcribed mothers' speech during play and divided speech into utterances, defined as strings of speech between two periods of silence lasting at least 400 msec. Utterances containing a name of one of the toy objects (e.g., "Is that a *penguin*?") were then marked as "naming utterances." Mothers produced a total of 335 naming utterances. A second trained coder completed speech transcription and coding of naming utterances for one randomly selected dyad. We computed, frame by frame, the reliability of the timing and the referent of naming utterances using Cohen's kappa. Reliability was considered high (.96) based on conventional guidelines (Bakeman & Gottman, 1997).

Object image size coding

Head camera video footage was sampled at a rate of 1 frame/sec.[3] Across all dyads, there were 3,866 frames coded. Using an in-house coding program, the image size of each object in toddler head camera images (on average, there were 4.4 objects in each frame) were annotated by a trained coder frame by frame. Image size coding was done by drawing a bounding box around each object in view (see Figure 2).[4] An object's image size was derived from computing the area of the bounding box divided by the area of the entire image. We then multiplied this value by 100, yielding a measure of an object's image size that reflects an estimate of the percentage of toddlers' FOV taken up by that object (see Figure 2). A second trained coder completed object image size coding for a randomly selected toddler. Estimates of object image size by the primary and secondary coders were on average within .2% FOV ($M_{\text{difference}}$ = .18% FOV, SD = .39%). To statistically assess the reliability of image size coding, we conducted a two-way mixed model of single-measure intra-class correlation (ICC; see Hallgren, 2012; McGraw & Wong, 1996). The ICC, which was based on absolute

[3]Our pilot work suggested that this sampling rate was sufficient to capture the visual information focused on in the current study.

[4]The bounding box method is common in computer vision research (Pirsiavash & Ramanan, 2012; Vondrick, Patterson, & Ramanan, 2013). Although there are a number of ways to extract visual object information, each with their pros and cons, our piloting work suggested that the different methods yielded very similar results for the types of analysis we are interested in.

agreement, was computed using the *IRR* package in *R* (version 0.84, Gamer, Lemon, Fellows, & Singh, 2015). The resulting ICC (.98) was in the excellent range (Cicchetti, 1994), indicating that the coders had a high degree of agreement and that object image size was estimated similarly across coders.

Manual activity coding

Trained coders also watched the play session frame by frame from both the toddler and parent head cameras and scored when parents and their toddlers touched each object. Reliability coding of toddler and parent manual activity was done for a randomly selected dyad. Reliability of manual action coding, as determined by Cohen's kappa, was high (toddler touch: .92; parent touch: .92).

Statistical analyses

Our primary analytic approach was to use mixed-effects regression models (Pinheiro & Bates, 2000). Models were implemented in *R Studio* (version 0.98.1103) using the *nlme* package (version 3.1-128; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2016).[5] In all models, subjects were considered random effects. Where means and 95% confidence intervals are reported, the mean reflects corpus-level grand means and 95% confidence intervals of those means were obtained via bootstrap resampling of observed data in a way that stayed true to the nested structure of the data (i.e., different subjects contributed different number of events/frames).

RESULTS

Visual referential clarity of naming utterances

To investigate whether naming utterances possessed visual referential clarity, we compared the image size of the named object to the average image sizes of all non-named objects that were part of the relevant object set. Because naming utterances were often longer than 1 sec (M = 1.77s, SD = 1.01s) and thus spanned multiple frames, to derive the visual object properties during naming, we computed the mean image size of objects across all frames falling within the utterance. In addition to comparing the image size of the named object to the mean image sizes of non-named objects, we also compared the image size of the named object to the image size of *that* object outside of naming contexts (e.g., the image size of the toy car when the parent uttered the word "car" to the mean image size of the toy car when parents were not naming). To compute the image size of objects outside of naming contexts, we simply averaged the image size of that object across all frames when no objects were named.[6]

[5]We also performed these models using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015), which from our reading of the developmental science literature is the more common package to perform these models. We chose the *nlme* package because it readily provided the statistical significance of model terms. All coefficient and variance estimates computed via the two packages were identical.

[6]In computing the mean image size of an object when not named, we also considered a version of the analysis that looked at only the moments when that object was not being named (as opposed to all moments when any object was being named). The two methods did not reveal any reliable differences.

Consistent with previous laboratory-based observations, named objects occupied a greater percentage of toddlers' FOV (M = 5.18, 95% CI = 4.57–5.70) than non-named competitors (M = 1.42, 95% CI = 1.30–1.55; see Figure 3a). Additionally, objects were larger in their image sizes when they were named than when they were not named (M = 2.01, 95% CI = 1.88–2.13). As a statistical test of these claims, we performed mixed-effects analyses on the difference scores between named objects and the average of the non-named competitors, as well as on the difference scores between objects when they were named and the average image size of those objects when they were not named. Of interest was the extent to which the intercept term in these models (*image size difference* ~ 1, *random* = ~1|*subject*) was statistically different from zero, suggesting that there was a reliable difference between named and non-named objects at parent naming moments, and between objects when they are named to when they are not named. Results of these analyses (named versus non-named objects: $M_{\text{diff.}}$ = 3.76, 95% CI = 3.15–4.31; B = 3.67, SE = .58, t = 6.27, p < .001; named object versus object when not named: $M_{\text{diff.}}$ = 3.17, 95% CI = 2.68–3.75; B = 3.05, SE = .55, t = 5.55, p < .001) confirmed what Figure 3a depicts visually: When objects were named during play in the home, they were visually more dominant than non-named competitor objects. Additionally, objects were visually more dominant when they were named than when they were not named. Importantly albeit numerically small, the differences in image size observed here at home are larger than those previously observed to be associated with object name learning in the laboratory (about 2.2% in Yu & Smith, 2012; about 1.5% in Pereira et al., 2014).

Because the above analysis considered all objects that were part of the playset (i.e., both objects that were in view and those that were out of view), the visual dominance of named objects compared to non-named competitors could be the result of two non-mutually exclusive visual properties of naming utterances. First, it could be that even among the objects that were in view, the named object was visually more dominant

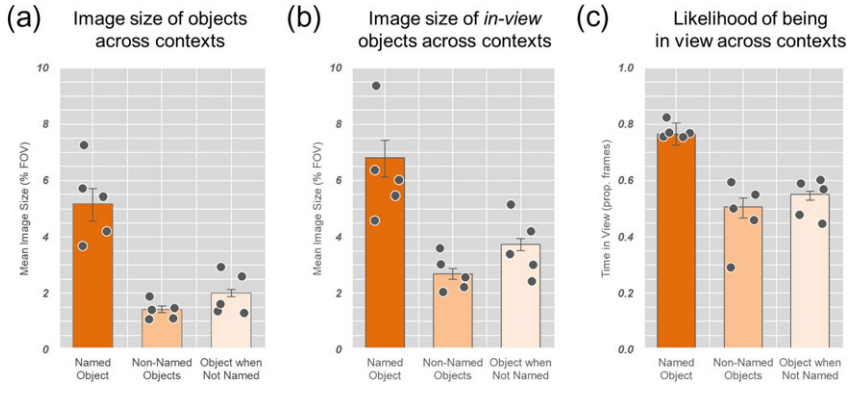

Figure 3 Referent visual clarity in toddlers' views as parents named objects. (a) Mean image size of the named objects and non-named objects during parent naming moments, as well as the mean image size of those named objects outside of naming moments. (b) Mean image size of in-view named objects and in-view non-named objects during parent naming moments, as well as the mean image size of in-view named objects outside of naming moments. (c) Mean proportion of time named objects and non-named objects was in view during parent naming moments, as well as the mean proportion named objects was in view outside of naming moments. Bars represent group means, dots represent individual subject's means, and error bars represent 95% confidence intervals.

than the non-named—but in view—competitors. A second possibility, however, is that the above result could more simply reflect the fact that named objects were more likely to be in view than non-named objects. Thus, the lower average image size of non-named objects could be driven primarily by the fact that many non-named objects were out of view (and thus were scored as 0% of field of view). Figure 3b,c illustrates how both visual properties are true. When we restricted our analyses to only the objects that were in view, the image sizes of named objects (M = 6.81, 95% CI = 6.12–7.42) were significantly larger than their non-named—but in view—competitors (M = 2.68, 95% CI = 2.49–2.88; B = 3.73, SE = .89, t = 4.17, $p < .001$), and were significantly larger than when they were in view but not named (M = 3.74, 95% CI = 3.49–3.96; B = 2.81, SE = .64, t = 4.43, $p < .001$). Additionally, named objects (M = .76, 95% CI = .72–.80) were more likely to be in view than non-named objects (M = .51, 95% CI = .47–.54; B = .29, SE = .05, t = 5.22, $p < .001$), and were more likely to be in view than when those objects were not named (M = .55, 95% CI = .53–.57, B = .24, SE = .03, t = 7.16, $p < .001$).

Visual referential clarity in naming and non-naming contexts

We next examined the degree to which visual clarity was contingent on the context of parent object naming. To do this, we first divided our corpus into frames that occurred during naming utterances (n = 745) and frames that occurred outside of naming utterances (n = 3,121). For each frame, we ordered objects by their image sizes (depicted in Figure 4a). Figure 4a highlights two key findings from this comparison. First, objects in toddlers' views were clearly not equal in their image size. The largest object in view occupied quite a bit more of toddlers' FOV relative to its closest competitor. Additionally, beyond the top few objects in view, other objects simply did not account for much of toddlers' views at all. These results highlight the selective nature of toddlers' views. Second, the shape of the distribution of objects in toddlers' FOV during naming and during non-naming frames was very similar, suggesting that visually clear object views may reflect a more general feature of toddlers' visual experience during object play rather than a visual property specific to the moments that parents choose to name objects.

Figure 4b, which depicts comparisons between the mean image size of the largest object (or the "focal" object; M_{naming} = 7.43, 95% CI = 7.01–7.90; $M_{\text{non-naming}}$ = 6.71, 95% CI = 6.5–6.94) and the mean image size of other objects in play (M_{naming} = 1.08, 95% CI = 1.02–1.15; $M_{\text{non-naming}}$ = .94, 95% CI = .91–.97), supports the above conclusions. In fact, the difference between focal and other objects across naming ($M_{\text{diff.}}$ = 6.35, 95% CI = 5.94–6.78) and non-naming contexts ($M_{\text{diff.}}$ = 5.76, 95% CI = 5.57–5.99) was not statistically different based on a mixed-effects regression analyses testing the effects of naming status (i.e., whether a frame was during versus outside of naming) on visual dominance (*image size difference ~ naming status, random = ~1| subject*): B = .38, SE = .23, t = 1.62, p = .10. It is possible, and perhaps likely, that with a larger sample size, the difference in focal object visual dominance between naming and non-naming contexts would reach statistical significance. However, the current data suggest that statistically reliable or not, the difference between naming and non-naming contexts is minor compared to the similarities between contexts.

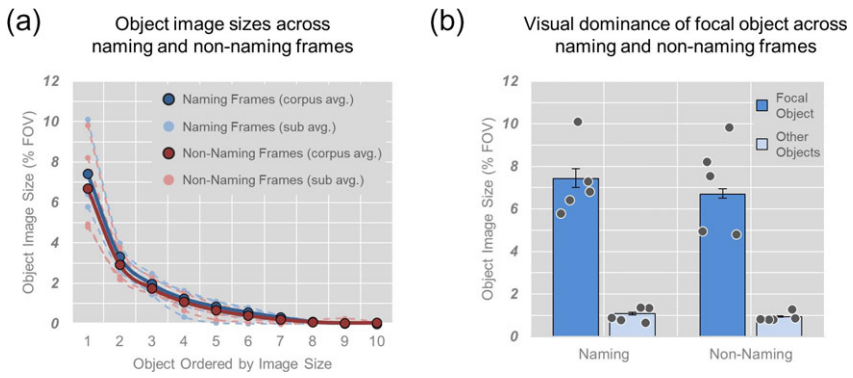

Figure 4 Toddler-perspective visual clarity as a function of naming status. (a) Distribution of object image sizes as a function of naming status. Full lines represent group means, and dotted lines represent individual subject means. (b) Object image sizes of the focal and other objects as a function of naming status. Bars represent group means, dots represent individual subject means, and error bars represent 95% confidence intervals.

The role of toddler and parent manual activity in shaping visual clarity

To examine whether manual activity shaped toddlers' visual clarity, we divided the corpus into three types of frames: frames in which only toddlers held at least one object (*toddler-held* frames; n = 1,477), frames in which only parents held at least one object (*parent-held* frames; n = 498), and frames in which neither toddler nor parent held any objects (*neither-held* frames; n = 828). For this analysis, we excluded frames in which both toddlers and their parents simultaneously held objects (either the same object or different objects; n = 1,063) because: (1) the source of the visual clarity in such frames would be ambiguous, and (2) including such frames would mask potential differences between the effect of toddlers' manual activity and the effect of parents' manual activity.

As Figure 5 illustrates, and in contrast to the negligible effect of parent naming on object visual dominance, toddler and parent manual actions clearly affected visual dominance. When toddlers or parents held objects, toddlers' views of objects had strongly skewed distributions with focal objects (M_{toddler} = 8.37, 95% CI = 8.00–8.73; M_{parent} = 5.73, 95% CI = 5.34–6.07) occupying large amounts of toddlers' FOV relative to other objects (M_{toddler} = 1.05, 95% CI = 1.01–1.09; M_{parent} = .83; 95% CI = .79–.90). When neither toddlers nor parents held objects, objects were more evenly distributed in their image sizes ($M_{\text{focal object}}$ = 2.72, 95% CI = 2.51–2.96; $M_{\text{other objects}}$ = .60, 95% CI = .55–.65). We statistically evaluated the role of manual actions on visual clarity via a mixed-effects regression model that used toddler and parent manual activity status to predict the difference in image size between the focal object and the other objects (*image size difference* ~ *toddler-held* + *parent-held, random* = ~1| *subject*). The model revealed unique roles for both toddler (B = 4.66, SE = .24, t = 19.09, p < .001) and parent manual activity (B = 2.56, SE = .32, t = 8.00, p < .001) on visual dominance. To investigate whether there were any differences between the effects of toddlers' manual actions and the effects of parents' manual actions, we conducted a planned comparison between toddler-held and parent-held frames. In this

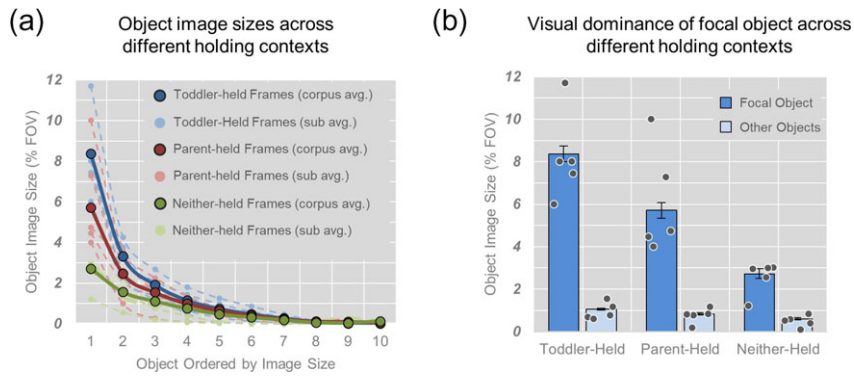

Figure 5 Toddler-perspective visual clarity as a function of manual activity status. (a) Distribution of object image sizes as a function of manual activity. Full lines represent group means, and dotted lines represent individual subject means. (b) Object image sizes of the focal and other objects as a function of manual activity. Bars represent group means, dots represent individual subject means, and error bars represent 95% confidence intervals.

model, we excluded frames when neither held an object and designated toddlers' manual activity as a fixed effect (toddler-held frames were scored as 1, parent-held frames were scored as 0) and subject as a random effect (*image size difference* ~ *toddler-held, random* = ~1|*subject*). Results revealed that toddler-held frames had greater focal object visual dominance than parent-held frames (B = 1.98, SE = .33, t = 5.95, $p < .001$), suggesting that although both toddlers' and parents' actions were associated with views in which one object dominated, the effect of toddlers' actions was more potent.

Finally, to provide deeper insight into the possible causal role manual actions play on an object's visual clarity, we analyzed the image sizes of held objects prior to manual actions, during manual actions, and after manual actions. Figure 6 illustrates the real-time dynamics of an object's image size time-locked to manual activity. The figure reveals that (1) object image sizes were consistently small leading up to the moment toddlers (M = 1.97, 95% CI = 1.73–2.20)[7] and parents (M = 1.64, 95% CI = 1.32–1.96) took hold of objects (Figure 6a,b), (2) object image sizes steeply increased the moment toddlers (M = 4.83, 95% CI = 4.40–5.39) and parents (M = 3.51, 95% CI = 3.13–3.96) took hold of objects (Figure 6a,b; Toddler: B = 2.90, SE = .39, t = 7.48, $p < .001$; Parent: B = 2.08, SE = .51, t = 4.09, $p < .001$), (3) object image sizes remained visually dominant up to the end of the holding event ($M_{toddler}$ = 4.36, 95% CI = 3.95–4.87; M_{parent} = 3.48, 95% CI = 3.03–4.01; Figure 6c,d), and (4) object image sizes dropped precipitously once toddlers (M = 1.52, 95% CI = 1.31–1.77) and parents (M = 2.13, 95% CI = 1.79–2.55) manually disengaged with those objects (Toddler: B = 2.83, SE = .42, t = 6.63, $p < .001$; Parent: B = 1.35, SE = .24, t = 5.56, $p < .001$; Figure 6c,d). The ebbs and flows of an object's image size as it relates to manual activity are strongly suggestive of the active role toddlers and parents play in shaping the toddlers' visual environment (see also Yu & Smith, 2012).

[7]The data reported here were based on the mean image size during 3s windows before and after holding onset and offset. The comparisons reported were based on the difference in mean object image size in the 3s window prior to holding onset (or offset) and the 3s window after holding onset (or offset). Means and analyses that were based on 1s and 5s windows revealed identical trends.

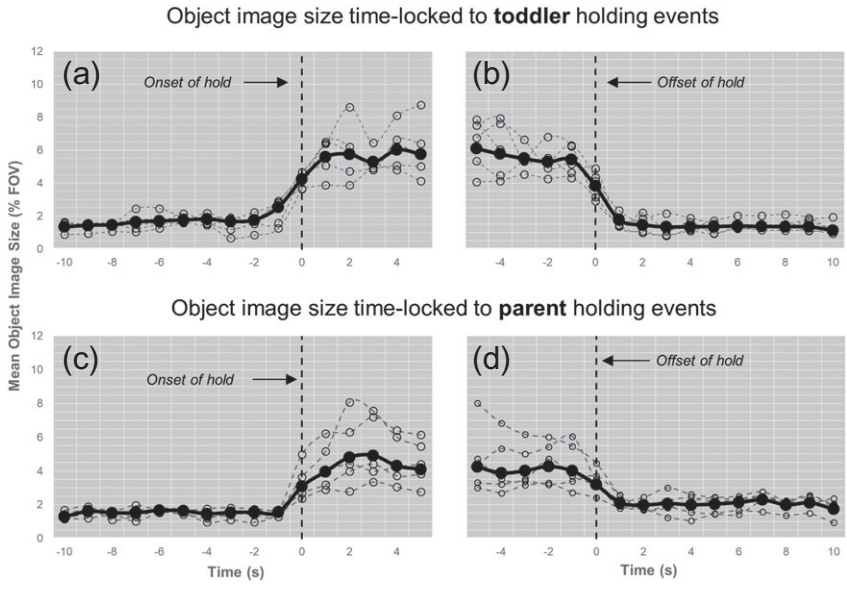

Figure 6 Real-time dynamics of toddler-perspective visual clarity time-locked to the onset and offset of toddler (a, b) and parent (c, d) object-holding events. Dark thick lines reflect group means of the image size of held objects before and after that object was held; dotted lines reflect individual subject means.

GENERAL DISCUSSION

Every day, toddlers are bombarded with many words and many objects. It is commonly assumed that this bombardment creates uncertainty about which words refer to which objects. For the past 40 years, this problem of referential uncertainty has been front and center in early word learning research (see Golinkoff et al., 2000). The present findings add to a growing body of literature that calls for a more careful look at the degree of uncertainty that toddlers actually face (e.g., Pereira et al., 2014; Yu & Smith, 2012; Yurovsky et al., 2013; see also Harris, Jones, & Grant, 1983; Masur, 1997; Messer, 1978). Consistent with recent laboratory-based findings (Pereira et al., 2014; Yu & Smith, 2012), when object naming is viewed from the toddlers' perspective—the perspective that matters most for learning—the referent of many parent object naming moments is often visually clear. These data raise the possibility that the learning task that toddlers face may be less problematic than commonly assumed. The current findings go beyond previous observations in that they show that: (1) referential visual clarity of parent object naming generalizes to contexts that better mirror toddlers' everyday environments, (2) the visual clarity reflects a more general aspect of toddlers' visual experience rather than something specific to object naming moments, and (3) the visual clarity in toddlers' everyday environments is tightly coupled to the actions of toddlers and their social partners. Together, these findings shed light on several current issues in the science of language development: the nature of toddlers' input and the methods employed to study it, the contributions of *visual* input quality to early word learning, and the role of non-linguistic developments (e.g., motor development) on language acquisition.

Visual referential clarity outside of the laboratory

A deeper understanding of the input to toddler word learners is central to many research foci in the word learning literature: the mechanisms of word-referent mapping (Cartmill et al., 2013; Smith, Suanda, & Yu, 2014), the contributions of the environment and the learner to vocabulary development (e.g., Bornstein, 1985), and the sources of inter-individual differences that pervade research on lexical development (e.g., Rowe, 2012). Here, we investigated how one recently discovered aspect of the input—referent visual clarity during parent object naming (Pereira et al., 2014; Yu & Smith, 2012)—scaled outside the laboratory context. We reasoned that if referent visual clarity is to be relevant to everyday learning of object names, then it would be important to know the extent that clarity occurs in contexts that better match toddlers' everyday environments. By documenting referential visual clarity of parent object naming during free-flowing object play outside the laboratory, the current study takes one step toward demonstrating the potential importance of a referent's visual properties for everyday learning. The next steps will be to understand the prevalence of referential visual clarity by expanding the current observations to contexts other than object play, and to understand its contributions to learning by directly examining the link between referential visual clarity in the home and toddlers' vocabulary growth (see *Limitations* below).

By mirroring the results of previous laboratory-based observations (Pereira et al., 2014; Yu & Smith, 2012), the current study adds to a body of evidence that has found parallels between data obtained from structured observations in the laboratory and data obtained from free-flowing observations in the home (Adolph et al., 2012; Bornstein, Haynes, Painter, & Genevro, 2000; Tamis-LeMonda et al., 2017). We suggest that the reason the current home-based findings on referential visual clarity mimic those obtained in the laboratory is because, as our work shows, toddlers' visual ecology is largely determined by the unique physical attributes of toddlers' bodies. For example, toddlers' shorter arms mean that when toddlers interact with objects, objects are naturally very close to toddlers' bodies. These close-to-the-body objects will in turn occupy a larger portion of the field of view and potentially even occlude other objects from view, creating the visually clear experiences we observed. Because these bodily and motor dynamics are constant across laboratory and home contexts, toddlers' visual experiences of objects—so long as they pick them up and manipulate them, and so long as their social partners move objects toward them—will also be largely constant.

Visual input quality and early word learning

When language development researchers write about input quality, they often mean something linguistic: the diversity of parents' speech (e.g., Rowe, 2012), the syntactic complexity of parents' sentences (e.g., Hoff & Naigles, 2002), or the coherence of parents' utterances within the larger discourse (Hirsh-Pasek et al., 2015). The current study, along with other recent work (e.g., Clerkin, Hart, Rehg, Yu, & Smith, 2017; Yu, Suanda, & Smith, 2019), highlights the potential value of thinking about input quality along the *visual* dimension as well. That is, toddlers' visual experiences may be just as relevant to the word–object mapping process as their linguistic experiences. Indeed, experimental studies of object name learning have shown how familiarization with objects (which entails, among other things, extended visual experience with

objects) actually improves the learning and retention of object names (Fennel, 2012; Graham, Turner, & Henderson, 2005; Kucker & Samuelson, 2012).

One contribution of the current study is in suggesting the pervasiveness of high-quality visual experiences with objects. That is, we demonstrate how visual object clarity may not only be a common occurrence when parents name objects, it may be a common occurrence more generally. One working hypothesis is that the pervasiveness of these clear views of objects may support a constellation of processes, including object segmentation (Metta & Fitzpatrick, 2003), object recognition (James, Jones, Swain, Pereira, & Smith, 2014), and object knowledge (Soska, Adolph, & Johnson, 2010), that make rich and robust object representations, which then in turn facilitates the process of mapping words onto those representations. Research that further documents the properties of infants' and toddlers' everyday *visual* experiences may thus prove to be central in attempts to understand early word learning (see Clerkin et al., 2017; Fausey, Jayaraman, & Smith, 2016; Jayaraman, Fausey, & Smith, 2015).

Linking motor processes to language development

Clear object views came about in large part through toddlers' own manual actions (see also Yu & Smith, 2012; Yu et al., 2009). At a broad level of analysis, this finding is consistent with the view that a host of non-linguistic processes are relevant for language development (Iverson, 2010; Smith, 2013). That is, the current study shows how developments in the motor system (including but not limited to developments in fine motor control, hand–eye coordination, posture control) that support mature actions on objects could be critical in creating optimal visual experiences for word learning (see Pereira et al., 2014; Yu & Smith, 2012). Other research suggests that advanced manual abilities may create tactile and multi-modal experiences that are ideal for learning (Chang, de Barbaro, & Deak, 2016; Suanda, Smith, & Yu, 2016; Yu & Smith, 2012). Thus, one pathway by which non-linguistic processes shape word learning is through the quality of the input. Considering the growing body of evidence demonstrating interconnections between non-linguistic and linguistic development in both typical and atypical populations (Collisson, Grela, Spaulding, Ruecki, & Magnuson, 2015; He, Walle, & Campos, 2015; Hellendoorn et al., 2015; James et al., 2014; Leonard, Bedford, Pickles, Hill, & The BASIS Team, 2015; Libertus & Violi, 2016; Oudgenoeg-Paz, Volman, & Leseman, 2016), future investigations that more precisely chart the pathways through which development in non-linguistic domains influence language development (see also Karasik, Tamis-LeMonda, & Adolph, 2014) will be important not just for theory building but also for diagnostic and interventional strategies.

Limitations

There are a number of limitations to the current study that we believe are worth future pursuit. First, although the current study goes beyond laboratory-based results on toddlers' visual ecology (Pereira et al., 2014; Smith et al., 2011; Yu & Smith, 2012), the current results are still constrained to one particular context (i.e., object play) and to a semi-structured setting (i.e., toddlers and parents played for a set amount of time and with a set of experimenter-provided toys). These constraints highlight that the current work represents only one step toward demonstrating the generalizability of previous laboratory-based research to toddlers' true everyday experiences. Research on toddlers'

visual ecology and referential experiences in the range of settings and activities commonly experienced by word learning toddlers (e.g., mealtime, grooming, book sharing) will go a long way toward testing the pervasiveness and limits of the current results (e.g., see Clerkin et al., 2017; Fausey et al., 2016; Hofferth & Sandberg, 2001; Jayaraman et al., 2015; Tamis-LeMonda, Custode, Kuchirko, Escobar, & Lo, in press). A second limitation of the current study is how toddlers' visual environments were measured, as well as which aspects of those environments were considered. As Smith, Yu, colleagues, and others have previously discussed, head camera images are an imperfect approximation of toddlers' visual environment (Aslin, 2008; Schmitow, Stenberg, Billard, & von Hofsten, 2013; Smith, Yu, Yoshida, & Fausey, 2015; Yoshida & Smith, 2008). Their imperfection comes from the fact that head direction, which determines the head camera image, is often but not always coupled with gaze direction and because the viewing angle of head cameras is smaller than toddlers' actual visual fields. Beyond this general limitation of head camera research, the current analyses are limited by the fact that we focused only on the set of objects with which toddlers and parents played. We suggest that this focus may have both underestimated and overestimated the clarity in toddlers' visual experience. That is, on the one hand we likely underestimated the clutter in toddlers' visual fields because we did not also consider any other object that may have been in toddlers' views (e.g., couches, tables, television sets). On the other hand, however, by coding the visual properties of objects that were in view as opposed to coding toddlers' visual attention, we believe we may have also overestimated the clutter that toddlers' attentional system actually processed. In support of this idea, a recent toddler eye-tracking study of cluttered scenes revealed that toddlers' visual attention (i.e., their gaze patterns) was focused on a much smaller subset of objects than what was available in view (Zhang & Yu, 2016). Future research employing head-mounted eye tracking (see Franchak et al., 2011) may provide a different, and perhaps more precise, measure of visual referential uncertainty.

A final limitation of the current study is its small sample. Future research employing larger sample sizes is needed to speak to the generalizability of these results and to issues of individual variability. Although we concede that a larger sample would have made for a more convincing result, we suggest three points in defense of these data. First, the current findings are robust *within* the sample. That is, the three key findings (visual referential clarity during naming, visual dominance outside of naming moments, and the effect of manual actions on object views) reached statistical significance not only at the level of the group but also at the level of individual toddlers (see Appendix A). These supplemental individual-level analyses highlight the statistical strength of the results and underscore that these findings were not shaped by a mere subset of the toddlers. Second, two of the key findings (visual referential clarity during naming and the effect of manual actions on object views) are not one-off results. That is, although the context in which these two findings were observed may be new, there is a sizeable body of evidence demonstrating these phenomena (with some employing sample sizes as large as 100 toddlers, see Suanda et al., 2016; see also Pereira et al., 2014; Yu & Smith, 2012; Yu et al., 2009). Finally, there is precedent for similar small sample research on toddlers' sensorimotor experiences to be reliable and generalizable. For example, in one of the earliest studies to employ toddler-worn head-mounted cameras, Yoshida and Smith observed in five toddlers the surprising finding that parents' faces were rarely present in the toddlers' views. At the time, these results were surprising considering the wealth of research on the role of gaze following and social

referencing in early development (Aslin, 2008). Since Yoshida and Smith's initial observation, the finding of minimal attention to parents' faces during object play has been repeatedly replicated across laboratories, tasks, methods, and sample sizes (see Deak, Krasno, Triesch, Lewis, & Sepeta, 2014; Fausey et al., 2016; Franchak et al., 2011; Yu & Smith, 2013; Yu & Smith, 2017).

CONCLUSION

What's the nature of the data for toddlers' word learning? Is the data noisy and unreliable, suggesting perhaps that the keys to understanding word learning are the powerful top-down cognitive and socio-cognitive mechanisms toddlers employ to filter through the noise? The current study suggests that there may be more signal in the noise than commonly assumed. The implication of this finding is neither that top-down processes do not matter for word learning nor that a high-quality visual signal solves all problems with determining reference. Instead, the implication of this work is to raise the possible need to rethink the role of top-down processes in situ given the nature of the real-world input. For example, in an environment rich with cues to reference, a good bit of learning could transpire despite fragile cognitive and socio-cognitive mechanisms. A deeper understanding of the environment may thus deepen our understanding of which of the many top-down processes are most critical for the developing learner.

ACKNOWLEDGMENTS

This research was supported in part by the National Science Foundation (SBE-BCS0924248, SBE-BCS15233982, SBE-SMA1203420) and the National Institutes of Health (R01-HD074601, K99/R00-HD082358). We thank Damien Fricker and Melissa Elston for assistance in data collection, Hao Lu and Seth Foster for developing the coding software, Sarah Suen, Ashley Meador, and Jessica Steinhiser for coding assistance, and the IU Computational Cognition and Learning Laboratory and the IU Cognitive Development Laboratory for discussion of this work. We are very grateful to the families who participated. The authors declare that there is no conflict of interest.

REFERENCES

Adolph, K. E., Cole, W. G., Komati, M., Garciaguirre, J. S., Badaly, D., Lingeman, J. M., … Sotsky, R. (2012). How do you learn to walk? Thousands of steps and dozens of falls per day. *Psychological Science*, *23*, 1387–1394.

Aslin, R. N. (2008). Headed in the right direction: A commentary on Yoshida and Smith. *Infancy*, *13*, 275–278.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge, UK: Cambridge University Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Belsky, J. (1980). Mother-infant interaction at home and in the laboratory: A comparative study. *The Journal of Genetic Psychology*, *137*, 37–47.

Bird, E. K. R., & Cleave, P. L. (2016). Mothers' talk to children with Down Syndrome, language impairment, or typical development about familiar and unfamiliar nouns and verbs. *Journal of Child Language*, *43*, 1072–1102.

Blythe, R. A., Smith, A. D. M., & Smith, K. (2016). Word learning under infinite uncertainty. *Cognition*, *151*, 18–27.

Bornstein, M. H. (1985). How infant and mother jointly contribute to developing cognitive competence in the child. *Proceedings of the National Academy of Sciences of the United States of America*, *82*, 7470–7473.

Bornstein, M. H., Haynes, O. M., Painter, K. M., & Genevro, J. L. (2000). Child language with mother and with stranger at home and in the laboratory: A methodological study. *Journal of Child Language*, *27*, 407–420.

Brand, R., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, *5*, 72–83.

Cartmill, E. A., Armstrong, B. F., III, Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parental input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 11278–11283.

Chang, L., de Barbaro, K., & Deak, G. (2016). Contingencies between infants' gaze, vocal, and manual actions and mothers' object naming: Longitudinal changes from 4 to 9 months. *Developmental Neuropsychology*, *41*, 342–361.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290.

Clark, E. V. (2010). Adult offer, word-class, and child uptake in early lexical acquisition. *First Language*, *303*, 250–269.

Clark, E. V., & Estigarribia, B. (2011). Using speech and gesture to introduce new objects to young children. *Gesture*, *11*, 1–23.

Cleave, P. L., & Bird, E. K. R. (2006). Effects of familiarity on mothers' talk about nouns and verbs. *Journal of Child Language*, *33*, 661–676.

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*, 20160055.

Collisson, B. A., Grela, B., Spaulding, T., Ruecki, J. G., & Magnuson, J. S. (2015). Individual differences in the shape bias in preschool children with specific language impairment and typical language development: Theoretical and clinical implications. *Developmental Science*, *18*, 373–388.

de Barbaro, K., Johnson, C. M., Forster, D., & Deak, G. O. (2013). Methodological considerations for investigating the microdynamics of social interaction development. *IEEE Transactions on Autonomous Mental Development*, *5*, 258–270.

Deak, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science*, *17*, 270–281.

Demuth, K., & McCullough, E. (2009). The prosodic (re)organization of children's early English articles. *Journal of Child Language*, *36*, 173–200.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107.

Fennel, C. T. (2012). Object familiarity enhances infants' use of phonetic detail in novel words. *Infancy*, *17*, 339–353.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development*, *82*, 1738–1750.

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2015). *irr: Various coefficients of interrater reliability and agreement*. R package version 0.84. Retrieved from http://CRAN.R-project.org/package = irr

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*, 878–894.

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby wordsmith: From associationist to social sophisticate. *Current Directions in Psychological Science*, *15*, 30–33.

Golinkoff, R. M., Hirsh-Pasek, K., Bloom, L., Smith, L. B., Woodward, A. L., Akhtar, N., … Hollich, G. (2000). *Becoming a word learner: A debate on lexical acquisition*. New York, NY: Oxford Press.

Graham, S. A., Turner, J. N., & Henderson, A. M. E. (2005). The influence of object pre-exposure on two-year-olds' disambiguation of novel labels. *Journal of Child Language*, *32*, 207–222.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*, 23–34.

Harris, M., Jones, D., & Grant, J. (1983). The nonverbal context of mothers' speech to infants. *First Language*, *4*, 21–30.

He, M., Walle, E. A., & Campos, J. J. (2015). A cross-national investigation of the relationship between infant walking and language development. *Infancy*, *20*, 283–305.

Hellendoorn, A., Wijnroks, L., van Daalen, E., Dietz, C., Buitelaar, J. K., & Leseman, P. (2015). Motor functioning, exploration, visuospatial cognition and language development in preschool children with autism. *Research in Developmental Disabilities*, *39*, 32–42.

Henderson, A. M. E., & Sabbagh, M. A. (2010). Parents' use of conventional and unconventional labels in conversations with their preschoolers. *Journal of Child Language*, *37*, 793–816.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., … Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, *26*, 1071–1083.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*, 418–433.

Hofferth, S. L., & Sandberg, J. F. (2001). How American children spend their time. *Journal of Marriage and Family*, *63*, 295–308.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, *37*, 229–261.

James, K. H., Jones, S. S., Swain, S., Pereira, A., & Smith, L. B. (2014). Some views are better than others: Evidence for a visual bias in object views self-generated by toddlers. *Developmental Science*, *17*, 338–351.

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PLoS One*, *10*, e0123780.

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2014). Crawling and walking infants elicit different verbal responses from mothers. *Developmental Science*, *17*, 388–395.

Kucker, S. C., & Samuelson, L. K. (2012). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy*, *17*, 295–323.

Leonard, H. C., Bedford, R., Pickles, A., Hill, E. L., & The BASIS Team (2015). Predicting the rate of language development from early motor skills in at-risk infants who develop autism spectrum disorder. *Research in Autism Spectrum Disorders*, *13–14*, 15–24.

Libertus, K., & Violi, D. A. (2016). Sit to talk: Relation between motor skills and language development in infancy. *Frontiers in Psychology*, *7*, 475.

Masur, E. F. (1997). Maternal labelling of novel and familiar objects: Implications for children's development of lexical constraints. *Journal of Child Language*, *24*, 427–439.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30–46.

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 9014–9019.

Messer, D. J. (1978). The integration of mothers' referential speech with joint play. *Child Development*, *49*, 781–787.

Metta, G., & Fitzpatrick, P. (2003). Better vision through manipulation. *Adaptive Behavior*, *11*, 109–128.

Ninio, A., & Bruner, J. (1978). The achievement and antecedents of labelling. *Journal of Child Language*, *5*, 1–15.

Oudgenoeg-Paz, O., Volman, M. C., & Leseman, P. P. (2016). First steps into language? Examining the specific longitudinal relations between walking, exploration and linguistic skills. *Frontiers in Psychology*, *7*, 1458.

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin and Review*, *21*, 178–185.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus*. New York, NY: Springer-Verlag.

Pinheiro, J. C., Bates, D. M., DebRoy, S., Sarkar, D., & R Core Team (2016). *nlme: Linear and nonlinear mixed effects models*. R package version 3.1-128. Retrieved from http://CRAN.R-project.org/package=nlme

Pirsiavash, H., & Ramanan, D. (2012). *Detecting activities of daily living in first-person camera views*. Paper presented at IEEE Conference on Computer Vision and Pattern Recognition. Retrieved from https://doi.org/10.1109/cvpr.2012.6248010

Rader, N. D., & Zukow-Goldring, P. (2012). Caregivers' gestures direct infant attention during early word learning: The importance of dynamic synchrony. *Language Sciences*, *34*, 559–568.

Rowe, M. L. (2012). A longitudinal investigation of the role of the quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*, 1762–1774.

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 12663–12668.

Schmitow, C., Stenberg, G., Billard, A., & von Hofsten, C. (2013). Using a head-mounted camera to infer attention direction. *International Journal of Behavioral Development*, *37*, 468–474.

Smith, L. B. (2013). It's all connected: Pathways in visual object recognition and early noun learning. *American Psychologist*, *68*, 618–629.

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word-referent learning. *Trends in Cognitive Sciences*, *18*, 251–258.

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: the dynamics of toddler visual experience. *Developmental Science*, *14*, 9–17.

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (2015). Contributions of head-mounted cameras to studying the visual environments of infants and young children. *Journal of Cognition and Development*, *16*, 407–419.

Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in development: Motor skill acquisition facilitates three-dimensional object completion. *Developmental Psychology*, *46*, 129–138.

Stevenson, M. B., Leavitt, L. A., Roach, M. A., Chapman, R. S., & Miller, J. F. (1986). Mothers' speech to their 1-year-old infants in home and laboratory settings. *Journal of Psycholinguistic Research*, *15*, 451–461.

Suanda, S. H., Smith, L. B., & Yu, C. (2016). The multisensory nature of verbal discourse in parent-toddler interactions. *Developmental Neuropsychology*, *41*, 324–341.

Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (in press). Routine language: Speech directed to infants during home activities. *Child Development*.

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science*, *20*, e12456.

Thelen, E. (1986). Treadmill-elicited stepping in seven-month-old infants. *Child Development*, *57*, 1498–1506.

Trueswell, J. C., Lin, Y., Armstrong, B., III, Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent-child interactions. *Cognition*, *148*, 117–135.

Vondrick, C., Patterson, D., & Ramanan, D. (2013). Efficiently scaling up crowdsourced video annotation – A set of best practices for high quality, economical video labeling. *International Journal of Computer Vision*, *101*, 184–204.

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using head camera to study visual experience. *Infancy*, *13*, 229–248.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*, 244–262.

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS ONE*, *8*, e79659.

Yu, C., & Smith, L. B. (2017). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science*, *41*, 5–31.

Yu, C., Smith, L. B., Shen, H., Pereira, A. F., & Smith, T. G. (2009). Active information selection: Visual attention through the hands. *IEEE Transactions on Autonomous Mental Development*, *2*, 141–151.

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, *22*, e12735.

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, *16*, 959–966.

Zhang, Y., & Yu, C. (2016). Examining referential uncertainty in naturalistic contexts from the child's view: Evidence from an eye-tracking study with infants. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 2027–2032). Austin, TX: Cognitive Science Society.

APPENDIX A
SUBJECT-LEVEL DATA AND ANALYSES

Table A1
Visual Referential Clarity of Naming Utterances

Sub	*Named object (TRGT)*	*Non-named objects (DIST)*	*Object when not named (BASE)*	*TRGT versus DIST*		*TRGT versus BASE*	
				t	*p*	*t*	*p*
Overall visual referential clarity: Mean image size of objects across contexts							
1	4.22 (6.02)	1.48 (1.24)	1.31 (.52)	4.07	<.001	4.17	<.001
2	4.20 (3.97)	1.07 (.87)	1.37 (.74)	7.99	<.001	7.42	<.001
3	7.26 (6.65)	1.90 (1.35)	2.94 (1.35)	8.15	<.001	7.15	<.001
4	3.70 (3.13)	1.09 (.81)	2.62 (1.21)	4.96	<.001	2.07	.045
5	5.72 (5.71)	1.12 (.82)	1.62 (.78)	3.43	.003	3.36	.003
Mean image size of in-view objects across contexts							
1	6.02 (6.79)	2.57 (1.03)	3.03 (.96)	3.51	.001	3.05	.003
2	5.47 (4.29)	2.05 (1.17)	2.43 (1.09)	6.96	<.001	4.44	<.001
3	9.38 (6.82)	3.03 (2.14)	5.17 (2.92)	8.15	<.001	7.15	<.001
4	4.59 (2.78)	3.61 (2.04)	4.21 (1.64)	1.54	.130	.77	.447
5	6.38 (5.55)	2.24 (.96)	3.40 (1.37)	3.06	.007	2.40	.028
Mean likelihood of objects being in view across contexts							
1	.77 (.41)	.55 (.38)	.45 (.15)	3.94	<.001	7.60	<.001
2	.75 (.37)	.48 (.30)	.57 (.17)	6.67	<.001	6.67	<.001
3	.76 (.37)	.59 (.28)	.60 (.16)	4.34	<.001	4.25	<.001
4	.77 (.36)	.29 (.14)	.59 (.15)	7.51	<.001	2.90	.006
5	.82 (.32)	.46 (.22)	.48 (.08)	4.33	<.001	5.07	<.001

Note. Standard deviations in parentheses.

Table A2
Visual Dominance of the Focal Object Across Naming and Non-naming Frames

Sub	*Naming moments*	*Non-naming moments*	*Naming versus Non-naming*	
			t	*p*
1	5.47 (7.43)	3.94 (4.52)	3.29	.001
2	4.90 (3.94)	4.14 (3.95)	2.61	.009
3	8.73 (7.71)	8.54 (7.90)	.32	.750
4	5.75 (3.86)	6.75 (5.52)	−1.76	.079
5	6.50 (5.45)	7.41 (5.77)	−.99	.323

Note. Data represent the mean difference scores between focal and non-focal objects. Standard deviations in parentheses.

Table A3
Visual Dominance of the Focal Object Across Different Holding Contexts

				T versus N		*P versus N*		*T versus P*	
Sub	*Toddler (T)*	*Parent (P)*	*Neither (N)*	*t*	*p*	*t*	*p*	*t*	*p*
1	6.27 (6.15)	3.57 (1.73)	1.98 (2.15)	14.26	<.001	2.77	.006	3.26	.001
2	5.32 (4.30)	3.69 (3.44)	2.50 (3.94)	7.28	<.001	3.40	.001	4.96	<.001
3	10.15 (9.14)	6.44 (4.94)	2.17 (1.83)	9.47	<.001	8.77	<.001	3.94	<.001
4	7.24 (4.81)	3.82 (2.93)	2.56 (2.75)	4.41	<.001	1.26	.217	2.55	.011
5	7.41 (6.15)	9.19 (5.81)	1.10 (2.57)	4.28	<.001	5.72	<.001	−1.82	.071

Note. Data represent the mean difference scores between focal and non-focal objects. Standard deviations in parentheses.

Table A4
Mean Object Image Sizes Time-locked to Toddler and Parent Holding Events

			Pre versus Onset				*Offset versus Post*	
Sub	*Pre*	*Onset*	*t*	*p*	*Offset*	*Post*	*t*	*p*
Toddler holding events								
1	2.29 (1.78)	5.28 (4.65)	6.11	<.001	4.48 (3.43)	1.84 (1.98)	7.40	<.001
2	2.12 (2.70)	3.74 (3.16)	3.78	<.001	3.31 (4.00)	1.20 (1.99)	4.27	<.001
3	2.15 (2.18)	5.93 (5.11)	4.98	<.001	3.96 (3.97)	1.89 (2.69)	3.78	<.001
4	1.46 (2.25)	4.27 (3.48)	5.34	<.001	5.31 (4.68)	1.14 (2.18)	6.88	<.001
5	.87 (1.35)	4.77 (3.48)	4.74	<.001	4.12 (3.81)	.82 (1.93)	2.71	.017
Parent holding events								
1	1.73 (1.78)	3.23 (2.32)	3.90	<.001	3.44 (2.32)	2.11 (2.15)	3.17	.003
2	1.64 (2.95)	2.48 (2.77)	3.85	<.001	2.75 (2.83)	1.67 (2.57)	2.89	.005
3	1.82 (2.57)	4.62 (4.63)	4.54	<.001	4.07 (4.85)	2.61 (3.95)	2.11	.004
4	1.41 (2.12)	3.16 (2.44)	4.61	<.001	3.30 (2.67)	2.16 (2.75)	2.33	.023
5	1.62 (2.91)	5.69 (4.83)	6.10	<.001	5.48 (4.39)	2.40 (2.85)	4.00	.002

Note. Pre = mean object image size during the 3 sec before holding event; Onset = mean object image size during the 3 sec after onset of holding event; Offset = mean object image size during the 3 sec before offset of holding event; Post = mean object image size during the 3 sec after holding event. Standard deviations in parentheses.

APPENDIX B
Toy Object List

Table B1
The Physical Dimensions of All Toy Objects from Which Object Sets for Each Participant were Selected

Object	*L*	*W*	*H*	*Object*	*L*	*W*	*H*
Alligator	12.4	7.5	4.6	Giraffe	18.2	6.2	13.7
Apple	7.9	7.8	6.5	Green Beans	6.0	6.4	2.4
Baby	12.4	6.2	3.3	Hammer	6.0	6.4	2.4
Ball	8.6	8.6	8.6	Jeans	18.0	6.5	2.4
Banana	9.5	2.9	2.4	Keys	8.9	5.1	3.3
Bird	20.6	7.3	6.5	Moose	10.0	6.4	10.3
Boat	13.2	7.8	6.0	Motorcycle	11.8	5.4	4.9
Car	14.8	8.3	5.6	Penguin	14.6	5.2	13.5
Cat	17.8	11.9	5.6	Rooster	8.6	2.4	5.9
Comb	14.1	4.1	.3	Scissors	11.3	6.5	0.6
Cup	19.1	7.5	20.3	Shoe	11.1	5.6	6.5
Dog	15.1	15.2	7.3	Tractor	9.4	5.4	6.4
Duck	9.2	6.7	7.0	Truck	10.6	4.9	4.4
Fish	14.6	8.1	6.0	Zebra	8.9	2.8	8.1

Note. L = length, W = width, H = height; measurements are in centimeters.