Published in final edited form as:
Psychon Bull Rev. 2014 February ; 21(1): 178–185. doi:10.3758/s13423-013-0466-4.

A Bottom-up View of Toddler Word Learning

Alfredo F. Pereira[1], **Linda B. Smith**[2], and **Chen Yu**[2]
[1]Universidade do Minho

[2]Indiana University Bloomington

Abstract

A head-camera was used to examine the visual correlates of object name learning by toddlers, as they played with novel objects, and as the parent spontaneously named those objects. The toddlers' learning of the object names was tested after play and the visual properties of the head-camera images during naming events associated with learned and unlearned object names were analyzed. Naming events associated with learning had a clear visual signature, one in which the visual information itself was clean and visual competition among objects was minimized. Moreover, for learned object names the visual advantage of the named target over competitors was sustained, both before and after the heard name. The findings are discussed in terms of the visual and cognitive processes that may depend on clean sensory input for learning and also on the sensory-motor, cognitive and social processes that may create these optimal visual moments for learning.

Children learn their first object names by linking a heard word to a seen thing. Contemporary theories all assume that the learning environment is noisy, with scenes containing several potential referents for a heard name. Different theories posit different mechanisms through which young learners reduce this uncertainty, including social cues to speaker intent (Baldwin, 1995; Tomasello & Aktar, 1995), innate linking functions between linguistic categories and meanings (Booth & Waxman, 2009; Lidz, Waxman & Freedman, 2003), and statistical mechanisms that aggregate word-object co-occurrences across multiple naming events (Xu & Tenebaum, 2007; Smith & Yu, 2008; Frank, Goodman & Tenenbaum, 2009). Here we present new evidence on the nature of the learning environment at the sensory level, in terms of the moment-to-moment visual information available to the learner about potential referents for a heard name. The findings raise questions about the starting assumption of rampant ambiguity in the early object name-learning environment and suggest new hypotheses about how visual clutter and competition may limit early word learning.

Our interest in and approach to studying the dynamic visual correlates of object-name learning stem from four considerations. First, the everyday visual world not only offers potentially many referents but it is also dynamically complex; objects in the scene move and change in relation to each other, and in relation to the sensors as the perceiver also acts and moves. Second, a large literature studying toddler attention shows how this everyday context of a moving body and moving objects is attentionally challenging (e.g., Kanass, Oakes, & Shady, 2006). Indeed, sustained attention during play with multiple objects is used to assess individual differences in attentional functioning in typically and atypically developing toddlers (e.g. Lawson & Ruff, 2004). Third, a growing literature on atypical development indicates the co-morbidity of sensory-motor, attention, and language delays (e.g., Iverson, 2010). These links are not well understood mechanistically. However, the significant changes in motor behavior that characterize the second year of life (e.g., Adolph & Berger,

Corresponding Author: Alfredo F. Pereira, alfredo.pereira@psi.uminho.pt.

2006) bring with them bodily instabilities and, as a result, large head and trunk movements (Berthenthal & von Hofsten, 1998). These movements directly affect the visual input and potentially destabilize attention, and may create special challenges to object name learning. Finally, several recent studies have used head cameras to capture the moment-to-moment visual dynamics as toddlers engage in various activities (Yoshida & Smith, 2008; Aslin, 2009; Cicchino, Aslin & Rakison, 2010). These studies show that toddlers' head-centered views during active play are not at all like adult views in that they are highly dynamic, with individual objects coming into and going out of view on times scales of seconds and fractions of seconds (Smith, Yu & Pereira, 2011). All four considerations suggest that value of studying the ambiguity of naming movements from the perspective of the dynamic properties of visual experience.

One particular result from the prior head-camera studies motivates our specific experimental question. Amidst the highly dynamic views that were found to characterize active toddlers' visual experiences were occasional less dynamic periods when, despite many objects being in near physical proximity to the child, there was just one object stably dominating the head camera image, being much larger in visual size because it was closer and unoccluded (Yoshida & Smith, 2008; Yu et al., 2009; Smith, Yu & Pereira, 2011). We ask: Are these periods of stable, clean, nearly one-object views optimal sensory moments for the early learning of object names? To answer this question, toddlers' first-person views were recorded by a head camera, as they played with several novel objects with a parent, and as the parent spontaneously named those objects. The toddlers' learning of the object names was tested after play and the visual properties of the head camera images during naming events associated with learned and unlearned object names were analyzed. Based on the prior head-camera studies, the main dependent measures were the temporal profile of the named object's image size before, during and after a naming and the same temporal profile for the un-named competitors. We also measured the centering of the objects in the image providing a dynamic profile of the spatial direction of attention with respect to the named target and un-named competitors around moments of naming.

Method

Participants

Twelve toddlers (7 male, 16 to 25 months, M = 20 months) were recruited to participate. Three additional children did not contribute data to the final analyses, either because of failure to tolerate the head-camera, or because of calibration difficulties.

Stimuli

Six novel objects (on average about 9.5 × 6.5 × 5 cm) were custom made from hardened clay to have unique shapes and textures. Each object was randomly paired with one name *zeebee, tema, dodi, habble, wawa, and mapoo* and the objects were organized in two sets of three. Within each set, one object was painted blue, one red, and one green.

Head camera

The mini head camera (KT&C model VSN500NH, f2.45, 768×494 pixels CCD resolution) was embedded in a custom headband and recorded a broad 97° visual field in the horizontal – approximately half of the visual field of infants (Mayer & Fulton, 1993) – and 87° in the vertical. A prior calibration study (Yoshida & Smith, 2008) independently measured eye-gaze direction and head-direction during toy play and found that non-correspondence between head and eye were generally infrequent (less than 17% of frames) and brief (less than 500 ms, see also Smith et al. 2011; Yu et al., 2009). To place the head-camera on the infant, one experimenter distracted the child, while the second placed the head camera on

the head. The child was then directed to push a button on a pop-up toy, and the camera was adjusted such that the button at the moment it was pushed was centered in the head-camera image. Additional third person cameras were used to record the play session, and to record the experimenter and the child during testing.

Experimental room

Parents and infants sat across from each other at a small table (61cm × 91cm × 64cm) that was illuminated from above. The average distance of the infant's eye to the center of the table was 43.2 cm. Parent and toddler wore white clothing and the walls, table, and floor were also white so that shadows were minimized.

Procedure

Prior to the play task, the parent was instructed as to the names of the six novel objects and asked to use these names during play. To remind the parents of the object labels, the labels and object pictures were attached to the boxes from which parents retrieved the two toy sets. Parents were *not* told that their task was to teach the names nor that infants would be later tested. They were only told to encourage their child's interaction with the objects in as natural a manner as possible. Parents were alone in the room with their child during the play period. There were 4 toy play trials, two with each set of three objects, lasting 1.5 min each. The start and stop of each play period was cued by an auditory signal. The parent's voice was recorded using a noise reduction microphone.

After the play trials, an experimenter entered the room and tested the toddlers' knowledge of the object names. On each test trial, three objects were placed on a tray, 44cm wide, such that one object was to the extreme right, one to the extreme left, and one at midline. The experimenter held the tray away from the infant, looked continually into the infant's eyes, never at the objects (as confirmed by video recording), and said "Show me the ____! Get the ____!" and then moved the tray forward for the infant to select an object. Each of the six object names was tested twice (with all three objects tested once before any object was tested a second time). The distracters on each trial were randomly selected from the other play objects with the following constraints: all objects served as distracters equally often, each trial was composed of one red, one blue, and one green object, and the distracters used for any target differed on the two testing trials. The location of the correct object varied (via a Latin Square) across trials for each infant.

Coding

A naming event was defined as any whole parent utterance (e.g., "*What are you doing with that habble*?") that contained an object name. A silence duration of more than 0.4 sec was used to mark the temporal boundaries of utterances, and human coders then identified utterances that included the object names. Agreement for two coders for a randomly selected set of utterances exceeded 90% and all disagreements were resolved by the two coders re-listening to the audio recordings.

The head-camera video was sampled at 10Hz and head-camera images were analyzed frame-by-frame for the 10 sec prior, during, and 10 sec after each naming event, yielding approximately 640 data points (frames) for each naming event. Measures of the visual properties were taken for each of the three play objects in the defined window using a custom image-analysis software (see Yu et al., 2009): (1) the image size of each of the three objects – measured by proportion of object pixels in the image; and (2) the centering of each object in the image – measured by computing the average distance of all object pixels to the image center and expressing that average distance as a proportion of the head-camera image's half diagonal, i.e. a fully centered object pixel corresponds to zero centering and an

head-camera image corner pixel has a centering value of one. All objects were the same physical size; thus image size and overlap with center varies with infant and object movements. For the statistical analyses, the 10 Hz time series were averaged within the utterance containing the naming event and within one-second windows for each of the 10 sec prior to and after the naming utterance.

The toddler's performance at test was scored by a naïve human coder who did not know the correct choice and who made an all-or-none decision as to the selected object on each object-name test trial. A second scorer scored a randomly selected 25% of the test trials and the level of agreement exceeded 94%. An object name was defined as "learned" if the toddler correctly selected it on two of the two testing trials; otherwise the object name was considered "not learned".

Results

Parents produced each of the 6 object names on average 9.7 times (SD = 4.4). At test infant choices indicated that on average 1.58 names were learned (range across the 12 toddlers, 1 – 6); overall this level of success exceeds that expected by random choice (0.67 correct names), $t(11) = 3.19$, $p < .01$ (two-tailed). However, the key issue is not whether infants could learn some object names, but rather the visual properties of the individual naming events that supported this learning. Accordingly, naming events were partitioned into those associated with learned versus unlearned object names. This is a noisy partition as not all naming events associated with learned object names may have contributed to learning. The mean number of naming utterances per parent associated with each learned object name was 11.0 (SD = 6.1) and was more than the number of naming utterances associated with unlearned object names, 8.0 (SD = 4.2), $t(70) = -2.27$, $p < .05$. The average duration of utterances containing a name was 1.25 sec, SD = 0.61 sec, and was slightly less than the average duration of utterances associated with unlearned object names of 1.38 sec (SD = 0.73 sec), $t(637) = -2.13$, $p < .05$. Both of these factors could contribute to learning; the key question for this study, however, concerns the dynamics of the visual properties of the naming events.

Do the visual properties of naming events associated with learned names differ from those associated with non-learned names? To answer this question, the two dependent measures, the image size of the objects and their centering in the image, were analyzed for 10 sec before and 10 sec after each naming event with the critical questions concerning the temporal profiles of these properties for the named target and for the other objects, the potential competitors. The analyses examine the properties of the head camera image for a 20 sec window around a naming utterance. More than one naming utterance (for the same or different objects) could potentially be contained in the same 20 sec window around a single naming utterance, yielding overlapping 20 sec windows for two different naming events. These were included in the analyses because they were relatively infrequent and did not differ for learned and unlearned object names. The proportion of naming events that overlapped each other within the 20 sec window was 8.6% for learned object names and 10.9% for unlearned object names.

The analyses were conducted on a total of 639 naming events (209 and 430, for learned and unlearned object names respectively) and used the methodology of GCA (Growth Curve Analysis). Separate GCAs were conducted for image size and object centering. GCA is a type of hierarchical linear modeling concerned with capturing time effects under assumptions of a continuous stochastic process, and is structured hierarchically at least two levels (see Mirman, Dixon, & Magnuson, 2008). In Level-1, the growth curve for each dependent variable is modeled by a linear regression using time as a predictor. The

regression model can include a zero-order (intercept), a first-order (slope), and higher-order polynomial *Time* terms. Because the polynomial terms are naturally collinear they were transformed into orthogonal polynomials so that the contribution of each polynomial term could be assessed independently of the others. The Level-2 model considers the Level-1 model as potentially explainable by a linear regression of population averages, fixed effects (typically the effects of interest), and random effects, and thus serves the role in the analyses of the more typical analysis of variance. To build the Level-1 and Level-2 model we followed the methodology of Baayen, Davidson, & Bates (2008). A model comparison approach based on a likelihood ratio test was used and models checked for possible over-fitting by examining the residuals of any random effects, and the correlations between fixed effects. Visual inspection of temporal profile for object size and centering measures revealed a clear U-curve, inverted for object size and U-shaped for centering, with the maximum (object size) or minimum (centering) point at the naming utterance. Consequentially we explored, for Level-1, models that included an intercept, a linear and a quadratic *Time* term. In order to account for individual and stimuli differences we considered a *Participant* random effect and a separate *Object Label* random effect (i.e. these were crossed random effects); we did not include interactions between *Time* and *Participant* or *Object Label*. The Level-2 model was constructed in two-steps: first, we used a series of model comparisons to determine different random effect structures (intercept, linear and quadratic terms) for *Participant* and *Object Label* effects; second, we added a full two-way interaction between the fixed effects of interest, (1) *Named Object* (target/competitor) and (2) *Learning* (learned/unlearned). Model parameters were estimated using the lme4 package (Bates, 2005, 2012; available in R, R development core team, 2008). Fixed effects were contrast-coded, and *p*-values for model parameter estimates computed using a Markov chain Monte Carlo (MCMC) simulation method (see Baayen, Davidson, & Bates, 2008).

In preview, the main conclusions that arise from the analyses, evident in Figure 1, are these: First, whether or not the name is learned, the visual properties of the named targets differed from unnamed ones, specifically, for both learned and unlearned object names the named target had an image-size advantage over competitors and was more centered in the visual field than the un-named competitor objects. Second, named targets that were learned differed from named targets that were not learned in the magnitude of the difference in these visual properties between the named target and the other, competitor, objects. Specifically, naming events for learned names showed a larger difference between named target and competitors, with the implication of less visual competition, that do the named targets that were not learned.

Object size

The temporal profiles for image size for target and (the average of) the competitors are shown in Figure 1C and D and the main results of the GCA are given in Table 1. The GCA yielded a best-fit model with a quadratic (B = −0.58, $p < .001$) *Time* term, indicating a rise and then fall of image size before and after a naming event and thus a clear dynamic link between image size and naming events. The GCA also yielded an average image size advantage for the named target versus the un-named competitors (B = 0.51, $p < .01$) but no main effect for learned versus unlearned words (B = 0.08, $p < .093$). Critically, the analysis yielded a reliable *Learning* X *Named Object* interaction (B = 0.92, $p < .001$), as the named target's image size advantage was greater for learned than unlearned object names. The maximum correlation between fixed effects was moderate, $r = 0.41$. The analysis also revealed a random intercept per *Participant* and a random intercept per *Object Label*. These indicate individual differences and stimulus differences (reflecting stimulus specific differences in how the infants held and interacted with the objects). The main conclusion, as apparent in Figure 1C and D, is: naming events associated with *learned object names*, more

than associated with *unlearned object names*, are characterized by temporal profile in which the image size for the named target is larger than that of the un-named competitors.

To determine when in the time series, the named target diverged in image size from the mean of the competitors, we determined the first and last significant difference in a series of ordered pairwise *t*-tests (Allopenna, Magnuson & Tannenhaus, 1998). For naming events associated with learning, the target advantage was stable and enduring: image size was reliably different for the target versus competitors at 6 seconds prior to the naming event and persisted until 5 seconds after the event. For naming events associated with unlearned object names, there was also a target advantage but is was much briefer; image size was reliably different for the target versus competitors only at 3 seconds prior to the naming event and persisted until 1 second after the event. In sum, for naming events associated with learning, the named object was more visually dominant than the competitors – larger in the field because it was closer and un-occluded – and this dominance was sustained over time.

Centering

The temporal profiles for centering for target and (the average of) the competitors are shown in Figure 1E and F and the main results of the GCA are also given in Table 1. The GCA for this measure yielded a best-fit model with a linear (B = −1.7, $p < .05$) and a quadratic (B = 4.1, $p < .001$) *Time* term. Centering, like image size, rises up to the naming event and then falls after then naming event. There was a reliable effect of *Named Object*, with an advantage in centering for the named target over competitors (B = −1.6, $p < .001$) and also an effect of *Learning* (B = −1.8, $p < .001$). Similar to the object size measure, a significant 2-way interaction of *Learning* X *Named Object* indicates that the target advantage in centering over the un-named competitors is larger for naming events associated with learning (B = −2.7, $p < .001$). The maximum correlation between fixed effects was moderate, $r = 0.44$. The analysis also yielded a random intercept per *Participant and* a random intercept per *Object Label*, again showing individual differences and stimulus differences in centering. Overall this pattern indicates that parents sensibly named objects when the child's spatial attention was directed to the target. Finally, by the method of first and last reliable pairwise differences, the overlap with the image center was reliably different for the target versus competitors at 4 seconds prior to the naming event and persisted until 1 seconds after the event for the naming events associated with learned object names, and was reliably different for the target versus competitors at 3 seconds prior to the naming event and persisted until 1 second after the event for naming events associated with unlearned object names. The main results of the centering analyses are these: (1) the named target shows a clear temporal profile in which the named target – but not the competitors – is increasingly more centered in the child's view prior to the naming event and that this centering declines after naming and (2) naming events associated with learned show a higher centering advantage of the named target over the competitors than did the unlearned named targets.

The joint consideration of both the image size and centering analyses yields the following conclusion: Both centering and image size are dynamically related to the naming of an object by a parent, and indicate that parents named objects when the target was being attended to by the child. However, learning also depended on the sustained visual dominance, as measured by image size and centering, of the named target over competitors.

It is likely that the two visual measures are not orthogonal but are co-dependent in a context of free-flow interaction. For example, child or parent holding of an object so that the child is actively examining it during naming could bring the object closer to the child's view with the result of both a larger and more centered image of the object in the head camera. To determine the degree to which these two measures might be dynamically linked, and thus

redundant measures of the very same visual event, we repeated the GCA analysis by partialing out the effect of the second measure. Specifically we estimated the parameters of two models: the best-fit model structure for image size and for centering, but with the residuals of image size predicted by centering (using a linear regression), and the residuals of centering predicted by image size as the dependent variable.

In summary, this analysis revealed that though moderately correlated, $r = 0.48$, $p < .001$, image size and centering are not entirely redundant. The parameters that remained significant were the *Named Object* fixed effect, and the *Learning* X *Named Object* 2-way interaction, when predicting residuals of image size ($p < .001$), and the quadratic *Time* term, *Learning* and *Named Object* fixed effects, and the *Learning* X *Named Object* 2-way interaction when predicting residuals of centering ($p < .05$). Comparing these findings with the main results in Table 1, this analysis yielded the same general conclusion: object size and centering of the target relative to competitors distinguished naming events associated with learned object names from those associated with unlearned object names, the visual dominance effect of named target vs. competitors, and the higher visual. The sole qualitative difference was in the *Time* terms, perhaps reflecting the similarities in the temporal pattern of both measures (a U-shaped curved that peaks at the naming event). This overall pattern suggests that object size and centering, though likely interdependent visually and in the sensory-motor aspects of the interaction that give rise to them, are also somewhat separable in their effects on learning and also perhaps in the specific behaviors by parents and infants that give rise to them.

Finally, to ensure that these conclusions did not depend on averaging the image sizes and centering of the competitors, a third set of analyses used the maximal value of the two competitors rather than the mean; these analyses revealed the same basic findings.

Discussion

The results reveal the properties of visually optimal moments for toddlers to learn an object name: when the named object is visually larger and more centered than competitors, and when that visual advantage is sustained for several seconds before and also after the naming event. The results are correlational and as such cannot specify the factors that created the observed visual signature for learned object names nor the mechanisms through which limited visual competition and sustained attention benefit learning. However, the findings suggest that the sensory properties of naming moments matter. They also provide new insights into the assumptions about ambiguity in the input and also raise new hypotheses – at the visual level – about the specific challenges posed by scenes with multiple objects.

Contemporary theories of early object name learning begin with the problem of referential ambiguity and offer cognitive solutions to that problem: the inference of a speaker's intended referent from social cues (e.g., Baldwin, 1995), the use of linguistic cues and innate biases (e.g., Lidz et al., 2003), and powerful statistical learning mechanisms (e.g., Xu & Tenenbaum, 2007). However, the present results tell us that for young learners there is sometimes *little* ambiguity, and that these moments of minimal visual ambiguity are strongly associated with object name learning. Not all naming moments had this property; many naming events associated with unlearned names were associated with multiple and nearly equal competitors for that name. Thus, the present results affirm the ambiguity often assumed and show that it also characterizes the visual level and the first person view; and the results show that such ambiguity does make learning more difficulty. But they also show there are very clean sensory moments when no additional cognitive processes would seem to be needed to determine the relevant object; no cognitive processes are needed because there is a sustained view in which just one object is much more salient in image size and centering

than possible competitors. One might conclude from these findings that there is no need to propose higher cognitive learning mechanisms, as young word learners might only learn words when there is minimal ambiguity at the visual level. Alternatively, these visually optimal moments may play a bootstrapping role, helping the child acquire or tune more cognitive and inferential processes that can succeed even given noisy input.

The dynamic visual properties of naming events associated with learning versus not learning the object name also suggests that there are *visual* limits on object name learning. This is a perspective that has not been considered in previous research but that is critical to understanding the mechanisms that underlie early object name learning, and the properties of the learning environment that matter. Previous studies of adult visual processing show that multiple objects that are visually close to each other perturb both visual selection and representation in adults (e.g. Henderson, Chaneaux & Smith, 2009). Recent studies suggest that the negative effects of clutter and crowding may be even more pervasive in toddlers (Oakes, Hurley, Ross-Sheehy & Luck, 2010). Movement and change in the visual field can mandatorily capture attention in adults (see Knudsen, 2007) and also in toddlers (Columbo, 2001). Clearly, we need to understand these visual limits on early object name learning in a greater detail. Indeed, the key factors in parent-child interactions with respect to early object name learning may be in limiting visual clutter and in sustaining selective attention on one object. Infant behavior itself may matter as previous head camera studies suggest that views in which one-object dominate are often linked to the toddlers' holding of the object (Yu et al., 2009; Smith et al., 2011). A large literature also suggests an important role for parent behavior, both as a top down cue to attention (e.g. Tomasello & Aktar, 1995) and also in terms of behaviors – holding, moving, and gesturing – that may directly structure the visual input. The present findings also suggest clear limits on what parents can do: Parents named objects when their infants heads were spatially directed to object (and the object was close to the child and centered in the view), and sometimes infants learned and sometimes they did not. Parents sometimes named the object when one object was visually dominant over competitors (and their infants learned) but they also sometimes named the object when the target and competitors were more equal in visual size (and their infants did not learn). This suggests that child's view and its properties at the sensory level are not completely transparent to parents. Detailing the role of parent behavior and child behavior in structuring the bottom up information and parent sensitivity to that information are key issues for future research.

One potentially important finding with respect to the mechanisms underlying early word learning is the temporal duration of the visual advantage of the target over competitors for naming events associated with learning: beginning 6 secs prior to the naming event and lasting 5 sec after. This long duration could be indicative of the kind of factors – child activity and interest, parent activity in structuring the learning moment – that create optimal visual moments for learning object names and need not be essential to the mechanisms of learning. However, the increased stickiness of attention over time has been hypothesized to be important to sustained attention in toddlers (e.g. Richards, 2000). Alternatively, the internal processes that bind a name to an object may themselves take time, and might, for example, require the formation of a stable visual representation of the object (Fennell, 2011 and Ramscar et al., 2010) prior to the naming event and/or maintenance of that visual representation (without replacement by another attended object) for some time after the heard name. These are hypotheses that need to be experimentally evaluated. In summary, the duration of sustained visual dominance of the target over the competitor observed in the present results may provide important clues as to how these optimal visual moments were created and also the mechanisms through which they benefit object-name learning.

In conclusion, some early naming events are not ambiguous, not from the learner's view as there is but one dominant object in view. These may be optimal visual moments for mapping a name to an object and play a particular critical role for very young word learners. The differences in the visual properties of naming events associated with learned and unlearned object names also suggests potential visual limits on learning –in terms of clutter and in terms of sustained selective attention that endures over several seconds, limits that merit detailed experimental study.

Acknowledgments

We thank Charlotte Wozniak, Amanda Favata, Amara Stuehling, and Andrew Filipowicz for collection of the data and Thomas Smith for developing data management and preprocessing software. This research was supported by National Science Foundation Grant 0924248, AFOSR FA9550-09-1-0665, National Institutes of Child Health and Development grant R21HD068475, and Portuguese Ministry of Education and Science Postdoctoral fellowship SFRH/BPD/70122/2010 awarded to Alfredo F. Pereira.

References

Adolph, KE.; Berger, SE. Motor development. In: Damon, W.; Lerner, R., editors. Handbook of child psychology. 6th ed.. Vol. Vol 2. New York: Wiley; 2006. p. 161-213.

Allopenna PD, Magnuson JS, Tanenhaus MK. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of Memory and Language. 1998; 38(4):419–439.

Aslin RN. How infants view natural scenes gathered from a head-mounted camera. Optometry and Vision Science: American Academy of Optometry. 2009; 86(6):561.

Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. Journal of memory and language. 2008; 59(4):390–412.

Baldwin, DA. Understanding the link between joint attention and language. In: Moore, C.; Dunham, JP., editors. Joint attention: Its origins and role in development. Erlbaum; 1995. p. 131-158.

Bates D. Fitting linear mixed models in R. R news. 2005; 5(1):27–30.

Bates, D. Linear mixed model implementation in lme4. Ms., University of Wisconsin; 2012.

Bertenthal B, Von Hofsten C. Eye, head and trunk control: The foundation for manual development. Neuroscience & Biobehavioral Reviews. 1998; 22(4):515–520. [PubMed: 9595563]

Booth AE, Waxman SR. A horse of a different color: Specifying with precision infants' mapping of novel nouns and adjectives. Child Development. 2009; 80:15–22. [PubMed: 19236389]

Cicchino JB, Aslin RN, Rakison DH. Correspondences between what infants see and know about causal and self-propelled motion. Cognition. 2010; 118:171–192. [PubMed: 21122832]

Colombo J. The development of visual attention in infancy. Annual Review of Psychology. 2001; 52(1):337–367.

Fennell, CT. Infancy. 2011. Object familiarity enhances infants' use of phonetic detail in novel words.

Frank MC, Goodman ND, Tenenbaum JB. Using speakers' referential intentions to model early cross-situational word learning. Psychological Science. 2009; 20(5):578–585. [PubMed: 19389131]

Henderson JM, Chanceaux M, Smith TJ. The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. Journal of Vision. 2009; 9(1)

Huttenlocher J, Haight W, Bryk A, Seltzer M, Lyons T. Early vocabulary growth: Relation to language input and gender. Developmental Psychology. 1991; 27(2):236–248.

Iverson JM. Developing language in a developing body: The relationship between motor development and language development. Journal of Child Language. 2010; 37(02):229–261. [PubMed: 20096145]

Kannass KN, Oakes LM, Shaddy DJ. A longitudinal investigation of the development of attention and distractibility. Journal of Cognition and Development. 2006; 7(3):381–409.

Knudsen EI. Fundamental components of attention. Annu.Rev.Neurosci. 2007; 30:57–78. [PubMed: 17417935]

Lawson KR, Ruff HA. Early focused attention predicts outcome for children born prematurely. Journal of Developmental & Behavioral Pediatrics. 2004; 25(6):399–406. [PubMed: 15613988]

Lidz J, Waxman SR, Freedman J. What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18 months. Cognition. 2003; 89:295–303.

Mayer, DL.; Fulton, AN. Development of the human visual field. In: Simons, K., editor. Early Visual Development, Normal and Abnormal. New York, NY: Oxford University Press; 1993. p. 117-129.

Mirman D, Dixon JA, Magnuson JS. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. Journal of memory and language. 2008; 59(4):475–494. [PubMed: 19060958]

Mirman D, Dixon JA, Magnuson JS. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. Journal of Memory and Language. 2008; 59:475–494. [PubMed: 19060958]

Oakes LM, Hurley KB, Ross-Sheehy S, Luck SJ. Developmental changes in infants' visual short-term memory for location. Cognition. 2010; 118:293–305. [PubMed: 21168832]

R development core team. R: A language and environment for statistical computing. R Foundation Statistical Computing. 2008

Ramscar M, Yarlett D, Dye M, Denny K, Thorpe K. The effects of feature label order and their implications for symbolic learning. Cognitive Science. 2010; 34(6):909–957. [PubMed: 21564239]

Richards JE, Cronise K. Extended visual fixation in the early preschool years: Look duration, heart rate changes, and attentional inertia. Child Development. 2000; 71(3):602–620. [PubMed: 10953928]

Ross-Sheehy S, Oakes LM, Luck SJ. Exogenous attention influences visual short-term memory in infants. Developmental Science. 2010; 14:490–501. [PubMed: 21477189]

Ruff HA, Saltarelli LM. Exploratory play with objects: Basic cognitive processes and individual differences. New Directions for Child and Adolescent Development. 1993; 1993(59):5–16.

Smith LB, Yu C, Pereira AF. Not your mother's view: The dynamics of toddler visual experience. Developmental Science. 2011; 14(1):9–17. [PubMed: 21159083]

Smith L, Yu C. Infants rapidly learn word-referent mappings via cross-situational statistics. Cognition. 2008; 106:1558–1568. [PubMed: 17692305]

Tomasello M, Akhtar N. Two-year-olds use pragmatic cues to differentiate reference to objects and actions. Cognitive Development. 1995; 10(2):201–224.

Xu F, Tenenbaum JB. Word learning as Bayesian inference. Psychological Review. 2007; 114:245–272. [PubMed: 17500627]

Yoshida H, Smith LB. What's in view for toddlers? using a head camera to study visual experience. Infancy. 2008; 13(3):229–248. [PubMed: 20585411]

Yu C, Smith LB, Shen H, Pereira AF, Smith T. Active information selection: Visual attention through the hands. Autonomous Mental Development, IEEE Transactions on. 2009; 1(2):141–151.

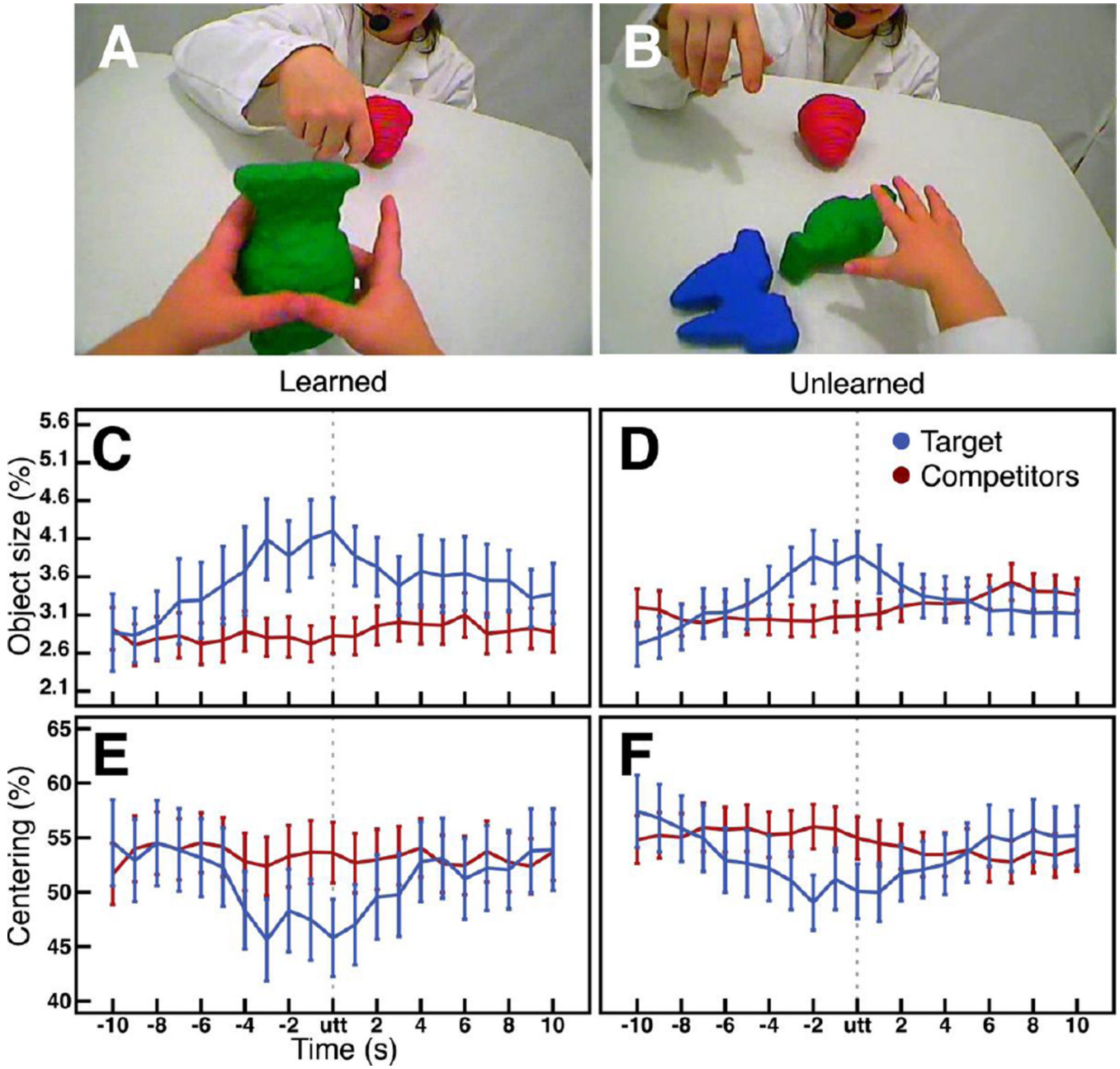

Figure 1.

(A) Example of a visual scene, while the parent labeled a target referent (green), for a referent the child learned the object name, and (B) for a target referent (blue) the child did not learn the object name. Temporal profiles for object size – measured as proportion of the image – for target and (the average of) the competitors are shown in (C) for naming events associates with learned words and (D) for naming events associated with unlearned words. Temporal profiles for centering – measured as average object pixel distance to center expressed as proportion of half-diagonal – for target and (the average of) the competitors are shown in (E) for naming events associates with learned words and (F) for naming events associated with unlearned words.

Table 1

Results of the Growth Curve Analysis for object image size (left section) and centering (right section). B-parameter estimates are presented with its respective 95% Highest Posterior Density interval and p-value calculated using an MCMC method; when meaningful, the category the estimate refers to is in parentheses.

	Object Image Size (% of image)				**Centering (average distance to center) (% of half-diagonal)**			
		95% HPD interval				**95% HPD interval**		
	Estimate	***Lower***	***Upper***	***p* <**	**Estimate**	***Lower***	***Upper***	***p* <**
Intercept	3.20	2.72	3.69	.001	52.9	46.7	58.8	.001
Time	Not included in best-fit model				−1.7	−3.2	−0.1	.05
$Time^2$	−0.58	−0.74	−0.41	.001	4.1	2.5	5.6	.001
Object(Target)	0.51	0.43	0.59	.001	−1.6	−2.3	−0.8	.001
Learning(Learned)	0.08	−0.01	0.17	.093	−1.8	−2.7	−0.9	.001
Object(Target) X Learning(Learned)	0.92	0.75	1.08	.001	−2.7	−4.2	−1.2	.001