CHILD DEVELOPMENT

Child Development, July/August 2012, Volume 83, Number 4, Pages 1122–1128

Using Saliency Maps to Separate Competing Processes in Infant Visual Cognition

Nadja Althaus and Denis Mareschal

Birkbeck University of London

This article presents an eye-tracking study using a novel combination of visual saliency maps and ''area-of-interest'' analyses to explore online feature extraction during category learning in infants. Category learning in 12-month-olds (N = 22) involved a transition from looking at high-saliency image regions to looking at more informative, highly variable object parts. In contrast, 4-month-olds (N = 27) exhibited a different pattern displaying a similar decreasing impact of saliency accompanied by a steady focus on the object's center, indicating that targeted feature extraction during category learning develops across the 1st year of life. These results illustrate how the effects of lower and higher level processes may be disentangled using a combined saliency map and area-of-interest analysis.

Categorization is ubiquitous in human cognition. The emergence of a category learning system has therefore been a major field of interest in early cognitive development research (Mareschal & Quinn, 2001; Rakison & Oakes, 2003). One line of research has revealed that early category learning is strongly influenced by the statistical distribution of features in the specific exemplars infants are exposed to during the course of familiarization or habituation. For example, Oakes and Spalding (1997) found that infants familiarized with a sequence of toys containing only very similar items formed a narrow, highly exclusive category, subsequently exhibiting a strong preference for an out-of-category test object. In contrast, infants familiarized with sequences containing a high number of dissimilar objects developed a much less constrained category.

Similarly, Mareschal and colleagues (French, Mareschal, Mermillod, & Quinn, 2004; Mareschal, French, & Quinn, 2000; Mareschal, Quinn, & French, 2002) found that looking times in visual preference categorization tasks could be explained by the distribution of feature values present in the stimuli 3- to 4-month-olds encountered during familiarization. Quinn, Eimas, and Rosenkrantz (1993) had reported an asymmetric categorization result: Infants previously familiarized with cat photographs preferred looking at a dog image over a novel cat, but infants familiarized with dogs showed no preference for cats. Mareschal et al. (2000) demonstrated, using computational modeling, that the observed behavior could be explained by the feature distributions present in the stimuli. In particular, with regard to geometrical surface features, the cats' feature distributions were subsumed by those of the dogs. The authors reasoned that infants familiarized with dogs treated the cat image as yet another exemplar of the previously viewed category, whereas the dog presented after familiarization with cat images did not match the infants' mental representation of the familiar category, causing preferential looking.

Thus, while there is growing evidence that early infant category learning is closely yoked to feature distributions in the environment, little is known about how infants identify and extract the features during learning. Schyns and Rodet (1997) provided evidence that adults extract visual features actively while learning a category, a process modulated by the regularities present in encountered exemplars. If infants also actively discover features from the stimuli they observe, this should be reflected in

We wish to thank Wolf Kienzle and Felix Wichmann for making their ''nonparametric saliency'' software available for our analysis, as well as the parents and infants who took part in this study. This work was supported by European Commission contract MEST-CT-2005-020725.

Correspondence concerning this article should be addressed to Nadja Althaus, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD, United Kingdom. Electronic mail may be sent to nadja.althaus@psy.ox.ac.uk.

© 2012 The Authors
Child Development © 2012 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2012/8304-0002
DOI: 10.1111/j.1467-8624.2012.01766.x

changes to their scan patterns across familiarization. Eye-tracking therefore promises to illuminate the learning process underlying categorization.

To our knowledge, only Quinn, Doran, Reiss, and Hoffman (2009) have previously used eye-tracking to investigate categorization in infancy. These authors assessed whether 6-month-olds' fixation targets during category learning correspond to earlier results suggesting that categorization performance is mainly predicted by head feature distributions (Quinn & Eimas, 1996; Quinn, Eimas, & Tarr, 2001; Spencer, Quinn, Johnson, & Karmiloff-Smith, 1997). Quinn et al. (2009) confirmed the presence of a head preference, but found no changes in looking patterns across familiarization. We propose that the ability to extract features actively may develop over time, and therefore compare looking behavior during category formation in two age groups (4- and 12-month-olds). Previous research suggests that categorization abilities undergo fundamental development between 4 and 12 months (Mareschal & Quinn, 2001). It has, for example, been argued that 4-month-olds do not rely on feature correlation for category formation, whereas 10-month-olds readily do so (Younger, 1985).

Moreover, the recent explosion in the use of eye-tracking methods in infancy research suggests a need for rigorous assessment methods (cf. Oakes, 2010). While a link between foveation and attentional focus is regarded as established (e.g., Kowler, 2008), eye movements may be initiated by a number of neural processes. The top-down impact of higher level cognitive processes on eye movements has been demonstrated in the task dependence of scan patterns (e.g., Richardson, Dale, & Kirkham, 2007; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Yarbus, 1967). Yet at the same time, low-level saliency may equally guide fixations (e.g., Reinagel & Zador, 1999). In order to gain insight into categorization, an analysis of fixation patterns must aim to disentangle the effects of low-level saliency and higher order processes. This is especially true for data acquired from infants who cannot be instructed to perform a specific task.

The impact of low-level saliency has been studied extensively in recent years. Several computational methods have been introduced to calculate so-called saliency maps, which predict, for any given image, where a human observer is likely to look. Itti, Koch, and Niebur (1998; see also Itti & Koch, 2000) introduced a biologically inspired approach, which integrates saliency maps with respect to intensity, color, and orientation changes. Itti et al.'s algorithms have, for instance, been used by computational approaches aiming to quantify the saliency of the spatial location of moving or colorful objects in order to construct models of attention in cognitive and social development (e.g., Nagai & Rohlfing, 2009; Schlesinger, Amso, & Johnson, 2007a, 2007b). In contrast to Itti et al.'s approach, Kienzle, Franz, Schölkopf, and Wichmann (2009) derived a saliency filter directly from adult eye movement data acquired during a natural scene viewing task. Image information from typical fixation targets versus nontargets was used to train a Support Vector classifier (Schölkopf & Smola, 2002). This resulted in a decision boundary in the space of luminance patterns of an image patch, which separates typical fixation targets from patches that are never fixated. This decision boundary serves as a filter for novel images, assigning a saliency score to every location in the image based on its similarity to typical fixation targets and nontargets. The advantage of this approach is that it avoids prior assumptions about what image properties determine saliency. This method of constructing a saliency filter furthermore allowed the authors to extract optimally salient stimuli from the learned decision boundary, which were found to be center-surround patterns. Saliency filters like this allow an assessment of the impact that low-level features have on the subject's eye movements.

The current article explores how infants tune in to relevant features during perceptual category learning, by using a combined saliency and area-of-interest (AOI) analysis on infants' eye-tracking data. Groups of 4- and 12-month-olds were familiarized with the category ''deer,'' and successful categorization was assessed in terms of novelty preference for an out-of-category item (a horse). To examine the impact of feature distributions on scan patterns, the deer stimuli were chosen such that only half possessed antlers. We predicted that the high variability of this feature should cause infants to develop a preference for the AOI ''antlers.'' From a computational point of view, variable features carry more information than static features (Liu & Motoda, 1998), as every instance of a variable feature contributes to the similarity structure of the category. Recognition of each depiction of ''antlers'' as an instance of the same feature, and extracting this feature's distributional properties across the whole familiarization sequence, requires processing at the category level, where comparisons between objects can be made. The representation of such complex higher order features goes beyond the early visual processes, and corresponding feature representations have been located in the inferotemporal cortex in monkeys (Logothetis, Pauls, Bülthoff, & Poggio, 1995; Tanaka, 1993; see also Martin, 2007, for similar discussions in the human adult brain). Feature extraction should therefore be regarded as a higher level process in comparison to low-level saliency. Saliency maps may be calculated as early as primary visual cortex (Li, 2002), or even subcortically, as suggested by the fact that features optimally predicting fixation targets have been found to resemble center-surround patches (Kienzle et al., 2009). We hypothesized therefore that low-level saliency would have a higher impact on fixations at the onset of familiarization, when the target category is not yet established.

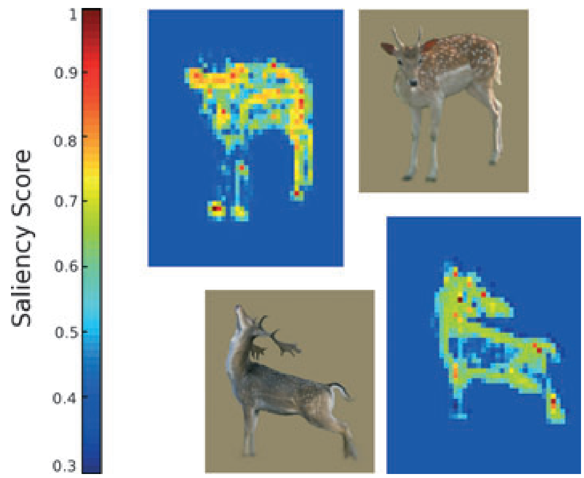

Figure 1. Examples of familiarization stimuli and corresponding saliency maps.

Method

Participants

Twenty-seven 4-month-olds (mean age = 125 days, range = 109–138 days; 11 girls) and twenty-two 12-month-olds (mean age = 363 days, range = 349–379 days; 13 girls) were included in the following analyses. Ten additional 4-month-olds and 8 additional 12-month-olds completed the procedure but were not included in the analysis due to fussiness or failure to reach the looking time criterion. Infants were recruited through advertisements in local parenting magazines and were from the Greater London area.

Stimuli

The familiarization stimuli for each individual stimulus sequence were randomly drawn from a set of 20 photographs of deer in varying postures (see Figure 1). Of the deer depicted in the photographs, only 10 had antlers. Images were depicted against a beige background (subtending 720 pixels/31 degrees visual angle in the horizontal, 480 pixels/22 degrees visual angle in the vertical direction) on either the right or left half of the screen. The stimulus location was counterbalanced across trials. The out-of-category stimuli were randomly drawn from a set of four horse photographs. Test images consisted of a horse paired with a deer. The location of the out-of-category stimulus (left–right) was counterbalanced across subjects.

Procedure

After a short warm-up session in the reception room, infants were seated on the caregiver's lap at a distance of approximately 55 cm from a 17.5 in. screen. The caregiver was instructed not to interfere with the infant in any way. Infants' eye movements were recorded throughout the procedure using a Tobii 1750 remote eye-tracker (sampling frequency: 50 Hz, system accuracy: 0.5 degrees visual angle). A 5-point infant calibration was conducted prior to the experiment, using attractive animated objects. Calibration was repeated up to three times or until all 5 points had been calibrated successfully. Only infants with at least three (4-month-olds) or four (12-month-olds) successfully calibrated points were included in the subsequent analyses. Infants were then familiarized with a sequence of eight randomly selected deer images displayed for 5 s each. Prior to each trial animated videos of objects were displayed in the center of the screen for 1 s, accompanied by sounds, to reattract attention. After familiarization, infants saw two test trials, lasting 10 s each, on which a previously unseen deer was shown side-by-side with a horse. The horse's position on the first trial was counterbalanced across infants. The second test trial was identical to the first, but positions of the stimuli were reversed. All fixations recorded were included for both the saliency and AOI analyses described below.

Saliency Map Analysis

Saliency maps for each stimulus were obtained using the saliency filter software developed by Kienzle et al. (2009). This software assigns a saliency value to each location in the image, indicating likely saccade targets. Figure 1 shows the saliency maps calculated for two of the deer images. In using these saliency maps to analyze infants' looking data, we make the tacit assumption that bottom-up visual saliency is the same for adults and infants. We consider this to be appropriate since the algorithm deals with images on a coarse resolution scale for which 4-month-olds' contrast sensitivity is equivalent to that of adults. All stimuli were rescaled to match the conditions under which the saliency filter was derived, and saliency values were obtained for all fixation locations. To evaluate the degree to which infants' fixation targets corresponded to highly salient image regions, we used the Wilcoxon–Mann–Whitney (WMW) statistic, which gives the probability that a randomly chosen fixation location has a higher saliency value than a randomly chosen location that was not fixated. To calculate WMW scores, a set of nontarget locations was constructed for each trial: These were randomly selected coordinates recorded as fixation targets for other stimuli that had never been fixated in the current image. This was to ensure that nontarget locations came from the same spatial distribution as target locations (cf. Kienzle et al., 2009). In order to avoid sampling biases, the analysis was based on average WMW scores obtained from 10 different random samples of nontarget locations. Trials with less than four fixations were excluded since the WMW score is not reliable for such few data points. Fixations falling within 1 cm of the object were mapped onto the nearest within-object pixel to avoid underestimating the saliency value of contour fixations.

We also calculated the mean saliency of each AOI as the average saliency value assigned to pixels within that AOI. A one-way ANOVA with factor AOI revealed that average saliency scores differed between the AOIs, $F(4, 95) = 16.13$, $p < .0001$, with ''antlers'' being the least salient AOI overall ($M = .38$, $SD = .05$), and ''body'' ($M = .49$, $SD = .05$) and ''head'' ($M = .47$, $SD = .03$) being the most salient AOIs.

AOI Analysis

To assess infants' looking toward individual object parts, AOIs were constructed to represent legs, tail, body, head, and antlers. This was achieved by the following semiautomatic procedure. ''Anchor points'' representing the center of each object part were first obtained for each image from adult annotation. A Matlab script then constructed nonoverlapping AOIs by iteratively expanding the lengths and widths of a small rectangle centered on each anchor point until the resulting AOI ceased to cover more fixation points than the previous AOI. Thus, borders were constructed in a data-driven way. This prevented data points from being excluded due to an arbitrary selection of AOI borders. Importantly, the automatic procedure ensured that AOI construction was based on the same, objective criteria for all stimuli.

Results

We first examined whether infants had formed a category of deer, then analyzed their inspection patterns during learning.

Categorization Performance

Because of order effects apparent on the second test trial (a common finding in habituation or familiarization designs; see Schöner & Thelen, 2006, for a full discussion), we focus here on the first test trial (see online Supporting Information Appendix S1 for results from Test Trial 2). Two-tailed t tests revealed that both 4-month-olds, $t(22) = 3.39$, $p < .01$, and 12-month-olds, $t(18) = 2.15$, $p < .05$, exhibited a significantly greater than chance level novelty preference on the test trial, spending 64.7% and 58.0% respectively of the total looking time fixating the novel item (horse).

Saliency Analysis

Figure 2 shows WMW scores for infants in the two age groups during the two familiarization blocks. A mixed design ANOVA with factors block (Trials 1–3 vs. Trials 6–8) and age (4-month-olds vs. 12-month-olds) revealed significant main effects of block, $F(1, 44) = 6.93$, $p < .05$, and age, $F(1, 44) = 10.70$, $p < .005$. The interaction of Block × Age was not significant, $F(1, 44) = .58$, $p > .40$. A planned comparison confirmed that WMW scores were

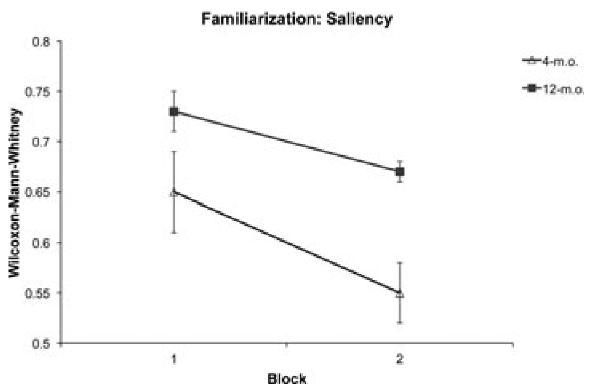

Figure 2. Mean Wilcoxon-Mann–Whitney statistics for familiarization Blocks 1 (Trials 1–3) and 2 (Trials 6–8).

higher during Block 1 than during Block 2, $t(45) = 2.81$, $p < .01$. Thus, infants in both age groups were influenced by saliency to a greater degree at the beginning of familiarization than at the end. Furthermore, the 12-month-olds had higher WMW scores than the 4-month-olds on average, $t(39.019) = 3.68$, $p < .005$, compensated for unequal variances.

AOI Analysis

Looking patterns exhibited during familiarization by infants in the two age groups are depicted in Figures 3 and 4. These data were subjected to an ANOVA with Age (4-month-olds vs. 12-month-olds) as a between-subjects factor, and with Block (trials 1–3 vs. 6–8) and AOI (legs, tail, body, head, and antlers) as within-subject factors. The ANOVA revealed a significant effect of AOI, $F(4, 188) = 43.59$, $p < .0001$, and significant interactions of AOI × Age, $F(4, 188) = 12.49$, $p < .0001$, and AOI × Age × Block, $F(4, 188) = 2.48$, $p < .05$. The three-way AOI × Age × Block interaction could be explained by a significant decrease across blocks in looking toward the ''head,'' $F(1, 84) = 6.49$, $p < .05$, coupled with a significant increase in looking to ''antlers,'' $F(1, 84) = 5.64$, $p < .05$, in 12-month-olds, with no such changes appearing in the 4-month-olds' looking. All other effects and interactions were nonsignificant (see online Supporting Information Appendix S1 for notes on the statistical analysis).

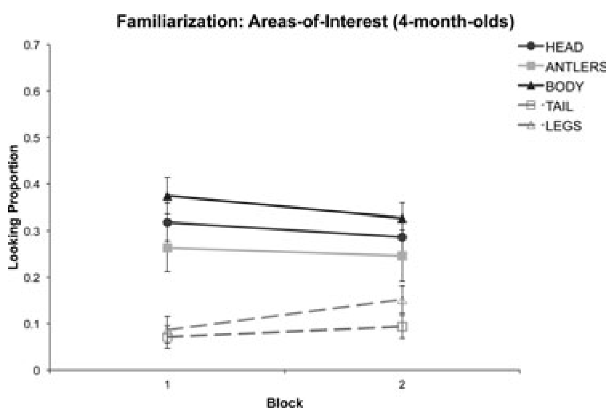

Figure 3. Four-month-olds' looking at areas of interest during Blocks 1 and 2 of familiarization.

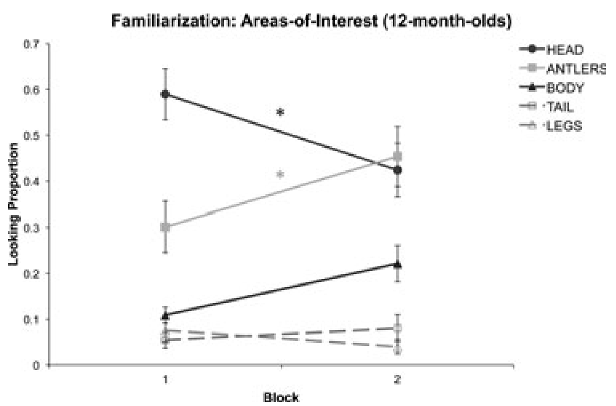

Figure 4. Twelve-month-olds' looking at areas of interest during Blocks 1 and 2 of familiarization.

Discussion

Results from both saliency and AOI analyses revealed converging evidence for the learning processes involved in familiarization with the deer category at two points in development. At the onset of familiarization, eye movements in both age groups were to a large extent driven by bottom-up saliency, but the influence of these early visual processes decreased across familiarization. The 4-month-olds had lower WMW scores overall. Thus, it is possible that the younger infants are less sensitive, less responsive, or slower at responding to local saliency than the 12-month-olds. Alternatively, the lower WMW scores may reflect a higher proportion of looks directed at the object's center, as suggested by the high looking proportions recorded for the AOI ''body.'' In this age group, the decrease in WMW scores was accompanied by only small changes in looking to individual AOIs. In contrast, 12-month-olds underwent a transition from a strong preference for the AOI ''head'' (often containing highly salient regions) to a preference for the high-variability feature ''antlers,'' which indicates the presence of an information-driven feature extraction process occurring during familiarization. Taken together, the results from saliency and AOI analyses suggest that the learning process behind categorization involves a change from low-level to higher order dominance in steering eye movements. Prior to the transition, looking is mainly driven by bottom-up saliency, whereas afterward it is driven by higher level, higher order processes where semantically meaningful comparisons between objects can be made. Given the low average saliency scores of the antlers, the increased fixations toward this object part are unlikely to be due to low-level attractiveness. Extracting feature variability, on the other hand, requires higher order processes because it involves computations across complex features whose distribution only emerges across multiple exemplars. Familiarization may therefore enable infants to disengage from looking at highly salient regions and systematically process object parts, which in turn allows extracting feature distribution information useful for constructing a category representation. The differences we found between 4- and 12-month-olds further suggest that the ability to engage in this systematic learning develops over time. While the 12-month-olds' head-preference is consistent with Quinn et al.'s (2009) report on familiarization with cats and dogs, the lack of changes in looking patterns across learning in their data from 6-month-olds supports the hypothesis that targeted feature extraction develops only toward the end of the 1st year of life.

In summary, the presented work has given a first direct insight into infants' on-line feature processing during the acquisition of a perceptual category. Our findings highlight the interaction between bottom-up visual processes and top-down selection of looking targets driven by factors related to category learning. Finally, this work underscores the importance of combining saliency and AOI analyses to disentangle the effects of low-level visual saliency and higher level cognitively relevant processes on infants' eye movements in order to identify learning-related behaviors.

References

French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3- to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology: General*, *133*, 382–397.

Harrell, F. E. (2001). *Regression modeling strategies*. New York: Springer.

Howell, D. C. (2007). The treatment of missing data. In W. Outhwaite & S. P. Turner (Eds.), *The Sage handbook of social science methodology* (pp. 208–224). London: Sage.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 1254–1259.

Kienzle, W., Franz, M., Schölkopf, B., & Wichmann, F. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of Vision*, *9*, 1–15.

Kowler, E. (2008). Attention and eye movements. In R. Krauzlis (Ed.), *Encyclopedia of neuroscience* (pp. 605–616). Amsterdam: Elsevier.

Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*, 9–16.

Liu, H., & Motoda, H. (1998). *Feature selection for knowledge discovery and data mining*. Norwell, MA: Kluwer Academic.

Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, *5*, 552–563.

Mareschal, D., French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology*, *36*, 635–645.

Mareschal, D., & Quinn, P. (2001). Categorization in infancy. *Trends in Cognitive Sciences*, *5*, 443–450.

Mareschal, D., Quinn, P., & French, R. (2002). Asymmetric interference in 3-to-4-month-olds' sequential category learning. *Cognitive Science*, *26*, 377–389.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45.

Nagai, Y., & Rohlfing, K. (2009). Computational analysis of motionese toward scaffolding robot action learning. *IEEE Transactions on Autonomous Mental Development*, *1*, 44–54.

Oakes, L. M. (2010). Infancy guidelines for publishing eye-tracking data. *Infancy*, *15*, 1–5.

Oakes, L. M., & Spalding, T. (1997). The role of exemplar distribution in infants' difierentiation of categories. *Infant Behavior and Development*, *20*, 457–475.

Quinn, P., Doran, M., Reiss, J., & Hoffman, J. (2009). Time course of visual attention in infant categorization of cats versus dogs: Evidence for a head bias as revealed through eye tracking. *Child Development*, *80*, 151–161.

Quinn, P., & Eimas, P. (1996). Perceptual cues that permit categorical differentiation of animal species by infants. *Journal of Experimental Child Psychology*, *63*, 189–211.

Quinn, P., Eimas, P., & Rosenkrantz, S. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, *22*, 463–475.

Quinn, P., Eimas, P., & Tarr, M. (2001). Perceptual categorization of cat and dog silhouettes by 3-to 4-month-old infants. *Journal of Experimental Child Psychology*, *79*, 78–94.

Rakison, D. H., & Oakes, L. M. (Eds.). (2003). *Early category and concept development: Making sense of the blooming buzzing confusion*. New York: Oxford University Press.

Reinagel, P., & Zador, A. (1999). Natural scene statistics at the centre of gaze. *Computational Neural Systems*, *10*, 1–10.

Richardson, D., Dale, R., & Kirkham, N. (2007). The art of conversation is coordination: common ground and the coupling of eye movements during dialogue. *Psychological Science*, *18*, 407–413.

Schlesinger, M., Amso, D., & Johnson, S. P. (2007a). Simulating infants' gaze patterns during the development of perceptual completion. In L. Berthouze, C. G. Prince, M. Littman, H. Kozima, & C. Balkenius (Eds.), *Proceedings of the seventh international workshop on epigenetic robotics: Modeling cognitive development in robotic systems*

(pp. 157–164). Sweden: Lund University Cognitive Studies.
Schlesinger, M., Amso, D., & Johnson, S. P. (2007b). The neural basis for visual selective attention in young infants: A computational account. *Adaptive Behavior*, *15*, 135–148.
Schölkopf, B., & Smola, A. (2002). *Learning with Kernels: Support vector machines, regularization, optimization, and beyond*. Cambridge, MA: MIT Press.
Schöner, G., & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review*, *113*, 273–298.
Schyns, P., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology. Learning, Memory and Cognition*, *23*, 681–696.
Spencer, J., Quinn, P., Johnson, M., & Karmiloff-Smith, A. (1997). Heads you win, tails you lose: Evidence for young infants categorizing mammals by head and facial attributes. *Early Development and Parenting*, *6*, 113–126.
Spivey, M., Tanenhaus, M., Eberhard, K., & Sedivy, J. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*, 447–481.
Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, *262*, 685–688.
Yarbus, A. (1967). *Eye movements and vision*, New York: Plenum.
Younger, B. (1985). The segregation of items into categories by ten-month-old infants. *Child Development*, *56*, 1574–1583.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix S1. Additional Test Data and Notes on Statistical Analysis.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.