Consistency and variability in two children's home visual environment across development

Anonymous CogSci submission

Abstract

Include no author information in the initial submission, to facilitate blind review. The abstract should be one paragraph, indented 1/8 inch on both sides, in 9˜point font with single spacing. The heading 'Abstract' should be 10˜point, bold, centered, with one line of space below it. This one-paragraph abstract section is required only for standard six page proceedings papers. Following the abstract should be a blank line, followed by the header 'Keywords' and a list of descriptive keywords separated by semicolons, all in 9˜point font, as shown below.

Keywords: Add your choice of indexing terms or keywords; kindly use a semi-colon; between each term.

Introduction

What do children tend to see in their everyday lives? While an understanding of children's visual environment is central to both theories of language acquisition and visual development, we know remarkably little about the categories and objects that tend to be in the infant view, or how they are experienced. For example, how often do infants tend to see animals in real-life vs. in storybooks or as toys? How consistent are the broad characteristics of children's visual environment across individuals and across developmental time?

Over the past decade, researchers have begun to answer these questions by documenting the infant egocentric perspective using head-mounted cameras (Franchak, Kretch, Soska, & Adolph, 2011; Yoshida & Smith, 2008),quantifying the degree to which there are substantial shifts in infants viewpoints that may have downstream developmental consequences.

Indeed, as adults it is hard to intuit how strange this viewpoint can be, and how much it varies across development, transitioning over the first two years of life from close-up views of faces to restricted views of hands manipulating objects (**???**; Fausey, Jayaraman, & Smith, 2016) and children's postural developments shaping what they see (**???**). Most work, however, has focused on documenting the social information that infants and children have access to across early development (**???**; Fausey et al., 2016; Yoshida & Smith, 2008).

More recent research has made progress towards understanding what objects tend to be the infant view, starting with annotating the basic-level categories (e.g., spoons, cups) in the view of 8-month-olds during mealtime. This research suggests that a small number of objects are both pervasively present and among infants' first-learned words (Clerkin, Hart, Rehg, Yu, & Smith, 2017), pointing towards a link between visual experience and early word learning. These findings suggest that a more complete understanding of the visual environment of infants and young children could yield insights about the inputs to both category learning and word learning.

Here, we take a step towards characterizing the visual environment of young children by analyzing the categories of objects (e.g., animals, vehicles, toys, people) present in the infant view in a longitudinal corpus of head-mounted camera data (**???**). We choose to annotate these broad categories – rather than the basic-level identities of objects – for three reasons. First, as our corpus is sampled densely from only a few children, an understanding of the particular objects in their viewpoints may not be particularly generalizable. Second, the basic-level identities of many objects – especially certain toys or objects viewed from odd angles – are sometimes ambiguous (and therefore would be inconsistent across annotators). Finally, these less detailed annotations allow us to collect annotations for more frames from the dataset and therefore license broader coverage. We thus collected human annotations of a randomly sampled set of 24,000 frames from two children in the longitudinal dataset, allowing the analysis of the proportion of broad categories in the infant view from 6-32 months of age.

Using these annotations, we conducted three sets of analyses. First, we examined whether the proportion of animals vs. inanimate objects would be relatively equal in the infant view. A long literature has documented that even newborns have a tendency to attend to animate agents (**???**), and visual cortex dedicates a remarkable amount of space to making fine-grained distinctions among different animals (**???**). Furthermore, animal words tend to be among children's first-learned words (CITES). However, at present, it is unknown whether (non-human) animates (e.g., cats, dogs, other animals) are prevalent in the infant viewpoint, and whether they tend to be exemplars of real-life animals (e.g., ducks at a park, pet dogs) or mostly illustrations in storybooks or as toy stuffed animals.

Second, we examined the co-occurrence between these broad object categories across visual scenes. While some activity contexts (e.g., storytime) and lead to intuitive co-occurrences between object categories (e.g., between books and people), not all activities are intuitive or consistent. We conducted a set of data-driven, exploratory analyses of the

co-occurrence statistics of these broad categories to identify other, reliable patterns in how infants experience their visual world.

Finally, prior work documenting the proportion of faces/hands in view has suggested some developmental changes in how children experience their visual across this same age range–including in this same dataset (**???**). Thus, one possibility is that as children learn to crawl and walk on their own (**???**, **???**; Franchak et al., 2011), some categories that children are likely to interact with (i.e., toys, small objects) could become more prevalent in the child's view across age, whereas other, more stable categories (i.e., furniture) might show relative consistency. On the other hand, the broad characteristics children's visual environments may be relatively stable and determined mostly by the activities that they tend to engage in. We thus examined this hypothesis by exploring the prevalence of each of the broad categories that were annotated across developmental time and across both children.

Method

Dataset

Annotation procedure

Preprocessing

Results

Part 1: Category prevalence

Lots of people generally speaking (and partial views of furniture) Equal proportion of toys vs. real for animals/vehicle Not that much variance in exemplars for reals – mostly dog/cats for animals Real vehicles are often interiors of vehicles (can quantify how many easily) Animacy/size prevalence comparisons More objects > animals in general Even though animal names are learned early Suggests role for heightened attention to animals Small objects broad category of stuff they see Some of this measured in mealtime, but a lot more variability here than just food/utensils (toys/other small objects, etc) No developmental change in prevalence AFAWCT (maybe in stereotyped activities, not done yet) Is there a dip for toys? (less play as learning locomotion?)

Part 2: Co-occurrence of categories in visual scenes

Stereotyped activities with/without parent Separation of these activities by child-alone vs. with parent (adult-hand or adult-face present) Mealtime (food + utensil?) Storytime (book + person) or book on own Joint playtime (not book + yes toy + person) Data-driven: Correlogram of different categories (and hierarchical clustering to identify similar/dissimilar contexts) Might identify the candidate contexts that we can think of, and maybe others that we can't (e.g., perhaps playtime outside)

General Discussion

Discussion Main findings Relative consistency in the visual environment over age, posture/other changes may be having more fine-grained impacts on HOW children see objects (and the diversity at the basic-level) Stereotyped activities are not that frequent in kid's experience, even though we think of them often (and this is parents putting headcam's on their kids, so maybe they are even overrepresented relative to normal) Kids learning from both real and toy examples in equal measure toys/depictions are real part of kid's representations Role of extra attention to animals (i.e. saliency of animate objs) in learning animals/their names, not just pure frequency

Future work Relationship to early learned words for these categories need to drill down to basic-level categories to assess this hypothesis Generalization to other datasets Fine-tuning/training of models to do so and make broader conclusions

Limitations WEIRD kids who have a lot of books and toys and may not go outside as much may really vary across kids, SES, contexts, etc Yet suggests that at least some kids are learning quite a lot from books and depictions Only 2 kids – though we see consistency here When parents choose to wear headcam – not getting time at daycare,etc Nonetheless, broad findings that objects > animals and consistency across age are things we predict will generalize broadly beyond this small population

Acknowledgements

Place acknowledgments (including funding information) in a section at the end of the paper.

References

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. (2017). Real-world visual statistics and infants' first-learned object names. *Phil. Trans. R. Soc. B*, *372*(1711), 20160055.

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

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*(6), 1738–1750.

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

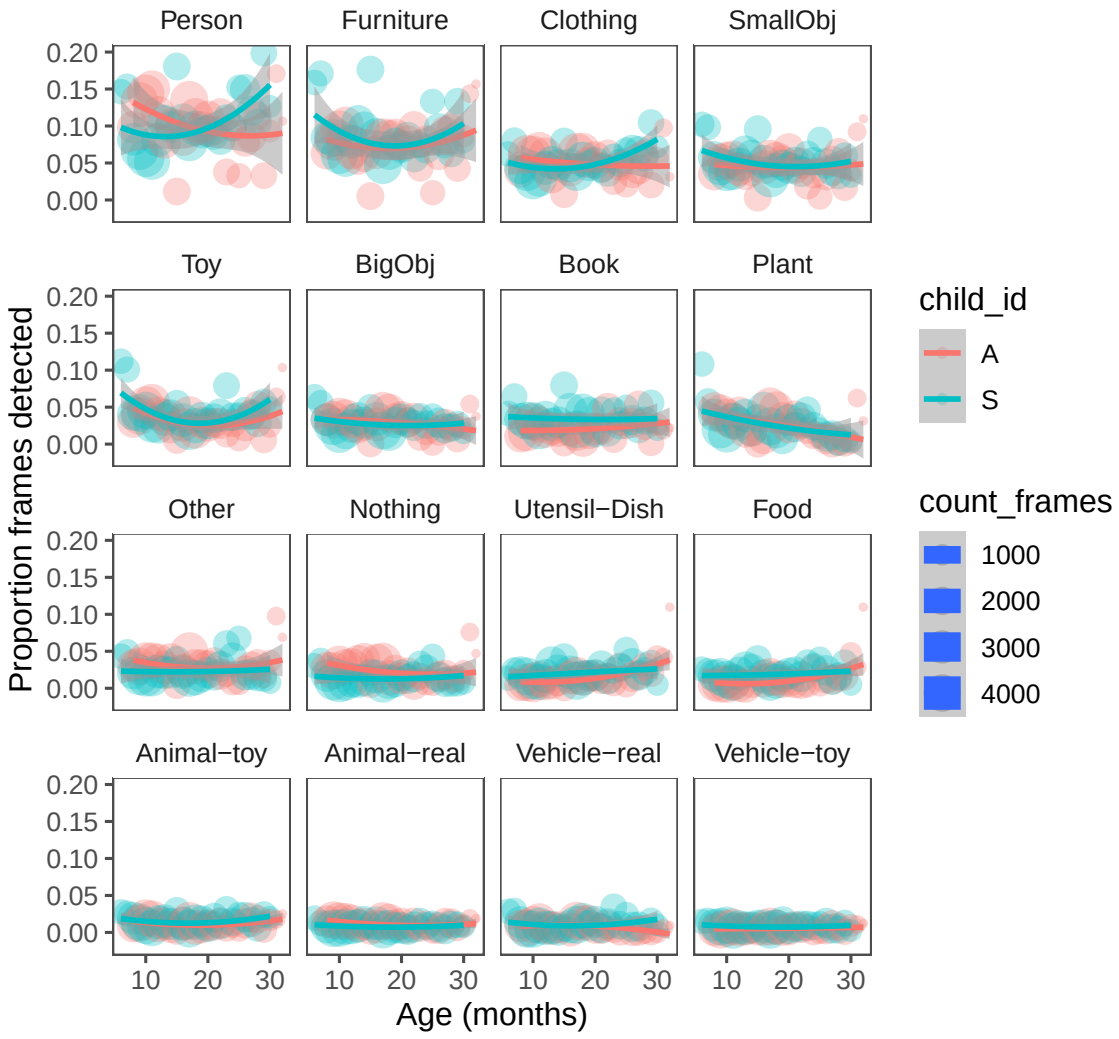

Figure 1: Frequency of categories annotated across the 24K random frames plotted as a function of each child's age (in months).

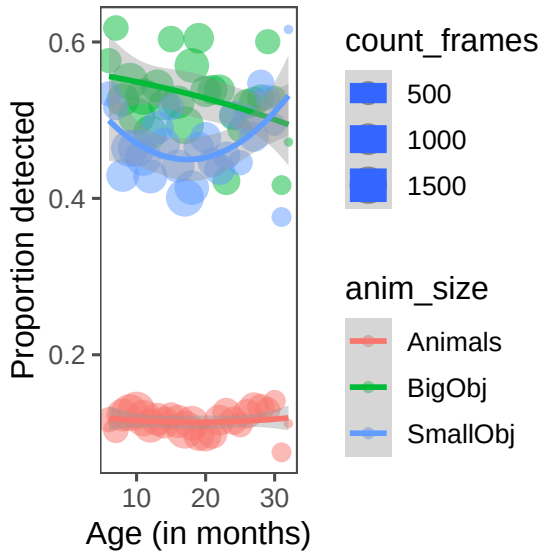

Figure 2: Frequency of animals (including toys) relative to big and small inaniamte objects detected in the dataset.

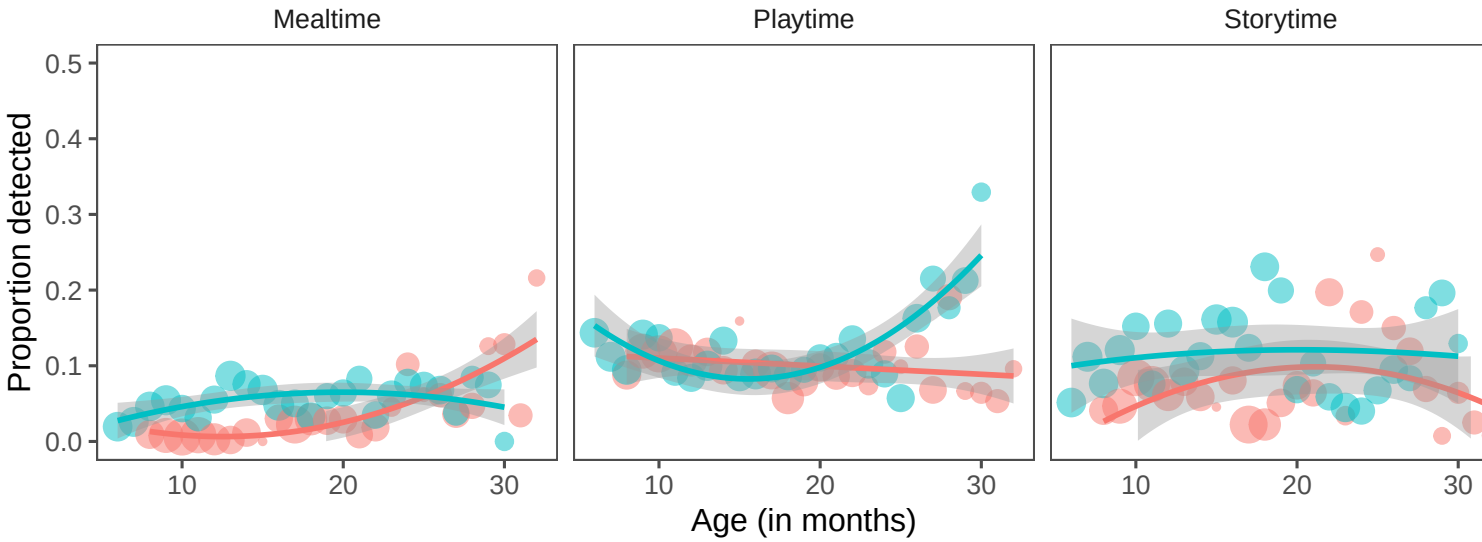

Figure 3: Frequency of mealtime (food = utensils), storytime (book + person), and playtime (e.g., toy + people) activities for each child across age in the randomly sampled frames.

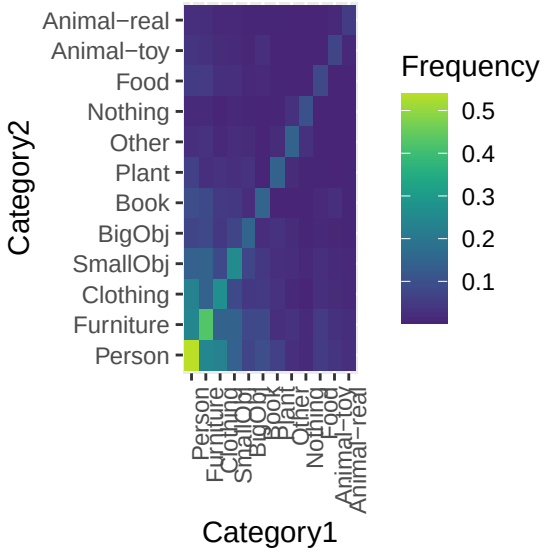

Figure 4: Co-ccurance between different categories detected in the dataset across all frames.