# Knowledge differences affect gaze behavior during naturalistic object exploration

**Amanda J. Haskins, Tarun Sepuri, Bria Long**
{ajhaskins, tsepuri, brlong}@ucsd.edu
Department of Psychology, UC San Diego, 2941 Muir Lane, La Jolla, CA 92093

## Abstract

**How does knowledge affect visual exploration? Here, we investigated whether learning about an object would change participants' gaze behavior. Adults (*N* = 22) wore a head-mounted eye tracker while exploring real-world objects that were relatively unfamiliar (e.g., a rolodex). We assessed learning-related differences in both (1) oculomotor behavior (e.g., fixation durations) and (2) participants' fixated visual content by leveraging embeddings from a large vision-language model. We find evidence that before learning, gaze is more exploratory (i.e., shorter, more frequent fixations). Moreover, we find that differences in fixated content across participants are increased after learning about an object, suggesting that knowledge states may contribute to differences in gaze behavior. These results underscore the importance for future work of quantifying individual knowledge states to further leverage naturalistic eye tracking as a window into learning and cognitive development.**



**Keywords: naturalistic vision; head-mounted eye tracking**


## Introduction

How does what we know influence where and how we look? Research with adults provides some clues that individual differences in semantic knowledge can change gaze behavior: for example, individuals with subject-matter expertise in certain domains (e.g., medical imaging) tend to visually explore in different ways than non-experts (Drew et al., 2013). But differences in semantic knowledge are not limited to expertise: individuals' knowledge about particular objects or contexts can be shaped by the accumulation of everyday experiences. Indeed, in studies of cognitive development, infants' gaze has been a critical tool for inferring knowledge states (Kunin et al., 2024; Margoni et al., 2024); preverbal infants use eye movements to explore and learn about people and objects in their environments (Yu & Smith, 2011) well before they are able to locomote independently or express their thoughts with words.

However, it is unclear how or whether gaze behavior during everyday learning experiences reflects knowledge differences in either children or adults. Eye tracking paradigms often lack features of everyday viewing, in turn minimizing individual variability. For example, participants are often passively shown static images on a computer screen for a few seconds each (Xu et al., 2014). Moreover, stimuli are often curated such that images that individual participants may have interacted with extensively or not at all are excluded. Although some eye tracking studies have characterized stable individual differences in gaze (Broda & de Haas, 2024; De Haas et al., 2019), it is far less common to characterize variability in individuals' knowledge or examine how it changes gaze behavior.

Understanding how an individual's specific semantic knowledge impacts their gaze is an important step for further advancing gaze behavior as a tool for both cognitive and developmental science. As a first step, here we aimed to understand how adults' gaze during naturalistic viewing is influenced by knowledge differences about everyday but uncommon objects. Adult participants (*N* = 22) engaged in an object learning task while wearing a head-mounted eye tracker. We compared participants' gaze behavior before and after they learned about objects with which we anticipated they would have limited or varying amounts of expertise (e.g., *rolodex, french press, floppy disk*). Participants explored each object visually and manually before and after reading a short informational passage about each object (Fig. 1A). We used these phases as a coarse proxy for participants' knowledge state differences about these objects.

We predicted that lower knowledge states would be marked by more exploratory gaze behavior, and we examined two potential signatures of exploration: 1) changes in oculomotor behavior (e.g., reduced duration, increased number of fixations (Gameiro et al., 2017)) and 2) changes in fixated content (e.g., variation in the specific object views selected by a participant).

## Methods

### Procedure

Participants (*N* = 22 adults; University undergraduates, average age = 20.59 years, range 18 - 25) wore a head-mounted eye tracker (Pupil Labs "Neon") while they explored up to 12 objects in a randomized order: (*stethoscope, french press, three hole punch, fishing reel, rolodex, blood pressure cuff, bulb planter, hand mixer, crank flashlight, shoe horn, pocket radio, floppy disk*). Each trial had exploration blocks, passage reading blocks, and verbal description blocks where participants described what they knew about each object (verbal descriptions not analyzed here). Participants were encouraged to learn as much as they could from the example object by visually exploring and handling it without time limits; participants spent up to an hour in the overall paradigm.

### Analysis

We analyzed gaze behavior before and after learning, using the passage reading block as a coarse proxy for learning across all participants. To characterize oculomotor behavior, we calculated the number and mean duration of fixations for each participant, object, and block (i.e., before and after learning). Fixations were defined using a dispersion threshold within a given window of time. We characterized the visual content that participants selectively attended with each fixation by identifying video frames that occurred during each fixation and then used the fixation coordinates to crop a small portion of the frame ($\sim$20% full FOV, see Figure 1B).

To quantify variation in participants' fixated content we leveraged the embedding space of a large vision-language model, CLIP (Radford et al., 2021), using a ViT-B/32 vision encoder. Thus, the information selected with each fixation in a given trial could be represented as a vector embedding, allowing us to characterize visual content similarity across fixations and participants. Vector embeddings for each fixation were correlated with each other, using cosine similarity, both within participants (i.e., participant's first fixation, second fix-

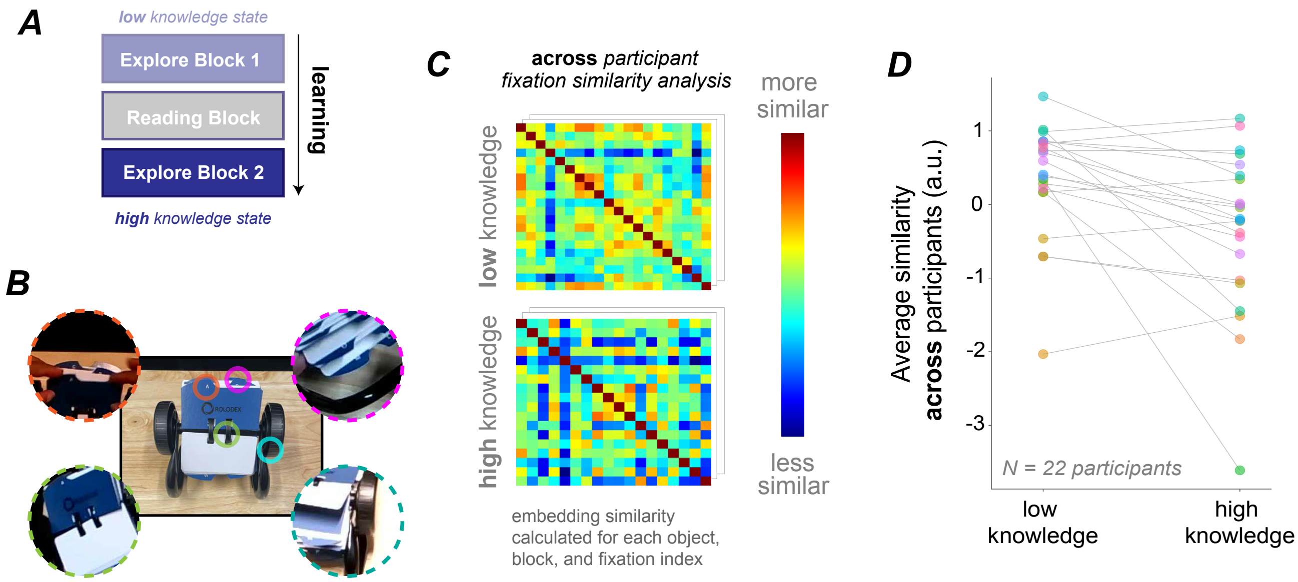


Figure 1: (A) Participants explored each object before and after reading a short informative passage; (B) Participant fixation coordinates were used to crop portions of video frames, which we then used to represent fixated content as vector embeddings within the vision encoder of CLIP (Radford et al., 2021); (C) Embeddings for each fixation were correlated both within and across participants; (D) Fixation embeddings were less similar across participants after they had read an informative passage (i.e. block 2, high knowledge state).

ation, etc.) and across participants (i.e., all participants' first fixations; Fig. 1C). Dependent variables (fixation duration, fixation frequency, fixation embedding similarity) were analyzed using linear mixed-effect models with random slopes of condition (i.e., low vs. high knowledge state block) by items and participants. To account for the generic impact of time (e.g., participants' gaze may change systematically after their initial exposure to the objects, unrelated to learning about the objects) all analyses included fixation index as a covariate. All code and analyses, as well as photographs of object stimuli, are available on OSF: https://osf.io/ndj2p/.

## Results

We found several pieces of evidence that learning influenced participants' gaze behavior. First, participants made significantly more fixations before learning about each object, during lower knowledge state blocks ($M$ = 55.73), than after learning, during higher knowledge state blocks ($M$ = 29.96; fixed effect of condition, $t(23.56) = 6.45$; $p < 0.001$). Moreover, fixation durations during lower knowledge state blocks ($M$ = 335.67ms) were significantly shorter than during higher knowledge state blocks ($M$ = 356.20ms, fixed effect of condition, $t(173.75) = 2.57$, $p = 0.011$).

Next, we explored what specific content participants fixated during object exploration. When examining the fixated visual content, we found that within a participant, fixation embeddings for a given trial tended to be more highly correlated after learning ($t(29.37) = 2.091$, $p = 0.0453$). However, this effect appeared to be driven by the overall reduced number of fixations made after learning. Next, we examined the variability of fixation embeddings *across* participants in different knowledge state blocks: we found that fixation embeddings were less similar across participants after learning, during high knowledge state blocks (fixed effect of condition, $t(25.43) = 2.329$; $p = 0.028$, see Figure 1D). This result could be driven by later-occurring fixation indices made by only a few participants, resulting in noisier estimates of across-participant similarity. To address this possibility, we repeated the across-participants analysis on only fixation indices that exceeded a minimum threshold of participants (N = 10). Critically, we observed the same results pattern in this control analysis.

## Discussion

We examined how differences in knowledge impact naturalistic gaze behavior before and after learning experiences with objects. We found that gaze was more exploratory when participants knew less about an object (i.e., shorter, more frequent fixations); however, we did not find clear within-participant differences in fixation variability before vs. after learning. In addition, across participants, fixations differed more after learning about an object. An important future direction will be to characterize finer-grained knowledge differences in individual participants, and to assess whether idiosyncratic knowledge gaps influence individual differences in gaze behavior. Gaze behavior may offer insight into newly learned knowledge during naturalistic object exploration.

## Acknowledgments

This work was supported by the National Institutes of Health under award number R00HD108386 to B.L.

## References

Broda, M. D., & de Haas, B. (2024). Individual differences in human gaze behavior generalize from faces to objects. *Proceedings of the National Academy of Sciences*, *121*(12), e2322149121.

De Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., & Gegenfurtner, K. R. (2019). Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences*, *116*(24), 11687–11692.

Drew, T., Evans, K., Võ, M. L.-H., Jacobson, F. L., & Wolfe, J. M. (2013). Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? *Radiographics*, *33*(1), 263–274.

Gameiro, R. R., Kaspar, K., König, S. U., Nordholt, S., & König, P. (2017). Exploration and exploitation in natural viewing behavior. *Sci. Rep.*, *7*(1), 1–23.

Kunin, L., Piccolo, S. H., Saxe, R., & Liu, S. (2024). Perceptual and conceptual novelty independently guide infant looking behaviour: a systematic review and meta-analysis. *Nature Human Behaviour*, *8*(12), 2342–2356.

Margoni, F., Surian, L., & Baillargeon, R. (2024). The violation-of-expectation paradigm: A conceptual overview. *Psychological Review*, *131*(3), 716.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., . . . others (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748–8763).

Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S., & Zhao, Q. (2014). Predicting human gaze beyond pixels. *Journal of vision*, *14*(1), 28–28.

Yu, C., & Smith, L. B. (2011). What you learn is what you see: using eye movements to study infant cross-situational word learning. *Developmental science*, *14*(2), 165–180.