How precise is infants' visual concept knowledge?

How precise is infants' visual concept knowledge?

Results

First, we examined the effect of target and distractor similarity on word recognition accuracy across our age range. Word recognition accuracy was assessed at the trial-level using the proportion of looking at the target over the distractor. This measure was incorporated as the difference between looking time in the critical window and baseline window (300 to 3500 ms and -2000 to 0 ms respectively, relative to target word onset), following the rationale described by [**?**]. Contrary to our expectation that word embedding similarity would inversely correlate with infants' looking time, and reveal the partiality of infants' visual concept knowledge, we found no significant change in looking time with change in word embedding similarity. However, exploratory image and multimodal similarity measures did correlate with looking time as expected (see Fig 1B for comparison). We confirmed this qualitative finding with a linear mixed-effects model (image: b=-0.07, p=<.05; text: b=-0.07, p=<.05; multimodal: b=-0.02, p=0.442). This difference in similarity measure predictivity is especially remarkable since the embedding spaces of a single multimodal model are highly correlated ([**?**]).

Next, we examined whether the similarity effect would change with age, expecting the effect of similarity to decrease with age as infants form more precise representations. Surprisingly, we did not find any interaction between infant age and embedding similarity. Corroborating this finding, looking time in general does not change with age with younger infants already exhibiting significant accuracy across all trials (Fig 1C). The nature of infants' representations of visual concepts remains relatively stable across our 14-24 month age range.

Why did text similarity and age not predict looking time? To ensure the robustness of our findings, we measured the effects of item-level differences. This examination is also pertinent since we use a broad range of naturalistic images, an atypical choice in looking-while-listening trials ([**?**]). We examined how differences in target word

age-of-acquisition (AoA) (measured using estimated AoA values from [**?**]) and image pair visual saliency differences (measured using the GBVS toolbox; [**?**]) affected accuracy. Target word difficulty did correlate inversely with looking time (see Fig 1D for split-half comparison), in line with our expectation that more difficult words are harder to recognize. Accuracy was at chance for our most difficult words like 'coconut' and 'swan'. Further, we did not find any interaction between AoA, infant age, and our measures of similarity, highlighting the separate roles that embedding similarity measures and word difficulty play in predicting infants' target looking time. In contrast, visual saliency did not predict infants' looking time. This finding strengthens our result, indicating that our model similarity measures' looking time predictivity cannot be explained only by lower-level perceptual features and establishes multimodal model embedding similarities as a future measure of infants' visual concept knowledge.