The integration of gaze information during real-time language comprehension and learning

Kyle MacDonald[1,2], Anne Fernald[1], Virginia A. Marchman[1], & Michael C. Frank[1]

[1]Stanford University, [2]UCLA

Background

Understanding language in real-time is hard. Even in grounded language comprehension, people can talk about many things, with no guarantee that they use familiar words. Listeners can overcome this ambiguity by integrating visual information available in face-to-face communication (e.g., **the direction of another speaker's gaze**) that constrains the interpretation of an utterance. [1,2,3]

For example, in prior work, we found that children and adults fixated longer on a speaker's face when processing familiar words in a "noisy" environments.[4]

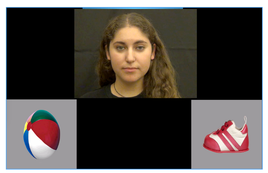
Research Questions

1. **Do listeners strategically seek supportive visual information from other speakers?**
2. **And does the tendency to seek information depend on the listener's uncertainty about word meanings?**

Methods

Hey! Where's the ball/dax ?

Images appear for 2.5s | Carrier phrase | Target noun onset | Gaze cue onset | Center stimulus off



straight ahead

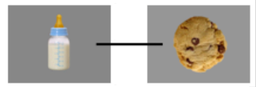
gaze

Average length of gaze cue was 2 sec

Data/code available at https://bit.ly/2FglbsW
E1 preregistration at https://osf.io/2q4gw/
E2 preregistration at https://osf.io/nfz85/

Experiment 1: Familiar words

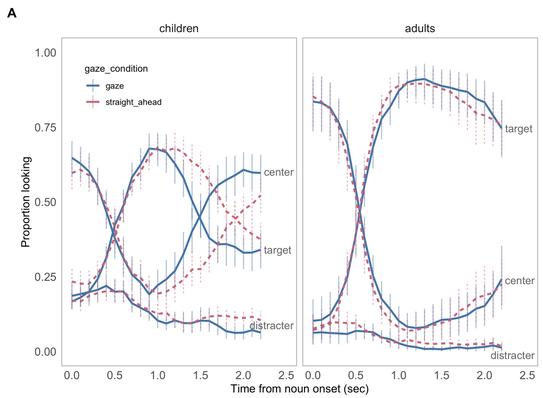
(38 children [3-6.5 y.o.]; 33 adults)



Measurements (SMI RED eye tracker; 30 Hz)

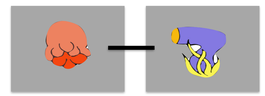
1. **Proportion looking:** mean proportion of trials on which participants fixated on the speaker every 33-ms
2. **First shift reaction time (RT):** latency of shifting gaze from speaker to either object from the onset of target noun
3. **First shift accuracy:** whether first gaze shift landed on the target or the distracter object

Results (32 trials; mixed-effects models; 95% Credible Intervals)

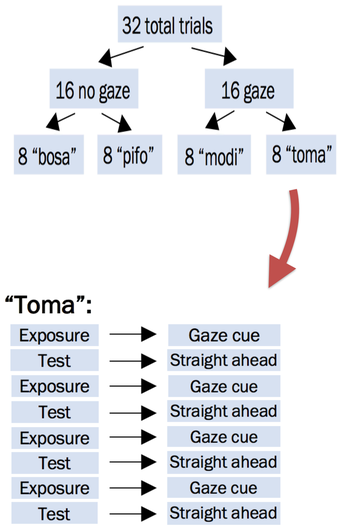


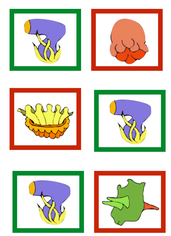
Experiment 2: Novel words

(54 children [3-6.5 y.o.]; 30 adults)



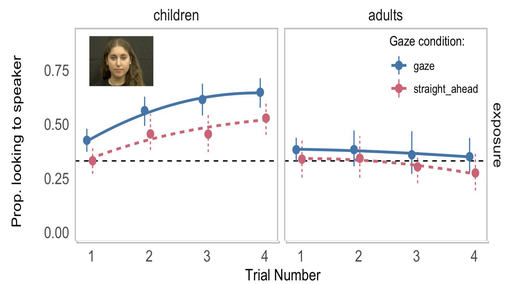
Design (within-subjects)



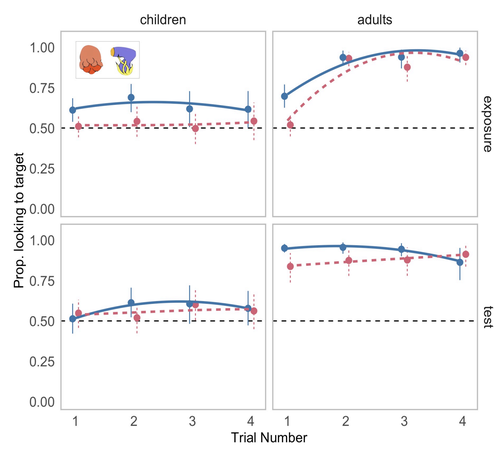


Results (proportion looking)

(A) Looks to speaker



(B) Looks to target object



Take Home

1. No delay of eye movements to seek gaze while processing familiar words
2. When processing novel words, slower shifts to seek a gaze cue. Adults, but not children, showed faster word learning in the presence of gaze

Listeners are sensitive to the informational tradeoffs in active information gathering, seeking social information when uncertainty was higher

1. Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. The Royal Society.
2. McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception?. *Trends in cognitive sciences*, *10*(8), 363-369.
3. Kelly, S.D., Ozyurek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260– 267.
4. MacDonald, K., Marchman, V. A., Fernald, A., & Frank, M. C. (under review). Children flexibly seek visual information during signed and spoken language comprehension.

kemacdonald@ucla.edu