LETTER

REPLY TO LEWIS ET AL.:

Inference is key to learning appearance from language, for humans and distributional semantic models alike

Judy S. Kim[a,1], Giulia V. Elli[a], and Marina Bedny[a]

Two major ways in which humans learn is by direct sensory observation and gathering information from other minds through language. In our original paper, we attempt to tease apart the contributions of sensory experience from other sources of information, including linguistic communication, by comparing knowledge of appearance among individuals blind from birth and those who are sighted. We report that blind and sighted people share structured knowledge of animal appearance (1). Lewis et al. (2) urge us not to "reject language as an important source of visual knowledge." We did not intend to do so and agree that language is indeed likely a key source of appearance information.

The question is how learning appearance from language works. Learning could entail memorizing verbally stipulated appearance facts (e.g., "hippos are gray"). Our data suggest that blind individuals learn animal appearance primarily by inference from taxonomy and habitat: Blind and sighted people are most likely to share appearance knowledge that can be inferred from these dimensions (e.g., "flamingos have feathers" but not "flamingos are pink"). In this case, appearance is gleaned from language indirectly. Language transmits taxonomy and habitat information and their relationship to appearance. The appearance of any particular animal is then inferred. If "florbs" are a type of bird, then they have feathers and wings.

Lewis et al. (2) show that a distributional semantics model (3) applied to text corpora can recover some information about animal appearance. However, on every dimension, blind and sighted people share more with each other than with the model. The pattern of what is learned by the model is also different. The model and humans agree for shape, but not texture, despite both shape and texture having high correspondence across blind and sighted groups. The model makes errors never made by people (e.g., goldfish have feathers). Better performance on shape compared to other dimensions is consistent with shape being predictable from taxonomy. Indeed, some shape features generated by sighted participants and used by Lewis et al. (2) in their analysis are highly diagnostic of taxonomy (e.g., fins, scales, and wings). For color, judgments are better predicted by the model for blind than sighted participants, perhaps also because color judgments are correlated with taxonomy only for the blind group.

These findings are consistent with the idea that language is an excellent transmitter of information about taxonomy but less so of explicit appearance features (4). Blind individuals may well learn taxonomy through language. However, we think this appearance-via-taxonomy hypothesis is more consistent with the available data than the idea that blind individuals learn that lions have fur and not feathers by tracking how often "lion" occurs in similar text contexts to "feathers" and "fur." Since humans do better using inference, we think models would also learn appearance from text better by implicitly or explicitly making inferences across dimensions (e.g., taxonomy to shape) and across exemplars within a class (e.g., if lions have fur then bears do as well).

[1] J. S. Kim, G. V. Elli, M. Bedny, Knowledge of animal appearance among sighted and blind adults. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11213–11222 (2019).

[2] M. Lewis, M. Zettersten, G. Lupyan, Distributional semantics as a source of visual knowledge. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 19237–19238 (2019).

[a]Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD 21218

Author contributions: J.S.K., G.V.E., and M.B. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: judyseinkim@gmail.com.

First published September 5, 2019.

3 T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space. https://arxiv.org/abs/1301.3781 (16 January 2013).

4 D. Rubinstein, E. Levi, R. Schwartz, A. Rappoport, "How well do distributional models capture different types of semantic knowledge?" in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (Association for Computational Linguistics, 2015), vol. 2, pp. 726–730.

19240 | www.pnas.org/cgi/doi/10.1073/pnas.1910410116 Kim et al.