

Distributional semantics as a source of visual knowledge

Molly Lewis^{a,1}, Martin Zettersten^a, and Gary Lupyan^a

In PNAS Kim et al. (1) detail congenitally blind individuals' extensive knowledge of the visual appearance of animals. This is exciting and important work speaking directly to long-standing questions about the role of direct perceptual experience in semantic knowledge. Despite lacking visual input, blind people show substantial alignment with one another and with sighted people in judging animal shape, skin texture, size, and, to a much lesser extent, color. Where does this knowledge come from? One possibility, advanced by the authors, is inferential reasoning. Knowing that birds have feathers and that ostriches are birds allows blind people to infer that ostriches have feathers despite never having seen an ostrich (or feathers). Another possibility is the distributional structure of language. The authors reject the "obvious idea . . . that blind individuals learn from sighted people's verbal descriptions" on the grounds that the lowest correspondence between the sighted and blind groups was found for color, a domain with high name agreement. The 2 groups showed higher alignment for shape and skin texture, domains that are harder to verbally describe.

We think the authors were premature to reject language as an important source of visual knowledge. We show that associative learning algorithms lacking inferential machinery can partially reproduce behaviors of sighted and blind people after being exposed to natural language. These algorithms learn word meanings by attempting to predict what words surround other words (2, 3). The resulting semantic representations show reasonably close correspondence to human judgments (4, 5). By measuring vector distances between representations of animal words

and target words used in ref. 1 (e.g., shark-skin vs. shark-feathers), we are able to subject these models to analogous tests performed by the human subjects. We find that semantic representations learned wholly from language correlate significantly with human judgments of animal similarity on the basis of shape, skin texture, and color (ref. 6 and Fig. 1 A and B). Despite in-principle high name agreement for animal colors, distributional semantics encode animal color much less than they encode shape. Nevertheless, color information encoded in language is still predictive of blind participants' responses. Even if participants' performance were partially based on explicit inference, the question remains: How do blind participants learn the taxonomic relationships in the first place? We find that this information is also embedded in the distributional structure of language (Fig. 1C).

The idea that distributional semantics are a rich source of visual knowledge also helps us to understand a related report (7) showing that blind people's semantic judgments of words like "twinkle," "flare," and "sparkle" were closely aligned with sighted people's judgments ($\rho = 0.90$). The authors again rely on explicit inference to word meanings as an explanation, but here too we find that word meanings learned through distributional semantics (using models lacking inferential machinery) show correspondences with both sighted ($\rho = 0.59$) and blind ($\rho = 0.63$) people's semantic judgments.

People, regardless of sight, doubtless use inferential reasoning to generate new knowledge. However, our results show that embedded within the statistical structure of language is a surprisingly rich repository of visual knowledge.

^aDepartment of Psychology, University of Wisconsin–Madison, Madison, WI, 53706

Author contributions: M.L., M.Z., and G.L. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

Published under the PNAS license.

Data deposition: The data, processing code, and analysis scripts reported in this paper have been deposited in a GitHub repository (https://github.com/mllewis/keb_2019_reanalysis) and posted on Open Science Framework (<https://osf.io/p84yx/>). Supporting results are available on RPNAS (<http://rpubs.com/ml/kebcommentaryS1>).

¹To whom correspondence may be addressed. Email: mollylewis@gmail.com.

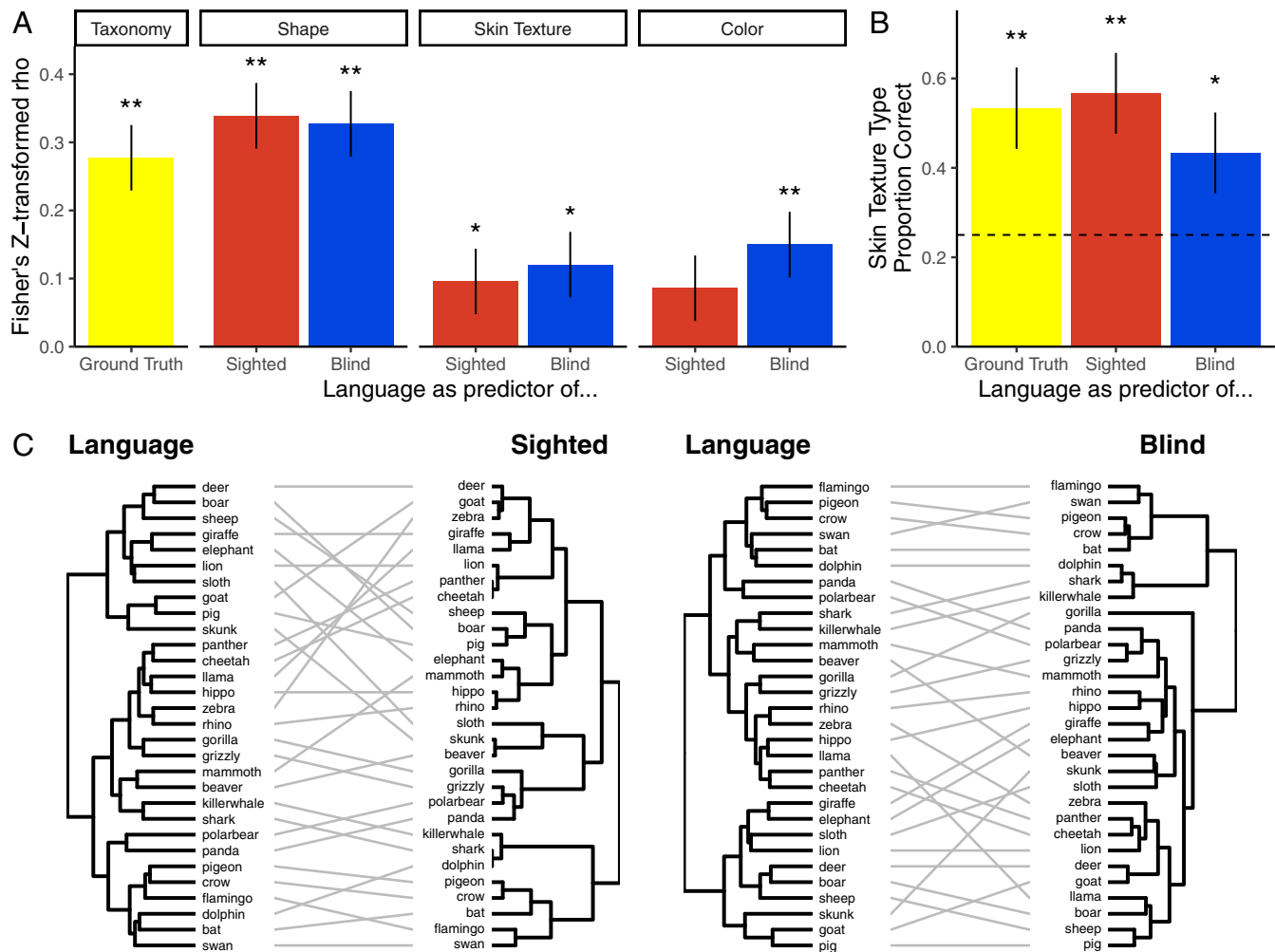


Fig. 1. (A) Pairwise correlations (Fisher-transformed as in ref. 1) between language-derived animal similarities (cosine distances) and evolutionary distances (yellow) and performance of sighted (red) and blind (blue) participants. Human comparisons are shown separately for shape, skin texture, and color card-sorting tasks in ref. 1. **(B)** Language statistics classify animals as having scales, skin, fur, or feathers at levels well above chance. Predictive accuracy is shown for ground truth (yellow) and performance of sighted (red) and blind (blue) participants. **(C)** Relationships among animals computed entirely from language statistics show considerable overlap with relationships derived from shape-based categorization produced by sighted (*Left*) and blind (*Right*) participants. The labels in the 2 language-based dendrograms are ordered to maximize alignment with behavioral data. * $P < 0.05$; ** $P < 0.01$.

- 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 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).
- 3 P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching word vectors with subword information. <https://arxiv.org/abs/1607.04606> (15 July 2016).
- 4 F. Hill, R. Reichart, A. Korhonen, Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Comput. Linguist.* **41**, 665–695 (2016).
- 5 D. Gerz, I. Vuli, F. Hill, R. Reichart, A. Korhonen, Simverb-3500: A large-scale evaluation set of verb similarity. <https://arxiv.org/abs/1608.00869> (2 August 2016).
- 6 M. Lewis, M. Zettersten, G. Luypan, Data from “Distributional semantics as a source of visual knowledge.” Open Science Framework. <https://osf.io/p84yx/>. Deposited 6 August 2019.
- 7 M. Bedny, J. Koster-Hale, G. Elli, L. Yazzolino, R. Saxe, There’s more to “sparkle” than meets the eye: Knowledge of vision and light verbs among congenitally blind and sighted individuals. *Cognition* **189**, 105–115 (2019).