Preserved Structure Across Vector Space Representations

Andrei Amatuni, Estelle He and Elika Bergelson | Department of Psychology and Neuroscience, Duke University

Research Questions:

1) Is cross-category structure preserved across visual & linguistic representations?

2) If so, does this influence infant learning?

Background

Concepts, Percepts, and Words:

- During language learning, infants carve the world into distinct lexical entities.
- This requires integrating sensory and linguistic input into appropriate language-specific categories.
- It is unclear how <u>independent</u> and <u>separable</u> conceptual and sensory input are.
- Previous work disagrees about how and when categories are formed vis-à-vis conceptual and perceptual input. e.g.:
  - percepts then concepts [6, 7, 8]
  - distinct processes in parallel [9]

Category judgments:

- functional information (e.g. animacy) can shift categories [11]
- preschoolers can override perceptual overlap in natural kinds [10]

Similarity and learning:

- Human learners integrate perceptual and linguistic information during categorization and learning [12, 13, 14]
- Similarity makes category/word learning more difficult [15, 16]

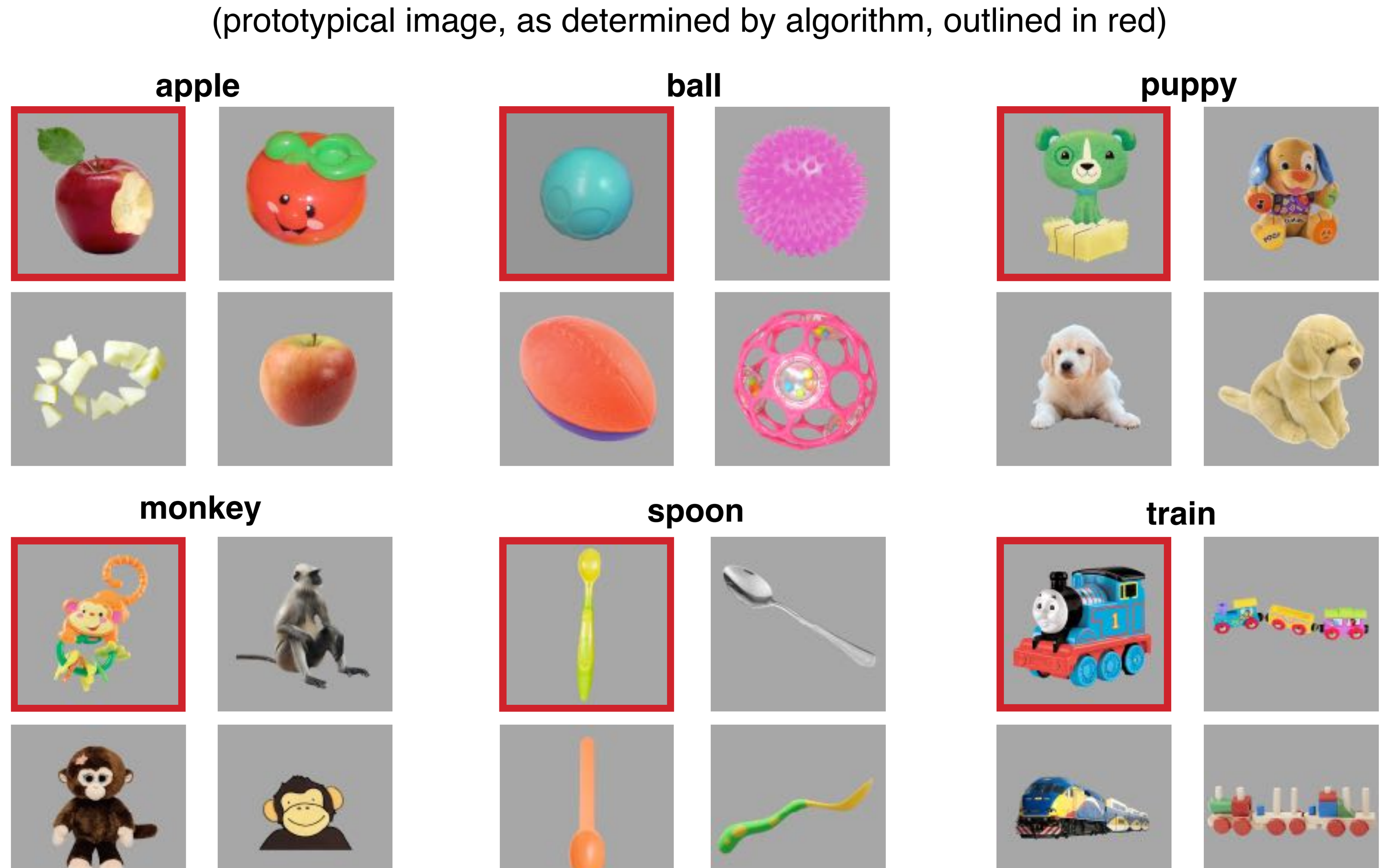
Methods

Items: common nouns (object words) heard & seen by 44 6-17m.o.'s in the SEEDLingS Corpus, an audio & video corpus of infants in their home environment [1, 2]

- Photos of top items from home recordings were used to test infants' word comprehension in-lab, creating an image database of exemplars of these nouns from children's lives.
- Items with at least <u>9 image exemplars</u> were used here (n=27).

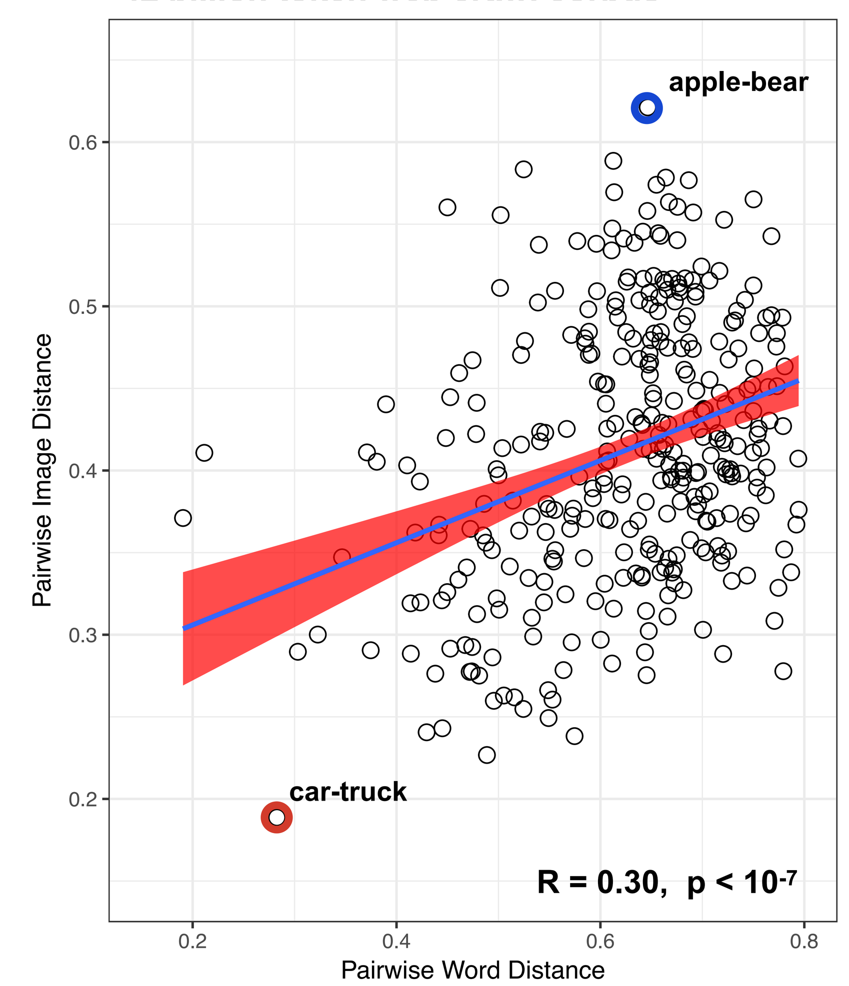
Each item had a corresponding **word,** i.e. the lemma for the relevant ***concept***, and a set of **images,** i.e. examples of the ***perceptual*** input infants saw when they heard these words.

Example Items (Images and Words)
(prototypical image, as determined by algorithm, outlined in red)

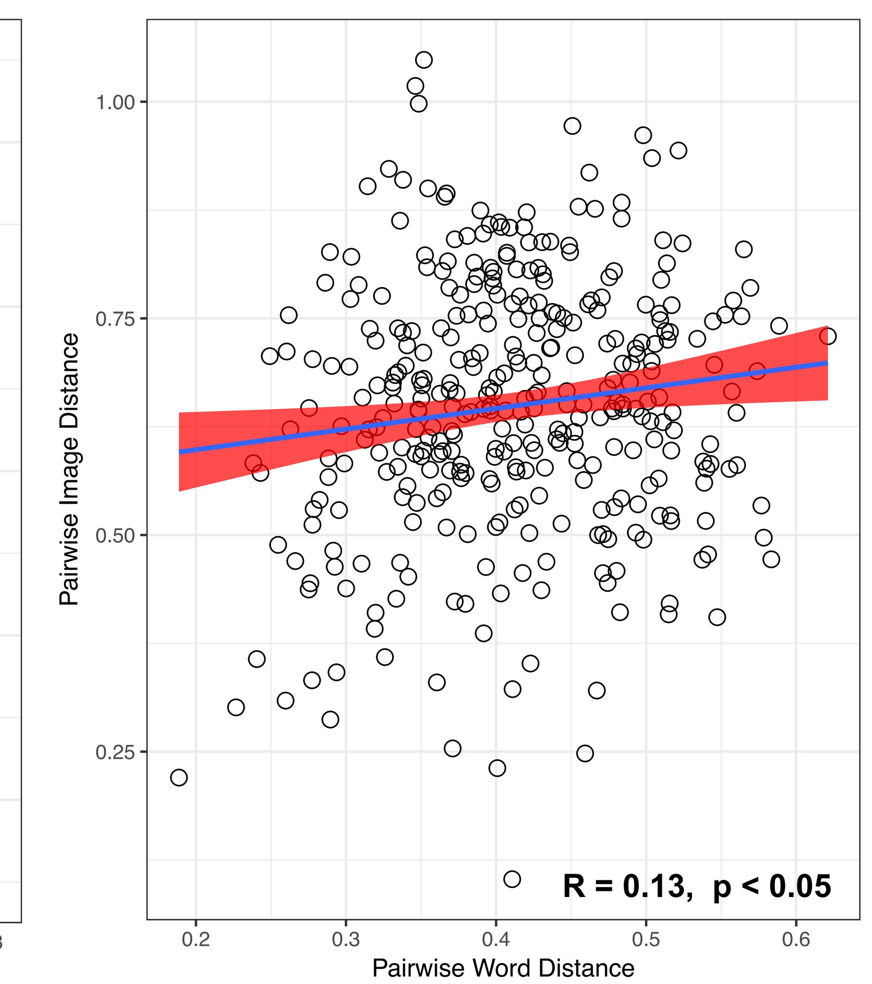


Results

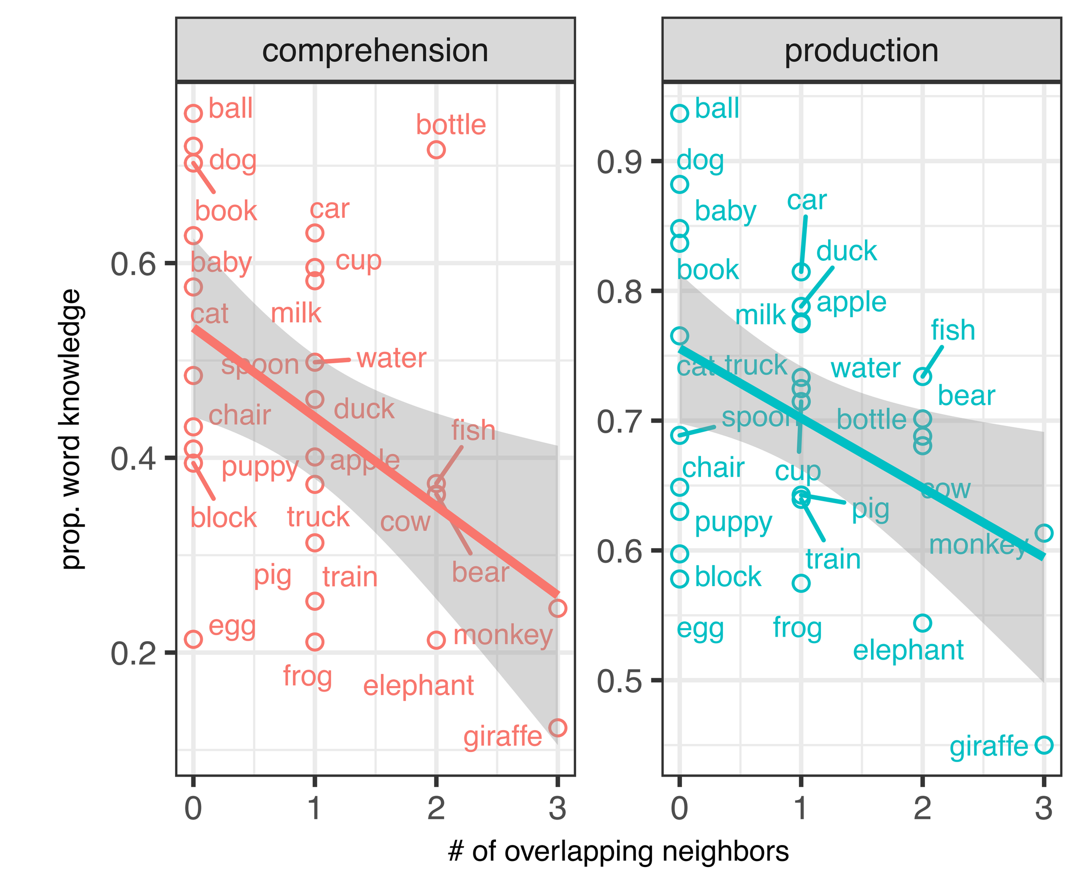
Common Crawl
42 billion token web crawl corpus



CHILDES
10.5 million token child directed speech corpus



- Pairwise distances between items in word-space and image-space are correlated.
- Correlation stronger in Common Crawl



- Objects with greater numbers of **overlapping** neighbors across *both* image and word space are later learned.
- This effect did <u>not</u> hold for # of neighbors in either space separately

word	ratio	neighbors
apple	0.17	**egg**, *puppy*, <u>milk, car, book, bear</u>
baby	0	*elephant, frog, cow, pig, train*, <u>puppy, dog, monkey, bear</u>
ball	0	*fish, frog, spoon, cow*, <u>dog, monkey</u>
bear	0.22	**monkey, elephant**, *puppy, cow*, <u>frog, baby, duck, dog, cat</u>
block	0	*egg, fish, train, duck, cow, puppy*, <u>water, ball, truck</u>
book	0	*fish, monkey, pig, puppy*, <u>car, baby, dog, train, cat</u>
bottle	0.33	**water, milk**, *egg*, <u>spoon, baby, cup</u>
car	0.5	**truck**, <u>train</u>
cat	0	*elephant, train, truck*, <u>dog, puppy, monkey</u>
chair	0	*frog, spoon, giraffe*, <u>baby, ball, dog, cat</u>
cow	0.22	**elephant, monkey**, *frog, egg, fish, giraffe*, <u>pig, dog, milk</u>
cup	0.14	**egg**, *duck, fish*, <u>spoon, milk, water, bottle</u>
dog	0	*cow, frog*, <u>puppy, cat</u>
duck	0.14	**fish**, *giraffe*, <u>cow, pig, egg, frog, elephant</u>
egg	0	*elephant, train*, <u>cow, duck, fish, milk, spoon</u>
elephant	0.25	**cow, giraffe**, *train, fish, egg*, <u>monkey, frog, bear</u>
fish	0.25	**duck, pig**, *cow, monkey, giraffe*, <u>frog, egg, water</u>
frog	0.1	**monkey**, *cow, giraffe, baby, puppy*, <u>elephant, fish, pig, duck, bear</u>
giraffe	0.43	**monkey, frog, cow**, *puppy, fish*, <u>elephant, baby</u>
milk	0.2	**water**, <u>bottle, egg, cow, cup</u>
monkey	0.38	**cow, frog, pig**, *giraffe, puppy, fish*, <u>elephant, cat</u>
pig	0.13	**monkey**, *puppy, fish, train*, <u>cow, duck, frog, dog</u>
puppy	0	*pig, monkey, cow, bear, giraffe*, <u>dog, baby, cat</u>
spoon	0	*frog, ball, giraffe*, <u>egg, cup, fish, milk, bottle</u>
train	0.25	**truck**, *elephant*, <u>car, dog</u>
truck	0.5	**train**, <u>car</u>
water	0.25	**milk**, *egg*, <u>bottle, fish</u>

- **overlap neighbors**, *image neighbors*, <u>word neighbors</u>
- overlap ratio (overlap/total neighbors) > 0 ($p<.05$)

Operationalization

Reducing items (words & images) to a common representational format
Two **vector space** models:

Words (Concepts) - GloVe word embeddings [4]
- Each word converted to a 300-dimensional vector = **word-vector** for that item

Images (Percepts) - ImageNet V3 CNN final layer activations [5]
- Each image converted to a 1024-dimensional vector, then the most **prototypical vector** chosen as the **image-vector** for that item

Defining a prototypical image vector
As opposed to word vectors, where every word-form is fixed and corresponds to a single vector, there is wide variance in images of any given category.

Choose the *most central* among a distribution: $\hat{x}_c = \underset{x \in U}{\arg\min} \sum_{y \in U} d(x, y)$

Subject to a *similarity/distance metric*: $d(x, y) = 1 - \frac{x \cdot y}{\|x\| \|y\|}$

Compute similarity or "distance" between 2 images or 2 words using the same $d(x, y)$ above.

Examining structure:
within a space
- compute the pairwise distances between all items in word- or image-space
- inter-object distances determine global structuring within the vector space

across spaces
- correlate pairwise distances between image and word space
- correlation implies ***conserved global structure across spaces*** learned by 2 unrelated algorithms

Relating to learning in humans
- Define a **neighbor** in word or image space as any item whose distance z-score < -1
- Test whether images or words with more neighbors are later-learned
- Measure degree of overlap in image- and word-space neighbors (**overlapping neighbors)**
- Link to comprehension/production norms on WordBank [3]
  - **Comprehension norms**: avg. percent of 8- to 18m.o. who **understand** each word
  - **Production norms:** average percent of 16 to 30m.o who **produce** each word

Discussion & Conclusions

1) Is cross-category structure preserved across visual & linguistic representations?
- **Yes**, pairwise distances between visual and semantic feature vectors are correlated ($R = 0.30$, $p < 10^{-7}$):
  - suggests that there is ***preserved structure*** across these two modality domains
    - similarity between items in one space predicts similarity in the other. <u>Could easily have been otherwise!</u>

2) If so, does this influence infant learning?
- **Yes**, overlap in visual and semantic feature spaces is linked to words' learnability ($R < -0.45$, $p < 0.05$)
  - i.e. the more cluttered an **item**'s neighborhood, the harder that item is to learn.

- Conceptual and sensory input are not independent or purely separable.
- Invariance relations across representational domains might be a useful cue for categorization/generalization/segmentation.
- Multi-dimensional clutter degrades learning.
- Vector space methods are a promising tool to model representational structure, without relying on (fraught) human judgments.

Ongoing & future directions
- Train image models on *decontextualized* input images so that residuals from background context do not enter the learning signal.
- Explore a wider space of possible items.
- Explore whether different classes of item (e.g. animate vs inanimate) preserve their position across spaces better than others.

References

- [1] Bergelson, E. (2016a). Bergelson seedlings homebank corpus. doi: 10.21415/T5PK6D
- [2] Bergelson, E. (2016b). Seedlings corpus. Retrieved 2017-01-29, from https://nyu.databrary.org/volume/228
- [3] Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman,V. A. (2016). Wordbank: An open repository for developmental vocabulary data. Journal of child language .
- [4] Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Empirical methods in natural language processing* (pp. 1532–1543).
- [5] Russakovsky, O., *et al.* (2015). ImageNet Large Scale Visual Recognition Challenge. Intenat'l Journal of Computer Vision, 115(3), 211–252. http://doi.org/ 10.1007/s11263-015-0816-y
- [6] Quinn, P., & Eimas, P. (1997). A reexamination of the perceptual-to-conceptual shift in mental representations. Review of General Psychology, 1(3), 271.
- [7] Quinn, P., & Eimas, P. (2000). The emergence of category representations during infancy: Are separate perceptual and conceptual processes required? J. Cog. & Dev., 1(1), 55–61.
- [8] Quinn, P. C., Johnson, M. H., Mareschal, D., Rakison, D. H., & Younger, B. A. (2000). Understanding early categorization: One process or two? Infancy, 1(1), 111–122.
- [9] Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. Journal of Cognition and Development, 1(1), 3–36.
- [10] Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. Cognition, 23(3), 183–209.
- [11] Trauble, B., & Pauen, S. (2007). The role of functional information for infant categorization. Cognition, 105(2), 362– 379.
- [12] Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. Psychological Review, 112(2), 347.
- [13] Sloutsky, V. M. (2003). The role of similarity in the development of categorization. Trends in Cognitive Sciences, 7(6), 246–251.
- [14] Sloutsky, V. M., Lo, Y.-F., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels, similarity, and the development of inductive inference. Child Development, 72(6), 1695–1709.
- [15] Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature, 388(6640), 381.
- [16] Rosch, E., & Lloyd, B. B. (1978). Cognition and categorization.

For the full paper, see: ***<u>https://arxiv.org/abs/1802.00840</u>***