A

Kiki or Bouba?

A.1 Full set of images

This appendix presents the full set of images with visual shapes that were used in the experiments. Besides the original image pair from Köhler (1929, 1947) which was shown in Figure 4.1, we used four image pairs from Maurer et al. (2006), displayed in Figure A.1, four from Westbury (2005), displayed in Figure A.2, and 8 additional pairs we newly generated using a method inspired by the one described by Nielsen and Rendall (2013), displayed in Figure A.3. For each image pair, the Curved version is displayed on the left and the Jagged version on the right.

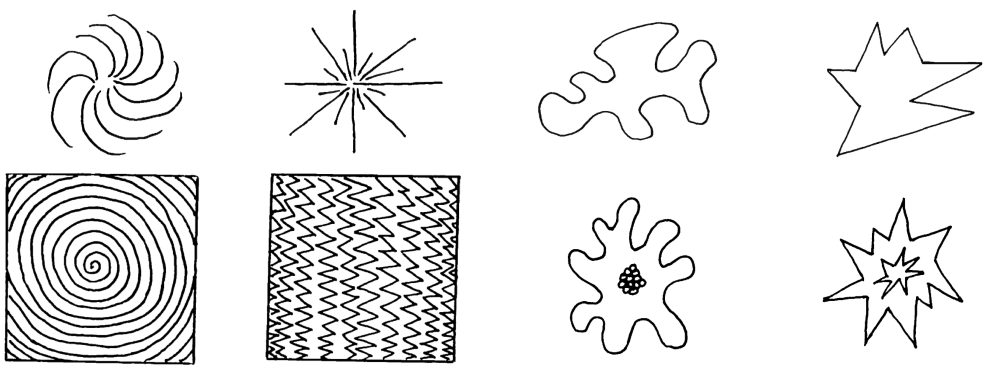


Figure A.1: Images from (Maurer et al., 2006)

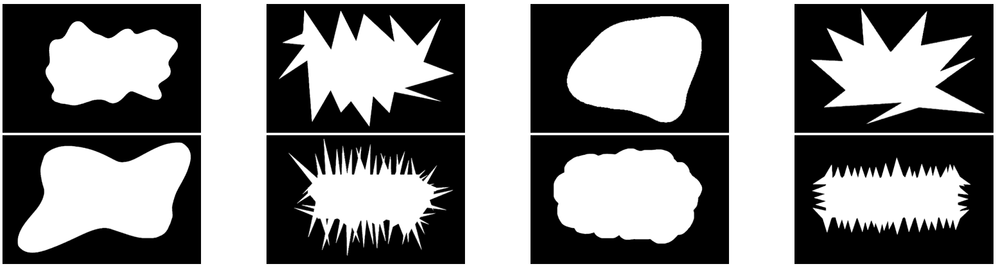


Figure A.2: Images from (Westbury, 2005)

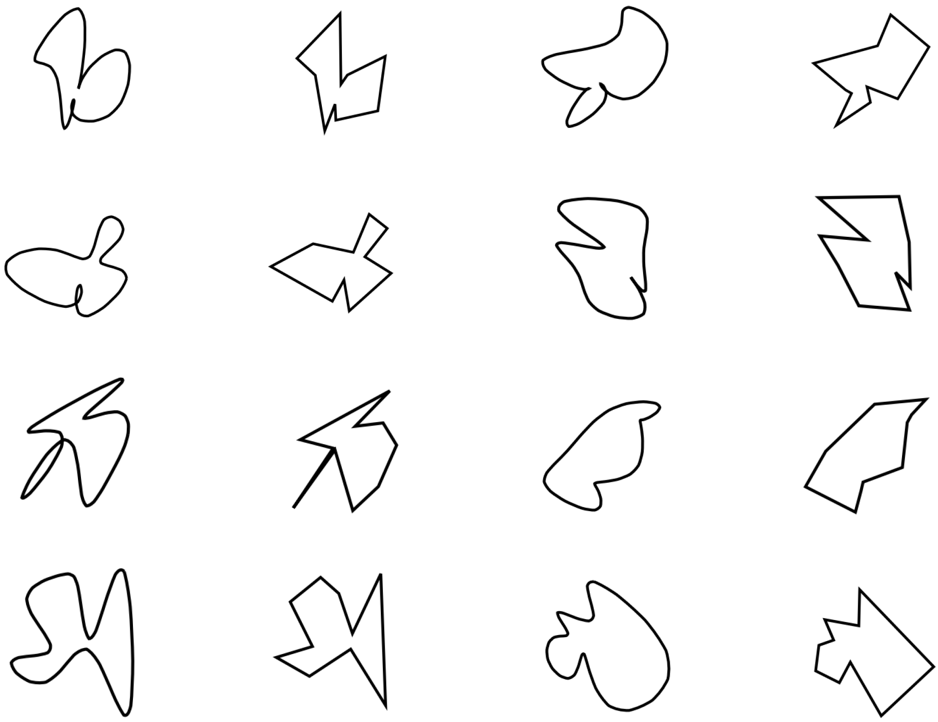


Figure A.3: Newly generated images

A.2 GPT-4o prompting

Image-label matching is not directly possible for GPT-4o since the probabilities of the input tokens cannot be accessed. We therefore prompt (Prompt A.2.1) this model, with the temperature being 0.0, to generate a syllable or pseudoword given an image and use the log probabilities of the generated tokens to calculate the probability for a label conditioned on an image. Just like in

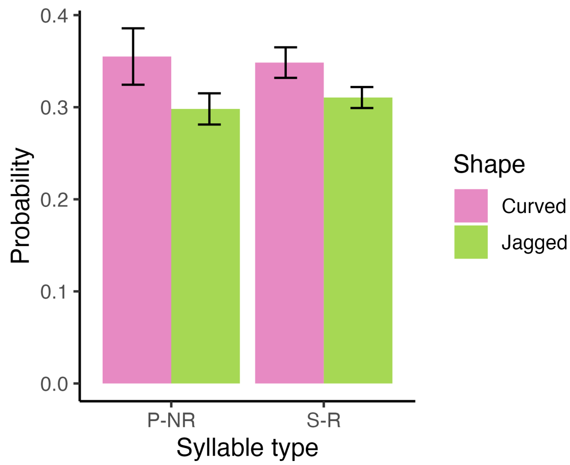


Figure A.4: Probability scores for GPT-4o when forced to generate a pseudoword for each image twice, once by combining two Jagged-associated syllables, and once with only the Curved-associated syllables as options.

the sentence setup used in the other models, our interest lies not primarily in the variability that may arise from using different prompts but rather focuses on the influence of the image on the predictions by using a simple and effective prompt that is identical for each image. Doing so allows us to use the resulting probabilities as a gauge for the models' preference of a label for a given image.

```
You are given an image for which you need to assign a label. Use {one
/two} of the following labels: {possible_labels}. Only respond with
the label.
```

Prompt A.2.1: The exact prompt used to obtain GPT-4o probabilities. $possible\_labels$ corresponds to the syllables of interest.

A.3 GPT-4o pseudoword probabilities

In Section 4.4.4 we describe the results of an experiment in which we asked GPT-4o to generate a pseudoword for each image twice, once when given only the set of Jagged-associated syllable options, and once with only the Curved-associated syllables as options. Figure A.4 shows the probabilities associated with these generated pseudowords. As concluded in the main text, no evidence for a preference to match P-NR syllables with Jagged shapes and S-R syllables with Curved shapes was found.

B

The Curious Case of Representational Alignment

B.1 Channel capacity

To test to what degree communicative success, *TopSim*, and representational alignment are confounded with the communication channel capacity, we ran simulations altering the vocabulary size ($V = \{3, 5, 10, 20, 40, 50, 100\}$) and message length ($L = \{2, 3, 5, 10, 50, 100\}$) resulting in 42 parameter settings per loss type. The parameters and seeds used to run the experiments in the main paper are displayed in Table B.1.

Overall, performance is relatively independent of the chosen configuration, but vocabulary size influences success more than message length (Figure B.1). The hyperparameters that resulted in the best validation accuracy (i.e., generalisation; Chaabouni et al., 2022) for the standard ce setup were $V = 40$ and $L = 2$. These parameters are used to produce the results

Parameter	**Value**
Batch size	32
Optimiser	Adam
Learning Rate (S & L)	0.01 & 0.001
Vocabulary size (V)	40
Message length (L)	2
Hidden size (S & L)	768 & 768
Embedding size	50
Listener cosine temperature	0.1
Seeds	16,22,41,56,67,77,14,78,99,23,82,40,51,37,62

Table B.1: Best-performing parameters resulting from the parameter sweep.

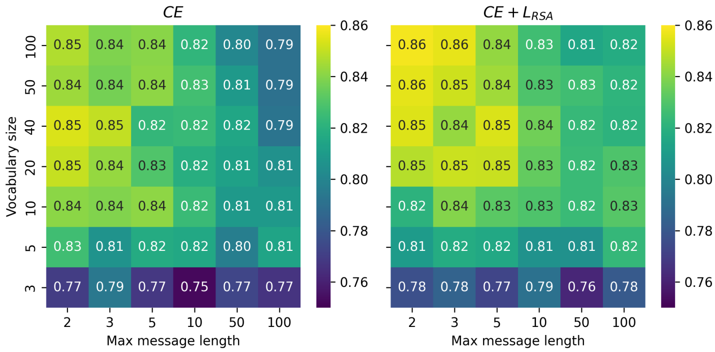


Figure B.1: The validation accuracy as a dependent factor of the vocabulary size and maximum message length. Values are averages across 15 seeds.

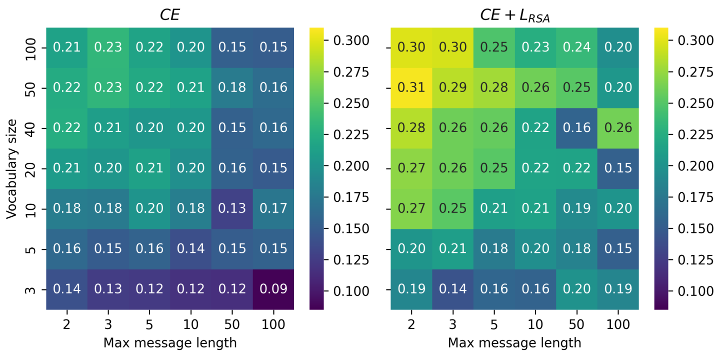


Figure B.2: *TopSim* as a dependent factor of the vocabulary size and maximum message length. Values are averages across 15 seeds.

in the main paper. Contra expectations, the vocabulary size also influenced *TopSim* more than message length. It, especially in the case of $ce + L_{RSA}$, is higher when messages are shorter but have access to a larger vocabulary (Figure B.2).

Figure B.3 shows that, regardless of the vocabulary capacity, inter-agent alignment (RSA_{sl}) increases while image-agent alignment (RSA_{si} and RSA_{li}) decreases with the ce loss. Interestingly, RSA_{sl} is agnostic to capacity but a larger vocabulary size, not message length, reduces the degree of drifting away from the input. We hypothesise this to result from lower pressure to compress rich continuous embeddings into smaller discrete vocabulary embeddings.

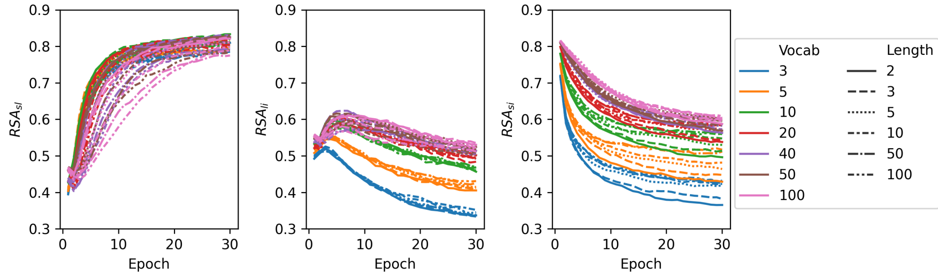


Figure B.3: Representational alignment metrics averaged over 15 simulations with the standard *ce* loss. Regardless of channel capacity, representational alignment always occurs while losing relation to the input.

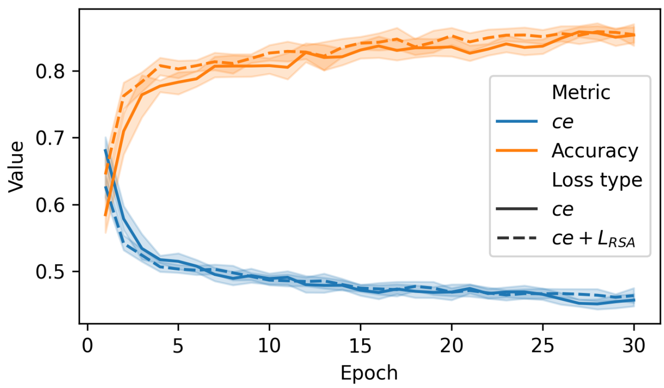


Figure B.4: Learning curves (accuracy) and cross-entropy loss (*ce*) for both loss settings. There is virtually no effect of the auxiliary term L_{RSA} on the cross entropy loss or communicative success.

B.2 Interaction between the alignment term and cross-entropy

To ensure that there is no impact of the alignment penalty on the pressure for communicative success, we ablated the L_{RSA} term of our proposed loss function and found that both, communicative success and *ce* are not affected by the alignment penalty (Figure B.4). Corroborating that only the *ce* term provides pressure for successful communication (Section 5.5.4).

B

B.3 Pre-trained vision modules

Although it is in principle possible to train the vision module of the agents from scratch (Dessi et al., 2021), in our work, agents' perception stems from a pre-trained vision-language model. Although there is reason to believe that DinoV2 embeddings capture high-level, conceptual image features useful for discriminating image pairs (Oquab et al., 2024), we assessed the degree to which the alignment problem occurs for different pre-trained models despite encoding the same objects. We ran additional simulations using image features obtained from ResNet (He et al., 2016) and CLIP (Radford et al., 2021) for 6 different parameter settings with the ce loss function. Here we used the parameters that resulted in the best, worst, mean, and quantile validation performance from the parameter sweep in Section B.1 (see Table B.2), and a sensible setup with $V = 10$ and $L = 5$.

Figure B.5 shows clearly that inter-agent alignment *increases* while agent-image alignment *decreases* for all models. In addition to the similar results reported by Bouchacourt and Baroni (2018) for VGG ConvNet embeddings, both 4096 and 1000 layers, our results confirm that the problem is agnostic to the input embeddings. Interestingly, agent representations drift most for CLIP embeddings. Nevertheless, the agents still develop a successful communication strategy, indicating that out-of-the-box CLIP embeddings are the least useful for agents in finding a (non-grounded) solution. No such differences are seen when the agents are trained with the additional alignment penalty term, inter-agent and image-agent alignment remain high for all models.

Message length (L)	Vocababulary Size (V)	Vision
2	40	
3	10	DinoV2
5	5	CLIP
5	10	ResNet
10	3	
50	100	

Table B.2: The parameters for running additional simulations with CLIP and ResNet to assess the robustness of our results. Each combination was run for 15 different seeds. Note: results for the DinoV2 simulations are from the sweep.

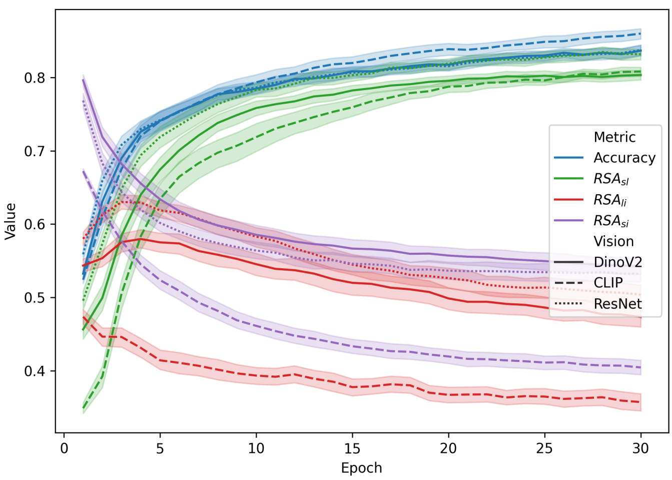


Figure B.5: Learning curves (accuracy) and *RSA* metrics for different vision models averaged over 6 parameter settings with 15 seeds each. The representational alignment problem always occurs. Line style corresponds to the vision module used to obtain image embeddings and colour indicates the metric. Areas indicate the 95% confidence intervals.

C
Searching for Structure

C.1 Additional results iterated learning

Figure C.1 shows how communicative success evolved across generations of language learners and users. There is no clear increase in communicative success. Figure C.2 shows that the average structure as measured by *TopSim* increases somewhat across generations, although not significantly. Interestingly, some generations display decreasing *TopSim*, indicating a loss of structure. This is reminiscent of findings in human iterated learning experiments, showing that processes of interaction and transmission sometimes generate fully systematic, compositional languages, but can also result in systems that lack structure entirely (Verhoef et al., 2022). In the case of *Ngram* diversity, we observe a decrease in the unique Ngrams produced, which indicates the languages re-use parts of signals more in later generations (Figure C.3).

C.2 Prompts

Our agents act based on prompts and system instructions. These are designed to be maximally close to the classical experimental setup and formatted similar to Galke et al. (2024). Prompt completion is used for labelling stimuli during the labelling and communication block. For the guessing task, we prefill the prompt with each possible word or distractor and pick the option with the highest probability. See the full prompts for labelling and guessing in Prompt C.2.1. Speaking during communication involved plain prompt completion (Prompt C.2.2). Discrimination during communication was done by prefilling the distractors attributes (Prompt C.2.3).

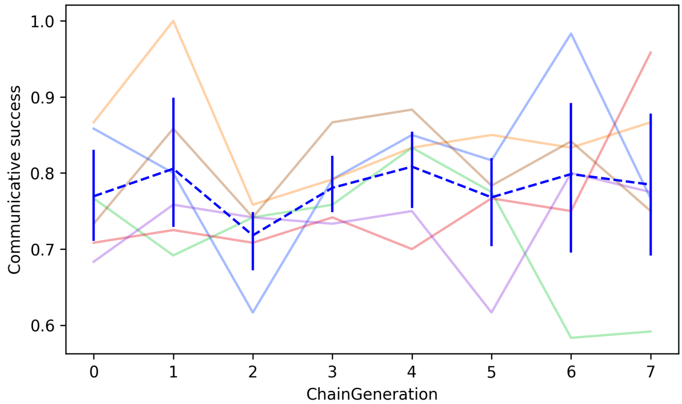


Figure C.1: The average communicative success across rounds for each generation. Each line represents a chain, and the dashed blue line indicates the average, with bars denoting the 95% confidence interval. See Table 6.2 for the descriptives of *PercCom*.

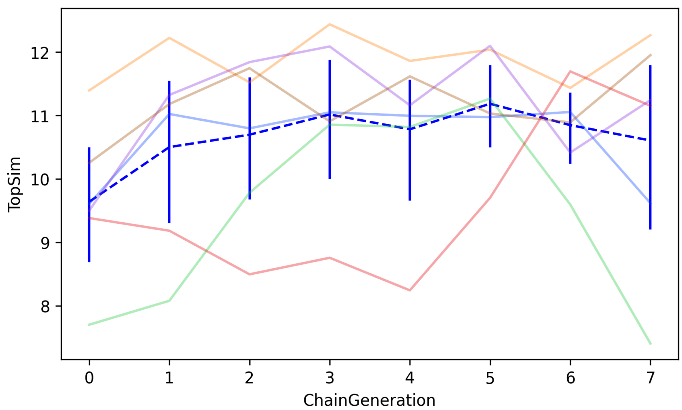


Figure C.2: The evolution of *TopSim* on the words produced in the testing block. Each line represents a chain, and the dashed blue line indicates the average, with bars denoting the 95% confidence interval. See Table 6.2 for the descriptives of *TopSim*.

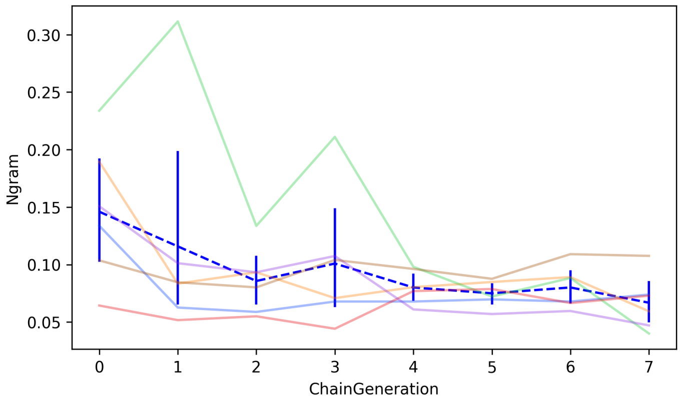


Figure C.3: The evolution of *Ngram* on the words produced in the testing block. Each line represents a chain, and the dashed blue line indicates the average, with bars denoting the 95% confidence interval. See Table 6.2 for the descriptives of *Ngram*.

```
<|begin_of_text|><|start_header_id|>system<|end_header_id|> You are a
 language learner who has to learn an artificial language with words
and their corresponding features. Your task is to complete the
vocabulary by generating a word that describes the last item. Only
respond with the word.<|eot_id|><|start_header_id|>user<|
end_header_id|>

{’shape’:2,’colour’:’orange’,’amount’:1,’word’:’giniwite’}
{’shape’:3,’colour’:’green’,’amount’:1,’word’:’ginisu’}
{’shape’:1,’colour’:’orange’,’amount’:2,’word’:’pinisugi’}
{’shape’:3,’colour’:’green’,’amount’:3,’word’:’sutepi’}
{’shape’:2,’colour’:’orange’,’amount’:2,’word’:’winisu’}
{’shape’:3,’colour’:’orange’,’amount’:1,’word’:’niwi’}
{’shape’:1,’colour’:’blue’,’amount’:2,’word’:’sutuwite’}
{’shape’:1,’colour’:’blue’,’amount’:3,’word’:’tupitene’}
{’shape’:3,’colour’:’blue’,’amount’:1,’word’:’wipinepi’}
{’shape’:2,’colour’:’orange’,’amount’:3,’word’:’gigi’}
{’shape’:1,’colour’:’green’,’amount’:2,’word’:’nite’}
{’shape’:3,’colour’:’blue’,’amount’:3,’word’:’wite’}
{’shape’:1,’colour’:’green’,’amount’:3,’word’:’sune’}
{’shape’:2,’colour’:’blue’,’amount’:2,’word’:’ninene’}
{’shape’:2,’colour’:’green’,’amount’:1,’word’:’tusetetu’}
{’shape’:1,’colour’:’green’,’amount’:3,’word’:’<|eot_id|><|
start_header_id|>assistant<|end_header_id|>
[COMPLETION OR PREFFILED]
```

Prompt C.2.1: Completion Prompt used for labelling and guessing.

```
<|begin_of_text|><|start_header_id|>system<|end_header_id|> You are a
 language learner who has to learn an artificial language with words
and their corresponding features. Your task is to generate a word
such that your communication partner can guess the correct meaning of
 the word. Communicative success is important. Only respond with the
word.<|eot_id|><|start_header_id|>user<|end_header_id|>

{’shape’:1,’colour’:’green’,’amount’:3,’word’:’sutupitite’,’
communicativeSuccess’:1}
{’shape’:2,’colour’:’orange’,’amount’:2,’word’:’ginupepi’,’
communicativeSuccess’:1}
{’shape’:1,’colour’:’orange’,’amount’:2,’word’:’sutupepi’,’
communicativeSuccess’:1}
{’shape’:1,’colour’:’green’,’amount’:2,’word’:’sutupepi’,’
communicativeSuccess’:0}
{’shape’:2,’colour’:’orange’,’amount’:1,’word’:’ginisu’,’
communicativeSuccess’:1}
{’shape’:2,’colour’:’orange’,’amount’:3,’word’:’ginupitite’,
’communicativeSuccess’:1}
{’shape’:3,’colour’:’green’,’amount’:1,’word’:’wipisu’,’
communicativeSuccess’:0}
{’shape’:2,’colour’:’green’,’amount’:1,’word’:’ginisu’,’
communicativeSuccess’:1}
{’shape’:1,’colour’:’blue’,’amount’:2,’word’:’sunupepi’,’
communicativeSuccess’:1}
{’shape’:3,’colour’:’green’,’amount’:3,’word’:’wipipitite’,’
communicativeSuccess’:1}
{’shape’:3,’colour’:’orange’,’amount’:1,’word’:’wipisu’,’
communicativeSuccess’:0}
{’shape’:1,’colour’:’blue’,’amount’:3,’word’:’sunupitite’,’
communicativeSuccess’:1}
{’shape’:3,’colour’:’blue’,’amount’:3,’word’:’wipipitite’,’
communicativeSuccess’:1}
{’shape’:3,’colour’:’blue’,’amount’:1,’word’:’wipisu’,’
communicativeSuccess’:1}
{’shape’:2,’colour’:’blue’,’amount’:2,’word’:’<|eot_id|><|
start_header_id|>assistant<|end_header_id|>
[COMPLETION]
```

Prompt C.2.2: Speaking Prompt during communication.

```
<|begin_of_text|><|start_header_id|>system<|end_header_id|> You are a
 language learner who has to learn an artificial language with words
and their corresponding features. Your task is to complete the
vocabulary by interpreting the intended meaning of the word generated
 by your communication partner. Communicative success is important.
Only respond with the complete last item.<|eot_id|><|start_header_id
|>user<|end_header_id|>

{’word’:’wipipitite’,’shape’:3,’colour’:’blue’,’amount’:3,’
communicativeSuccess’:1}
{’word’:’wipisu’,’shape’:3,’colour’:’orange’,’amount’:1,’
communicativeSuccess’:0}
{’word’:’wipisu’,’shape’:3,’colour’:’green’,’amount’:1,’
communicativeSuccess’:0}
{’word’:’sutupepi’,’shape’:1,’colour’:’orange’,’amount’:2,’
communicativeSuccess’:1}
{’word’:’ginupepi’,’shape’:2,’colour’:’orange’,’amount’:2,’
communicativeSuccess’:1}
{’word’:’sutupitite’,’shape’:1,’colour’:’green’,’amount’:3,’
communicativeSuccess’:1}
{’word’:’wipipitite’,’shape’:3,’colour’:’green’,’amount’:3,’
communicativeSuccess’:1}
{’word’:’wipisu’,’shape’:3,’colour’:’blue’,’amount’:1,’
communicativeSuccess’:1}
{’word’:’ginisu’,’shape’:2,’colour’:’green’,’amount’:1,’
communicativeSuccess’:1}
{’word’:’ginisu’,’shape’:2,’colour’:’orange’,’amount’:1,’
communicativeSuccess’:1}
{’word’:’sunupepi’,’shape’:1,’colour’:’blue’,’amount’:2,’
communicativeSuccess’:1}
{’word’:’sutupepi’,’shape’:1,’colour’:’green’,’amount’:2,’
communicativeSuccess’:0}
{’word’:’sunupitite’,’shape’:1,’colour’:’blue’,’amount’:3,’
communicativeSuccess’:1}
{’word’:’ginupitite’,’shape’:2,’colour’:’orange’,’amount’:3,
’communicativeSuccess’:1}
{’word’:’ginupepi’,’shape’:’<|eot_id|><|start_header_id|>assistant<|
end_header_id|>
[PREFILLED WITH DISTRACTOR ATTRIBUTES]
```

Prompt C.2.3: Guessing Prompt during communication.

D
Shaping Shared Languages

D.1 Prompts

The agents in our experiment act based on prompts and system instructions which are identical to those used in Chapter 6. These were designed to be maximally close to the classical experimental setup and formatted similar to Galke et al. (2024). During the labelling and guessing block, we use the completion Prompt D.1.1. In the labelling block, we simply ask the model to provide a completion. In the case of the guessing block, we prefill the word and pick the signal with the highest probability. See the full prompts for labelling and guessing (Prompt D.1.1), speaking (Prompt D.1.2), and discrimination (Prompt D.1.3) below.

As explained in the main body of our paper, we update the agent-specific vocabulary after each label prediction. This allows the vocabularies of signal-meaning mappings to evolve during the simulation. This entails that the prompts are also slightly different after each interaction or prediction. Moreover, given the observed bias for primacy and recency Liu et al. (2024) in LLMs, we shuffle the vocabulary before creating prompts to account for unwanted ordering effects.

D

```
<|begin_of_text|><|start_header_id|>system<|end_header_id|> You are a
 language learner who has to learn an artificial language with words
and their corresponding features. Your task is to complete the
vocabulary by generating a word that describes the last item. Only
respond with the word.<|eot_id|><|start_header_id|>user<|
end_header_id|>

{'shape':3,'colour':'green','amount':1,'word':'tego'}
{'shape':1,'colour':'green','amount':3,'word':'wananima'}
{'shape':2,'colour':'blue','amount':3,'word':'wumawaka'}
{'shape':2,'colour':'green','amount':3,'word':'mafa'}
{'shape':3,'colour':'orange','amount':2,'word':'wawa'}
{'shape':1,'colour':'orange','amount':1,'word':'gofa'}
{'shape':1,'colour':'blue','amount':1,'word':'maka'}
{'shape':3,'colour':'blue','amount':1,'word':'kama'}
{'shape':3,'colour':'blue','amount':3,'word':'mawa'}
{'shape':2,'colour':'orange','amount':2,'word':'nawa'}
{'shape':2,'colour':'blue','amount':1,'word':'kaka'}
{'shape':3,'colour':'green','amount':2,'word':'matefama'}
{'shape':1,'colour':'orange','amount':3,'word':'kagonigo'}
{'shape':2,'colour':'green','amount':2,'word':'nimaniwu'}
{'shape':1,'colour':'orange','amount':2,'word':'wago'}
{'shape':1,'colour':'orange','amount':2,'word':'<|eot_id|><|
start_header_id|>assistant<|end_header_id|>
[COMPLETION OR PREFFILED]
```

Prompt D.1.1: An example completion prompt used for labelling and guessing.

```
<|begin_of_text|><|start_header_id|>system<|end_header_id|> You are a
 language learner who has to learn an artificial language with words
and their corresponding features. Your task is to generate a word
such that your communication partner can guess the correct meaning of
 the word. Communicative success is important. Only respond with the
word.<|eot_id|><|start_header_id|>user<|end_header_id|>

{’shape’:1,’colour’:’green’,’amount’:3,’word’:’gofamama’,’
communicativeSuccess’:1}
{’shape’:2,’colour’:’orange’,’amount’:2,’word’:’kakafa’,’
communicativeSuccess’:1}
{’shape’:1,’colour’:’orange’,’amount’:2,’word’:’gofama’,’
communicativeSuccess’:1}
{’shape’:2,’colour’:’blue’,’amount’:1,’word’:’kaka’,’
communicativeSuccess’:1}
{’shape’:2,’colour’:’green’,’amount’:2,’word’:’kakafa’,’
communicativeSuccess’:1}
{’shape’:3,’colour’:’green’,’amount’:2,’word’:’tegoma’,’
communicativeSuccess’:1}
{’shape’:1,’colour’:’orange’,’amount’:1,’word’:’go’,’
communicativeSuccess’:1}
{’shape’:3,’colour’:’green’,’amount’:1,’word’:’tega’,’
communicativeSuccess’:1}
{’shape’:2,’colour’:’blue’,’amount’:3,’word’:’kakamama’,’
communicativeSuccess’:1}
{’shape’:3,’colour’:’blue’,’amount’:1,’word’:’tego’,’
communicativeSuccess’:1}
{’shape’:1,’colour’:’orange’,’amount’:3,’word’:’gofamama’,’
communicativeSuccess’:1}
{’shape’:2,’colour’:’green’,’amount’:3,’word’:’kakamama’,’
communicativeSuccess’:1}
{’shape’:3,’colour’:’orange’,’amount’:2,’word’:’tegoma’,’
communicativeSuccess’:0}
{’shape’:3,’colour’:’blue’,’amount’:3,’word’:’tegomama’,’
communicativeSuccess’:1}
{’shape’:1,’colour’:’blue’,’amount’:1,’word’:’<|eot_id|><|
start_header_id|>assistant<|end_header_id|>
[COMPLETION]
```

Prompt D.1.2: An example speaking prompt during communication. In this particular case, the speaker produced the label *’goa’* which was correctly interpreted by the human listener.

```
<|begin_of_text|><|start_header_id|>system<|end_header_id|> You are a
 language learner who has to learn an artificial language with words
and their corresponding features. Your task is to complete the
vocabulary by interpreting the intended meaning of the word generated
 by your communication partner. Communicative success is important.
Only respond with the complete last item.<|eot_id|><|start_header_id
|>user<|end_header_id|>

{'word':'kakafa','shape':2,'colour':'orange','amount':2,'
communicativeSuccess':1}
{'word':'go','shape':1,'colour':'orange','amount':1,'
communicativeSuccess':1}
{'word':'kakafa','shape':2,'colour':'green','amount':2,'
communicativeSuccess':1}
{'word':'goa','shape':1,'colour':'blue','amount':1,'
communicativeSuccess':1}
{'word':'kakamama','shape':2,'colour':'green','amount':3,'
communicativeSuccess':1}
{'word':'tego','shape':3,'colour':'blue','amount':1,'
communicativeSuccess':1}
{'word':'kaka','shape':2,'colour':'blue','amount':1,'
communicativeSuccess':1}
{'word':'tegoma','shape':3,'colour':'orange','amount':2,'
communicativeSuccess':0}
{'word':'gofamama','shape':1,'colour':'green','amount':3,'
communicativeSuccess':1}
{'word':'kakamama','shape':2,'colour':'blue','amount':3,'
communicativeSuccess':1}
{'word':'gofama','shape':1,'colour':'orange','amount':2,'
communicativeSuccess':1}
{'word':'tega','shape':3,'colour':'green','amount':1,'
communicativeSuccess':1}
{'word':'tegomama','shape':3,'colour':'blue','amount':3,'
communicativeSuccess':1}
{'word':'gofamama','shape':1,'colour':'orange','amount':3,'
communicativeSuccess':1}
{'word':'tegama','shape':'<|eot_id|><|start_header_id|>assistant<|
end_header_id|>
[PREFILLED WITH DISTRACTOR ATTRIBUTES]
```

Prompt D.1.3: An example guessing prompt during communication. Here the human speaker has produced the label *'tegama'* which was correctly interpreted by the listener.

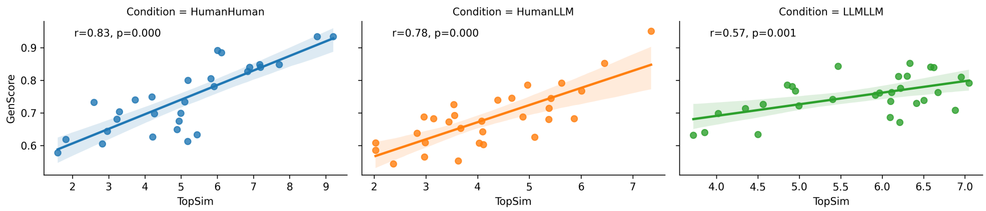


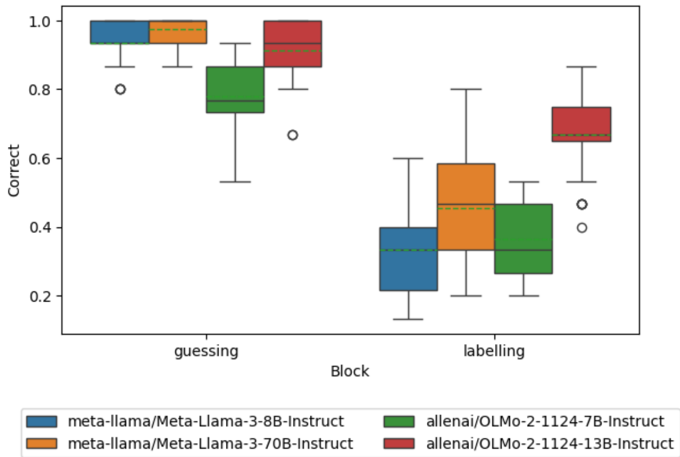
Figure D.1: Across conditions, generalisation to novel stimuli was more consistent with known samples when the labels produced during the last round of communication showed a higher degree of *TopSim*.

D.2 Generalisation to novel stimuli

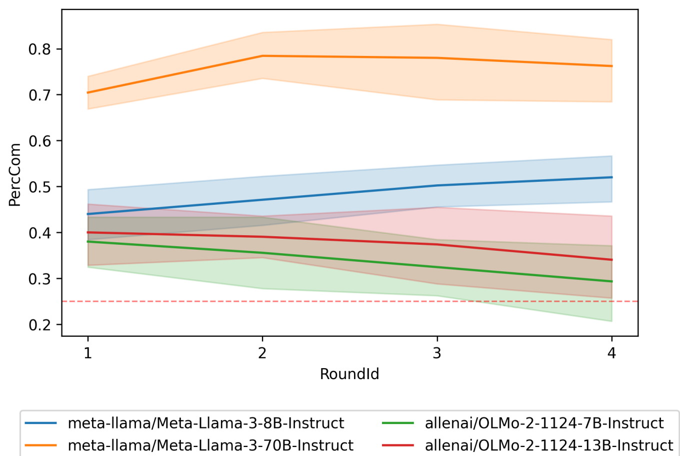
In the testing block of our experiments, we examined whether the participants and LLMs could generalise to novel stimuli. Figure D.1 shows the relationship between *TopSim* and *GenScore* that was modelled by a mixed effects model in Chapter 7. It reveals clearly that generalisation to novel items happened more consistently when the vocabulary evolved to be more structured.

D.3 Testing additional LLMs

To extend the findings of Chapter 6, we ran additional simulations with three different LLMs, Llama-3-8B (Llama Team, 2024), OLMo-2 7B, and OLMo-2 13B (Walsh et al., 2025) using the same 15 seeds. Figure D.2a shows that learning the artificial languages is also possible for smaller and different LLMs. Interestingly, out of all models, agents comprised of a OLMo-2 13B model perform best during the labelling task. While agents with OLMo-2 13B can also communicate reliably above change performance ($t(13) = 1.96, p = .036$), they struggle much more (*PercCom* $\approx 35\%$, Figure D.2b, chance performance amounts to 25%). Moreover, we observe that compared to their larger versions, smaller models struggle more. Llama-3 8B achieves $\approx 50\%$ of successful communication and OLMo-2 8B only 30%. The latter is not significantly above chance ($t(14) = 1.04, p = .158$). We take these results as additional evidence that our setup can be used to discover LLM-specific constraints and indicating that larger LLMs benefit more from instruction following Lou et al. (2024). We leave the precise dynamics of both to future work.



(a) Performance on the guessing and labelling task. The newly tested LLMs can learn the languages equally well or better (e.g. OLMO-2 13B outperforms Llama-3 70B).



(b) Communicative performance across rounds for different models. Note that the dashed red line indicates chance performance.

Figure D.2: The results of learning and using languages for two different LLMs with two different sizes. Figure D.2a shows the degree to which LLMs can learn the languages and Figure D.2b shows how well these models can use the language during communication.

Bibliography

Abramova, E. and Fernández, R. (2016). Questioning arbitrariness in language: a data-driven study of conventional iconicity. In Knight, K., Nenkova, A., and Rambow, O., editors, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 343–352, San Diego, California. Association for Computational Linguistics.

Acerbi, A. and Stubbersfield, J. M. (2023). Large language models show human-like content biases in transmission chain experiments. *Proceedings of the National Academy of Sciences*, 120(44):e2313790120.

Agrawal, A., Batra, D., Parikh, D., and Kembhavi, A. (2018). Don't just assume; look and answer: Overcoming priors for visual question answering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4971–4980.

Aher, G. V., Arriaga, R. I., and Kalai, A. T. (2023). Using large language models to simulate multiple humans and replicate human subject studies. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J., editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 337–371. PMLR.

Allen, K., Dasgupta, I., Kosoy, E., and Lampinen, A. K. (2025). The in-context inductive biases of vision-language models differ across modalities.

Alper, M. and Averbuch-Elor, H. (2023). Kiki or bouba? sound symbolism in vision-and-language models. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S., editors, *Advances in Neural Information Processing Systems*, volume 36, pages 78347–78359. Curran Associates, Inc.

Alper, M., Fiman, M., and Averbuch-Elor, H. (2023). Is BERT blind? exploring the effect of vision-and-language pretraining on visual language understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6778–6788.

Arnon, I. and Kirby, S. (2024). Cultural evolution creates the statistical structure of language. *Scientific Reports*, 14(1):5255.

Auersperger, M. and Pecina, P. (2022). Defending compositionality in emergent languages. In Ippolito, D., Li, L. H., Pacheco, M. L., Chen, D., and Xue, N., editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop*, pages 285–291, Hybrid: Seattle, Washington + Online. Association for Computational Linguistics.

Baker, N., Lu, H., Erlikhman, G., and Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLOS Computational Biology*, 14(12):1–43.

Ben Zion, R., Carmeli, B., Paradise, O., and Belinkov, Y. (2024). Semantics and spatiality of emergent communication. In Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., and Zhang, C., editors, *Advances in Neural Information Processing Systems*, volume 37, pages 110156–110196. Curran Associates, Inc.

Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 610–623, New York, NY, USA. Association for Computing Machinery.

Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, 80(2):290–311.

Beuls, K. and Van Eecke, P. (2024). Humans learn language from situated communicative interactions. what about machines? *Computational Linguistics*, 50(3):1277–1311.

Biderman, S., Schoelkopf, H., Anthony, Q. G., Bradley, H., O'Brien, K., Hallahan, E., Khan, M. A., Purohit, S., Prashanth, U. S., Raff, E., Skowron, A., Sutawika, L., and Van Der Wal, O. (2023). Pythia: A suite for analyzing large language models across training and scaling. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J., editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 2397–2430. PMLR.

Binz, M., Akata, E., Bethge, M., Brändle, F., Callaway, F., Coda-Forno, J., Dayan, P., Demircan, C., Eckstein, M. K., Éltető, N., Griffiths, T. L., Haridi, S., Jagadish, A. K., Ji-An, L., Kipnis, A., Kumar, S., Ludwig, T., Mathony, M., Mattar, M., Modirshanechi, A., Nath, S. S., Peterson, J. C., Rmus, M., Russek, E. M., Saanum, T., Scharfenberg, N., Schubert, J. A., Buschoff, L. M. S., Singhi, N., Sui, X., Thalmann, M., Theis, F., Truong, V., Udandarao, V., Voudouris, K., Wilson, R., Witte, K., Wu, S., Wulff, D., Xiong, H., and Schulz, E. (2025). A foundation model to predict and capture human cognition. *Nature*.

Binz, M. and Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, 120(6):e2218523120.

Binz, M. and Schulz, E. (2024). Turning large language models into cognitive models. In *The Twelfth International Conference on Learning Representations*.

Bisazza, A., Üstün, A., and Sportel, S. (2021). On the difficulty of translating free-order case-marking languages. *Transactions of the Association for Computational Linguistics*, 9:1233–1248.

Bisk, Y., Holtzman, A., Thomason, J., Andreas, J., Bengio, Y., Chai, J., Lapata, M., Lazaridou, A., May, J., Nisnevich, A., Pinto, N., and Turian, J. (2020). Experience grounds language. In Webber, B., Cohn, T., He, Y., and Liu, Y., editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8718–8735, Online. Association for Computational Linguistics.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., and Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, 113(39):10818–10823.

Blondel, M., Teboul, O., Berthet, Q., and Djolonga, J. (2020). Fast differentiable sorting and ranking. In *International Conference on Machine Learning*, pages 950–959. PMLR.

Bottini, R., Barilari, M., and Collignon, O. (2019). Sound symbolism in sighted and blind. the role of vision and orthography in sound-shape correspondences. *Cognition*, 185:62–70.

Bouchacourt, D. and Baroni, M. (2018). How agents see things: On visual representations in an emergent language game. In Riloff, E., Chiang, D., Hockenmaier, J., and Tsujii, J., editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 981–985, Brussels, Belgium. Association for Computational Linguistics.

Boyd, R., Richerson, P. J., et al. (1996). Why culture is common, but cultural evolution is rare. In *Proceedings-british academy*, volume 88, pages 77–94. Oxford University Press Inc.

Brandizzi, N. (2023). Toward more human-like ai communication: A review of emergent communication research. *IEEE Access*, 11:142317–142340.

Brandizzi, N. and Iocchi, L. (2022). Emergent communication in human-machine games. In *Emergent Communication Workshop at ICLR 2022*.

Brighton, H. and Kirby, S. (2006). Understanding linguistic evolution by visualizing the emergence of topographic mappings. *Artificial Life*, 12(2):229–242.

Brinkmann, L., Baumann, F., Bonnefon, J.-F., Derex, M., Müller, T. F., Nussberger, A.-M., Czaplicka, A., Acerbi, A., Griffiths, T. L., Henrich, J., et al. (2023). Machine culture. *Nature Human Behaviour*, 7(11):1855–1868.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, 100(5):1–54.

Carr, J. W., Smith, K., Cornish, H., and Kirby, S. (2017). The cultural evolution of structured languages in an open-ended, continuous world. *Cognitive Science*, 41(4):892–923.

Chaabouni, R., Kharitonov, E., Bouchacourt, D., Dupoux, E., and Baroni, M. (2020). Compositionality and generalization in emergent languages. In Jurafsky, D., Chai, J., Schluter, N., and Tetreault, J., editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4427–4442, Online. Association for Computational Linguistics.

Chaabouni, R., Kharitonov, E., Dupoux, E., and Baroni, M. (2019a). Anti-efficient encoding in emergent communication. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Chaabouni, R., Kharitonov, E., Lazaric, A., Dupoux, E., and Baroni, M. (2019b). Word-order biases in deep-agent emergent communication. In Korhonen, A., Traum, D., and Màrquez, L., editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5166–5175, Florence, Italy. Association for Computational Linguistics.

Chaabouni, R., Strub, F., Altché, F., Tarassov, E., Tallec, C., Davoodi, E., Mathewson, K. W., Tieleman, O., Lazaridou, A., and Piot, B. (2022). Emergent communication at scale. In *International Conference on Learning Representations*.

Chang, T. A. and Bergen, B. K. (2022). Word acquisition in neural language models. *Transactions of the Association for Computational Linguistics*, 10:1–16.

Chen, A., Shwartz-Ziv, R., Cho, K., Leavitt, M. L., and Saphra, N. (2024). Sudden drops in the loss: Syntax acquisition, phase transitions, and simplicity bias in MLMs. In *The Twelfth International Conference on Learning Representations*.

Cheng, E., Rita, M., and Poibeau, T. (2023). On the correspondence between compositionality and imitation in emergent neural communication. In Rogers, A., Boyd-Graber, J., and Okazaki, N., editors, *Findings of the Association for Computational Linguistics: ACL 2023*, pages 12432–12447, Toronto, Canada. Association for Computational Linguistics.

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Moschitti, A., Pang, B., and Daelemans, W., editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

Chomsky, N. (2023). The false promise of chatgpt. *The New York Times*.

Christiansen, M. and Chater, N. (2022). *The Language Game: How Improvisation Created Language and Changed the World*. Basic Books.

Christiansen, M. H. and Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31(5):489–509.

Christiansen, M. H. and Kirby, S. (2003). Language evolution: The hardest problem in science? In *Language Evolution*. Oxford University Press.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204.

Clark, H. and Brennan, S. (1991). *Grounding in Communication*, volume 13, pages 127–149. American Psychological Association.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Clark, K., Khandelwal, U., Levy, O., and Manning, C. D. (2019). What does BERT look at? an analysis of BERT's attention. In Linzen, T., Chrupała, G., Belinkov, Y., and Hupkes, D., editors, *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 276–286, Florence, Italy. Association for Computational Linguistics.

Cohen, R. G. and Rosenbaum, D. A. (2004). Where grasps are made reveals how grasps are planned: generation and recall of motor plans. *Experimental Brain Research*, 157(4):486–495.

Conklin, H. and Smith, K. (2023). Compositionality with variation reliably emerges in neural networks. In *The Eleventh International Conference on Learning Representations*.

Contreras Kallens, P. and Christiansen, M. H. (2024). Distributional semantics: Meaning through culture and interaction. *Topics in Cognitive Science*, n/a(n/a).

Contreras Kallens, P., Kristensen-McLachlan, R. D., and Christiansen, M. H. (2023). Large language models demonstrate the potential of statistical learning in language. *Cognitive Science*, 47(3):e13256.

Conwell, C., Prince, J. S., Kay, K. N., Alvarez, G. A., and Konkle, T. (2023). What can 1.8 billion regressions tell us about the pressures shaping high-level visual representation in brains and machines? *bioRxiv*.

Culbertson, J. and Smolensky, P. (2012). A bayesian model of biases in artificial language learning: The case of a word-order universal. *Cognitive Science*, 36(8):1468–1498.

Cuskley, C., Dingemanse, M., Kirby, S., and Van Leeuwen, T. M. (2019). Cross-modal associations and synesthesia: Categorical perception and structure in vowel–color mappings in a large online sample. *Behavior research methods*, 51:1651–1675.

Cuskley, C. and Kirby, S. (2013). Synesthesia, Cross-Modality, and Language Evolution. In *Oxford Handbook of Synesthesia*. Oxford University Press.

Cuskley, C., Simner, J., and Kirby, S. (2017). Phonological and orthographic influences in the bouba–kiki effect. *Psychological research*, 81:119–130.

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyan, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á., Ünal-Logacev, Ö., Zeller, J., Perlman, M., and Winter, B. (2022). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1841):20200390.

Dagan, G., Hupkes, D., and Bruni, E. (2021). Co-evolution of language and agents in referential games. In Merlo, P., Tiedemann, J., and Tsarfaty, R., editors, *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2993–3004, Online. Association for Computational Linguistics.

Darcet, T., Oquab, M., Mairal, J., and Bojanowski, P. (2024). Vision transformers need registers. In *The Twelfth International Conference on Learning Representations*.

Davis, C. P., Morrow, H. M., and Lupyan, G. (2019). What does a horgous look like? nonsense words elicit meaningful drawings. *Cognitive Science*, 43(10):e12791.

De Boer, B. (2000). Self-organization in vowel systems. *Journal of phonetics*, 28(4):441–465.

de Boer, B. (2006). *Computer modelling as a tool for understanding language evolution*, pages 381–406. Springer Netherlands, Dordrecht.

De Kleijn, R., Kachergis, G., and Hommel, B. (2014). Everyday robotic action: lessons from human action control. *Frontiers in Neurorobotics*, 8:13.

de Kleijn, R., Kachergis, G., and Hommel, B. (2018). Predictive movements and human reinforcement learning of sequential action. *Cognitive Science*, 42(S3):783–808.

de Kleijn, R., Sen, D., and Kachergis, G. (2022). A critical period for robust curriculum-based deep reinforcement learning of sequential action in a robot arm. *Topics in Cognitive Science*, 14(2):311–326.

Deacon, T. W. (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*. ISSR Library. W.W. Norton.

DeCaro, M. S., Thomas, R. D., and Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition*, 107(1):284–294.

Dessi, R., Kharitonov, E., and Marco, B. (2021). Interpretable agent communication from scratch (with a generic visual processor emerging on the side). In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 26937–26949. Curran Associates, Inc.

Dillion, D., Tandon, N., Gu, Y., and Gray, K. (2023). Can ai language models replace human participants? *Trends in Cognitive Sciences*, 27(7):597–600.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, 6(10):654–672.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., and Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19(10):603–615.

Diwan, A., Berry, L., Choi, E., Harwath, D., and Mahowald, K. (2022). Why is winoground hard? investigating failures in visuolinguistic compositionality. In Goldberg, Y., Kozareva, Z., and Zhang, Y., editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 2236–2250, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Dubova, M. and Moskvichev, A. (2020). Effects of supervision, population size, and self-play on multi-agent reinforcement learning to communicate. In *The 2020 Conference on Artificial Life*, volume ALIFE 2020: The 2020 Conference on Artificial Life of *Artificial Life Conference Proceedings*, pages 678–686.

Eva, S., Silvia, H., and Dáša, M. (2014). Personal need for structure in relation to language variables. *Procedia - Social and Behavioral Sciences*, 159:665–670. 5th World Conference on Psychology, Counseling and Guidance, WCPCG-2014, 1-3 May 2014, Dubrovnik, Croatia.

Evans, K. K. and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, 10(1):6–6.

Evans, N. and Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32(5):429–448.

Fay, N., Garrod, S., and Roberts, L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1509):3553–3561.

Fedzechkina, M., Jaeger, T. F., and Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, 109(44):17897–17902.

Fort, M. and Schwartz, J.-L. (2022). Resolving the bouba-kiki effect enigma by rooting iconic sound symbolism in physical properties of round and spiky objects. *Scientific reports*, 12(1):19172.

Futrell, R. and Mahowald, K. (2025). How linguistics learned to stop worrying and love the language models.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, 29(5):737–767.

Galantucci, B. and Garrod, S. (2010). Experimental semiotics: A new approach for studying the emergence and the evolution of human communication. *Interaction Studies*, 11(1):1–13.

Galke, L., Ram, Y., and Raviv, L. (2022). Emergent communication for understanding human language evolution: What's missing? In *Emergent Communication Workshop at ICLR 2022*.

Galke, L., Ram, Y., and Raviv, L. (2024). Deep neural networks and humans both benefit from compositional language structure. *Nature Communications*, 15(1):10816.

Galke, L. and Raviv, L. (2024). Emergent communication and learning pressures in language models: a language evolution perspective. *arXiv preprint arXiv:2403.14427*.

Galke, L. and Raviv, L. (2025). Learning and communication pressures in neural networks: Lessons from emergent communication. *Language Development Research*, 5(1):116–143.

Garrod, S., Fay, N., Lee, J., Oberlander, J., and MacLeod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, 31(6):961–987.

Geirhos, R., Narayanappa, K., Mitzkus, B., Thieringer, T., Bethge, M., Wichmann, F. A., and Brendel, W. (2021). Partial success in closing the gap between human and machine vision. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 23885–23899. Curran Associates, Inc.

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., and Brendel, W. (2019). Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *International Conference on Learning Representations*.

Gemini Team, G. (2024). Gemini: A family of highly capable multimodal models.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., and Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, 23(5):389–407.

Giulianelli, M., Malagutti, L., Gastaldi, J. L., DuSell, B., Vieira, T., and Cotterell, R. (2024). On the proper treatment of tokenization in psycholinguistics. In Al-Onaizan, Y., Bansal, M., and Chen, Y.-N., editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 18556–18572, Miami, Florida, USA. Association for Computational Linguistics.

Goyal, Y., Khot, T., Summers-Stay, D., Batra, D., and Parikh, D. (2017). Making the v in vqa matter: Elevating the role of image understanding in visual question answering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Griffiths, T. L. and Kalish, M. L. (2007a). Language evolution by iterated learning with bayesian agents. *Cognitive Science*, 31(3):441–480.

Griffiths, T. L. and Kalish, M. L. (2007b). Language evolution by iterated learning with bayesian agents. *Cognitive Science*, 31(3):441–480.

Griffiths, T. L., Zhu, J.-Q., Grant, E., and McCoy, R. T. (2024). Bayes in the age of intelligent machines. *Current Directions in Psychological Science*, 33(5):283–291.

Guo, S., Ren, Y., Havrylov, S., Frank, S., Titov, I., and Smith, K. (2019). The emergence of compositional languages for numeric concepts through iterated learning in neural agents.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3):335–346.

Havrylov, S. and Titov, I. (2017). Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30, page 2146–2156. Curran Associates, Inc.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., and Steinhardt, J. (2021). Measuring massive multitask language understanding. In *International Conference on Learning Representations*.

Hermann, K., Chen, T., and Kornblith, S. (2020). The origins and prevalence of texture bias in convolutional neural networks. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 19000–19015. Curran Associates, Inc.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.*, 9(8):1735–1780.

Hockett, C. F. (1959). Animal "languages" and human language. *Human Biology*, 31(1):32–39.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203(3):88–97.

Holtzman, A., Buys, J., Du, L., Forbes, M., and Choi, Y. (2020). The curious case of neural text degeneration. In *International Conference on Learning Representations*.

Hsieh, C.-Y., Zhang, J., Ma, Z., Kembhavi, A., and Krishna, R. (2023). Sugarcrepe: Fixing hackable benchmarks for vision-language compositionality. In Oh, A., Neumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S., editors, *Advances in Neural Information Processing Systems*, volume 36, pages 31096–31116. Curran Associates, Inc.

Hu, J. and Frank, M. (2024). Auxiliary task demands mask the capabilities of smaller language models. In *First Conference on Language Modeling*.

Hu, J. and Levy, R. (2023). Prompting is not a substitute for probability measurements in large language models. In Bouamor, H., Pino, J., and Bali, K., editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5040–5060, Singapore. Association for Computational Linguistics.

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American Journal of Psychology*, 109(2):219–238.

Hurford, J. (2007). *The Origins of Meaning: Language in the Light of Evolution*. James R. Hurford. Oxford University Press.

Iida, H. and Funakura, H. (2024). Investigating iconicity in vision-and-language models: A case study of the bouba/kiki effect in japanese models. In *Proceedings of the 46th Annual Conference of the Cognitive Science Society*, volume 46.

Imai, M. and Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical transactions of the Royal Society B: Biological sciences*, 369(1651):20130298.

Imai, M., Kita, S., Nagumo, M., and Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109(1):54–65.

Jabri, A., Joulin, A., and van der Maaten, L. (2016). Revisiting visual question answering baselines. In Leibe, B., Matas, J., Sebe, N., and Welling, M., editors, *Computer Vision – ECCV 2016*, pages 727–739, Cham. Springer International Publishing.

Jepma, M., Murphy, P. R., Nassar, M. R., Rangel-Gomez, M., Meeter, M., and Nieuwenhuis, S. (2016). Catecholaminergic regulation of learning rate in a dynamic environment. *PLOS Computational Biology*, 12(10):1–24.

Jones, C. R., Bergen, B., and Trott, S. (2024). Do multimodal large language models and humans ground language similarly? *Computational Linguistics*, 50(3):1415–1440.

Josserand, M., Pellegrino, F., Grosseck, O., Dediu, D., and Raviv, L. (2024). Adapting to individual differences: An experimental study of language evolution in heterogeneous populations. *Cognitive Science*, 48(11):e70011.

Juzek, T. S. and Ward, Z. B. (2025). Why does ChatGPT "delve" so much? exploring the sources of lexical overrepresentation in large language models. In Rambow, O., Wanner, L., Apidianaki, M., Al-Khalifa, H., Eugenio, B. D., and Schockaert, S., editors, *Proceedings of the 31st International Conference on Computational Linguistics*, pages 6397–6411, Abu Dhabi, UAE. Association for Computational Linguistics.

Kallini, J., Papadimitriou, I., Futrell, R., Mahowald, K., and Potts, C. (2024). Mission: Impossible language models. In Ku, L.-W., Martins, A., and Srikumar, V., editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14691–14714, Bangkok, Thailand. Association for Computational Linguistics.

Kamath, A., Hessel, J., and Chang, K.-W. (2023). What's "up" with vision-language models? investigating their struggle with spatial reasoning. In Bouamor, H., Pino, J., and Bali, K., editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 9161–9175, Singapore. Association for Computational Linguistics.

Karamcheti, S., Nair, S., Balakrishna, A., Liang, P., Kollar, T., and Sadigh, D. (2024). Prismatic VLMs: Investigating the design space of visually-conditioned language models. In *Forty-first International Conference on Machine Learning*.

Kemp, C. and Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, 336(6084):1049–1054.

Kharitonov, E., Chaabouni, R., Bouchacourt, D., and Baroni, M. (2020). Entropy minimization in emergent languages. In III, H. D. and Singh, A., editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 5220–5230. PMLR.

Kharitonov, E., Dessì, R., Chaabouni, R., Bouchacourt, D., and Baroni, M. (2021). EGG: a toolkit for research on Emergence of lanGuage in Games. `https://github.com/facebookresearch/EGG`.

Khatun, A. and Brown, D. G. (2024). A study on large language models' limitations in multiple-choice question answering.

Kim, W., Son, B., and Kim, I. (2021). Vilt: Vision-and-language transformer without convolution or region supervision. In Meila, M. and Zhang, T., editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 5583–5594. PMLR.

Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In Bengio, Y. and LeCun, Y., editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psychonomic bulletin & review*, 24(1):118–137.

Kirby, S. and Christiansen, M. H. (2003). From language learning to language evolution. In *Language Evolution*. Oxford University Press.

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31):10681–10686.

Kirby, S., Griffiths, T., and Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, 28:108–114.

Kirby, S. and Tamariz, M. (2022). Cumulative cultural evolution, population structure and the origin of combinatoriality in human language. *Philosophical Transactions of the Royal Society B*, 377(1843):20200319.

Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102.

Köhler, W. (1929). *Gestalt Psychology*. New York: Horace Liveright.

Köhler, W. (1947). *Gestalt Psychology*. (2nd ed.) New York: Horace Liveright.

Kosinski, M. (2024). Evaluating large language models in theory of mind tasks. *Proceedings of the National Academy of Sciences*, 121(45):e2405460121.

Kottur, S., Moura, J., Lee, S., and Batra, D. (2017). Natural language does not emerge 'naturally' in multi-agent dialog. In Palmer, M., Hwa, R., and Riedel, S., editors, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2962–2967, Copenhagen, Denmark. Association for Computational Linguistics.

Kouwenhoven, T., Shahrasbi, K., and Verhoef, T. (2025). Cross-modal associations in vision and language models: Revisiting the bouba-kiki effect.

Kozachkov, L., Kastanenka, K. V., and Krotov, D. (2023). Building transformers from neurons and astrocytes. *Proceedings of the National Academy of Sciences*, 120(34):e2219150120.

Kriegeskorte, N., Mur, M., and Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2:4.

Lashley, K. S. et al. (1951). *The problem of serial order in behavior*, volume 21. Bobbs-Merrill Oxford.

Lazaridou, A. and Baroni, M. (2020). Emergent multi-agent communication in the deep learning era.

Lazaridou, A., Hermann, K. M., Tuyls, K., and Clark, S. (2018). Emergence of linguistic communication from referential games with symbolic and pixel input. In *International Conference on Learning Representations*.

Lazaridou, A., Peysakhovich, A., and Baroni, M. (2017). Multi-agent cooperation and the emergence of (natural) language. In *International Conference on Learning Representations*.

Lazaridou, A., Pham, N. T., and Baroni, M. (2016). Towards multi-agent communication-based language learning. *arXiv preprint arXiv:1605.07133*.

Lazaridou, A., Potapenko, A., and Tieleman, O. (2020). Multi-agent communication meets natural language: Synergies between functional and structural language learning. In Jurafsky, D., Chai, J., Schluter, N., and Tetreault, J., editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7663–7674, Online. Association for Computational Linguistics.

Lee, H., Phatale, S., Mansoor, H., Mesnard, T., Ferret, J., Lu, K., Bishop, C., Hall, E., Carbune, V., Rastogi, A., and Prakash, S. (2024). Rlaif vs. rlhf: scaling reinforcement learning from human feedback with ai feedback. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org.

Lewis, D. (1969). Convention: A philosophical study. *Cambridge, MA*.

Li, F. and Bowling, M. (2019). Ease-of-teaching and language structure from emergent communication. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Li, H., Nourkhiz Mahjoub, H., Chalaki, B., Tadiparthi, V., Lee, K., Moradi Pari, E., Lewis, C., and Sycara, K. (2024). Language grounded multi-agent reinforcement learning with human-interpretable communication. In Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., and Zhang, C., editors, *Advances in Neural Information Processing Systems*, volume 37, pages 87908–87933. Curran Associates, Inc.

Li, J., Galley, M., Brockett, C., Gao, J., and Dolan, B. (2016). A diversity-promoting objective function for neural conversation models. In Knight, K., Nenkova, A., and Rambow, O., editors, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 110–119, San Diego, California. Association for Computational Linguistics.

Li, J., Li, D., Savarese, S., and Hoi, S. (2023). BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J., editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 19730–19742. PMLR.

Lian, Y., Bisazza, A., and Verhoef, T. (2021). The effect of efficient messaging and input variability on neural-agent iterated language learning. In Moens, M.-F., Huang, X., Specia, L., and Yih, S. W.-t., editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10121–10129, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Lian, Y., Bisazza, A., and Verhoef, T. (2023a). Communication Drives the Emergence of Language Universals in Neural Agents: Evidence from the Word-order/Case-marking Trade-off. *Transactions of the Association for Computational Linguistics*, 11:1033–1047.

Lian, Y., Bisazza, A., and Verhoef, T. (2023b). Communication drives the emergence of language universals in neural agents: Evidence from the word-order/case-marking trade-off. *Transactions of the Association for Computational Linguistics*, 11:1033–1047.

Lian, Y., Verhoef, T., and Bisazza, A. (2024). NeLLCom-X: A comprehensive neural-agent framework to simulate language learning and group communication. In Barak, L. and Alikhani, M., editors, *Proceedings of the 28th Conference on Computational Natural Language Learning*, pages 243–258, Miami, FL, USA. Association for Computational Linguistics.

Liljencrants, J. and Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48(4):839–862.

Lin, B. Y., Ravichander, A., Lu, X., Dziri, N., Sclar, M., Chandu, K., Bhagavatula, C., and Choi, Y. (2024). The unlocking spell on base LLMs: Rethinking alignment via in-context learning. In *The Twelfth International Conference on Learning Representations*.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *Computer Vision – ECCV 2014*, pages 740–755, Cham. Springer International Publishing.

Lindblom, B. and Maddieson, I. (1988). Phonetic universals in consonant systems. In Li, C. and Hyman, L. M., editors, *Language, Speech and Mind*, pages 62–78. Routledge, London.

Little, H., Eryılmaz, K., and de Boer, B. (2017). Signal dimensionality and the emergence of combinatorial structure. *Cognition*, 168:1–15.

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., and Liang, P. (2024). Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12:157–173.

Llama Team, M. (2024). The llama 3 herd of models.

Locke, J. (1847). *An essay concerning human understanding*, volume 114. Kay & Troutman.

Lockwood, G. and Dingemanse, M. (2015). Iconicity in the lab: a review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, 6.

Löhn, L., Kiehne, N., Ljapunov, A., and Balke, W.-T. (2024). Is machine psychology here? on requirements for using human psychological tests on large language models. In Mahamood, S., Minh, N. L., and Ippolito, D., editors, *Proceedings of the 17th International Natural Language Generation Conference*, pages 230–242, Tokyo, Japan. Association for Computational Linguistics.

Lou, R., Zhang, K., and Yin, W. (2024). Large language model instruction following: A survey of progresses and challenges. *Computational Linguistics*, 50(3):1053–1095.

Lowe, R., Gupta, A., Foerster, J., Kiela, D., and Pineau, J. (2020). On the interaction between supervision and self-play in emergent communication. In *International Conference on Learning Representations*.

Lupyan, G. and Dale, R. (2016). Why are there different languages? the role of adaptation in linguistic diversity. *Trends in Cognitive Sciences*, 20(9):649–660.

Mahaut, M., Dessi, R., Franzon, F., and Baroni, M. (2025). Referential communication in heterogeneous communities of pre-trained visual deep networks. *Transactions on Machine Learning Research*.

Manning, C. D., Clark, K., Hewitt, J., Khandelwal, U., and Levy, O. (2020). Emergent linguistic structure in artificial neural networks trained by self-supervision. *Proceedings of the National Academy of Sciences*, 117(48):30046–30054.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2_Part_1):209–220.

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *The American Journal of Psychology*, 87(1/2):173–188.

Maurer, D., Pathman, T., and Mondloch, C. J. (2006). The shape of boubas: Sound–shape correspondences in toddlers and adults. *Developmental science*, 9(3):316–322.

McCrae, R. R., Costa, Jr, P. T., and Martin, T. A. (2005). The NEO–PI–3: A more readable revised NEO personality inventory. *Journal of Personality Assessment*, 84(3):261–270.

Meister, C., Pimentel, T., Wiher, G., and Cotterell, R. (2023). Locally typical sampling. *Transactions of the Association for Computational Linguistics*, 11:102–121.

Meister, C., Wiher, G., Pimentel, T., and Cotterell, R. (2022). On the probability–quality paradox in language generation. In Muresan, S., Nakov, P., and Villavicencio, A., editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 36–45, Dublin, Ireland. Association for Computational Linguistics.

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance*, 15(1):69.

Meteyard, L., Stoppard, E., Snudden, D., Cappa, S. F., and Vigliocco, G. (2015). When semantics aids phonology: A processing advantage for iconic word forms in aphasia. *Neuropsychologia*, 76:264–275. Special Issue: Semantic Cognition.

Michel, P., Rita, M., Mathewson, K. W., Tieleman, O., and Lazaridou, A. (2023). Revisiting populations in multi-agent communication. In *The Eleventh International Conference on Learning Representations*.

Mikolov, T., Joulin, A., and Baroni, M. (2018). A roadmap towards machine intelligence. In *Computational Linguistics and Intelligent Text Processing*, pages 29–61, Cham.

Millière, R. and Rathkopf, C. (2024). Anthropocentric bias and the possibility of artificial cognition. In *ICML 2024 Workshop on LLMs and Cognition*.

Millière, R. (2024). Language models as models of language.

Mina, M., Ruiz-Fernández, V., Falcão, J., Vasquez-Reina, L., and Gonzalez-Agirre, A. (2025). Cognitive biases, task complexity, and result intepretability in large language models. In Rambow, O., Wanner, L., Apidianaki, M., Al-Khalifa, H., Eugenio, B. D., and Schockaert, S., editors, *Proceedings of the 31st International Conference on Computational Linguistics*, pages 1767–1784, Abu Dhabi, UAE. Association for Computational Linguistics.

Mishra, S., Khashabi, D., Baral, C., Choi, Y., and Hajishirzi, H. (2022). Reframing instructional prompts to GPTk's language. In Muresan, S., Nakov, P., and Villavicencio, A., editors, *Findings of the Association for Computational Linguistics: ACL 2022*, pages 589–612, Dublin, Ireland. Association for Computational Linguistics.

Misra, K. and Mahowald, K. (2024). Language models learn rare phenomena from less rare phenomena: The case of the missing AANNs. In Al-Onaizan, Y., Bansal, M., and Chen, Y.-N., editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 913–929, Miami, Florida, USA. Association for Computational Linguistics.

Mitchell, M. and Krakauer, D. C. (2023). The debate over understanding in ai's large language models. *Proceedings of the National Academy of Sciences*, 120(13):e2215907120.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., and Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In Balcan, M. F. and Weinberger, K. Q., editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1928–1937, New York, New York, USA. PMLR.

Mollo, D. C. and Millière, R. (2023). The vector grounding problem.

Montero, M. L., Ludwig, C. J., Costa, R. P., Malhotra, G., and Bowers, J. (2021). The role of disentanglement in generalisation. In *International Conference on Learning Representations*.

Mordatch, I. and Abbeel, P. (2018). Emergence of grounded compositional language in multi-agent populations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

Morey, R., Rouder, J., Jamil, T., Urbanek, S., Forner, K., and Ly, A. (2018). Bayesfactor: Computation of bayes factors for common designs (r package version 0.9. 12-4.2)[computer software]. *Retrieved form https://CRAN. R-project. org/package= BayesFactor*.

Mu, J. and Goodman, N. (2021). Emergent communication of generalizations. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 17994–18007. Curran Associates, Inc.

Muennighoff, N., Wang, T., Sutawika, L., Roberts, A., Biderman, S., Le Scao, T., Bari, M. S., Shen, S., Yong, Z. X., Schoelkopf, H., Tang, X., Radev, D., Aji, A. F., Almubarak, K., Albanie, S., Alyafeai, Z., Webson, A., Raff, E., and Raffel, C. (2023). Crosslingual generalization through multitask finetuning. In Rogers, A., Boyd-Graber, J., and Okazaki, N., editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15991–16111, Toronto, Canada. Association for Computational Linguistics.

Muttenthaler, L., Dippel, J., Linhardt, L., Vandermeulen, R. A., and Kornblith, S. (2023). Human alignment of neural network representations. In *The Eleventh International Conference on Learning Representations*.

Nakagawa, S. and Schielzeth, H. (2013). A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2):133–142.

Nastase, S. A., Goldstein, A., and Hasson, U. (2020). Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, 222:117254.

Neuberg, S. L. and Newsom, J. T. (1993). Personal need for structure: Individual differences in the desire for simpler structure. *Journal of Personality and Social Psychology*, 65(1):113.

Nielsen, A. and Rendall, D. (2011). The sound of round: evaluating the sound-symbolic role of consonants in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 65(2):115.

Nielsen, A. and Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition*, 4(2):115–125.

Nielsen, A. and Rendall, D. (2013). Parsing the role of consonants versus vowels in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 67(2):153.

Nikolaus, M. and Fourtassi, A. (2021). Evaluating the acquisition of semantic knowledge from cross-situational learning in artificial neural networks. In Chersoni, E., Hollenstein, N., Jacobs, C., Oseki, Y., Prévot, L., and Santus, E., editors, *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 200–210, Online. Association for Computational Linguistics.

Nussenbaum, K. and Hartley, C. A. (2019). Reinforcement learning across development: What insights can we draw from a decade of research? *Developmental Cognitive Neuroscience*, 40:100733.

Nölle, J., Staib, M., Fusaroli, R., and Tylén, K. (2018). The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition*, 181:93–104.

Oliphant, M. (2002). *Learned systems of arbitrary reference: The foundation of human linguistic uniqueness*. Cambridge University Press.

Onnis, L., Lim, A., Cheung, S., and Huettig, F. (2022). Is the mind inherently predicting? exploring forward and backward looking in language processing. *Cognitive Science*, 46(10):e13201.

OpenAI (2024). Gpt-4 technical report.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H. V., Szafraniec, M., Khalidov, V., Fernandez, P., HAZIZA, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.-Y., Li, S.-W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., and Bojanowski, P. (2024). DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*.

Ostrand, R. and Berger, S. E. (2024). Humans linguistically align to their conversational partners, and language models should too. In *ICML 2024 Workshop on LLMs and Cognition*.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P. F., Leike, J., and Lowe, R. (2022). Training language models to follow instructions with human feedback. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A., editors, *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744. Curran Associates, Inc.

Parise, C. V. and Spence, C. (2009). 'when birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLOS ONE*, 4(5):1–7.

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., and Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA. Association for Computing Machinery.

Peeperkorn, M., Kouwenhoven, T., Brown, D., and Jordanous, A. (2024). Is temperature the creativity parameter of large language models? In Grace, K., Llano, M. T., Martins, P., and Hedblom, M. M., editors, *Proceedings of the 15th International Conference on Computational Creativity*, pages 226–235. Association for Computational Creativity.

Peeperkorn, M., Kouwenhoven, T., Brown, D., and Jordanous, A. (2025). Mind the gap: Conformative decoding to improve output diversity of instruction-tuned large language models.

Pellert, M., Lechner, C. M., Wagner, C., Rammstedt, B., and Strohmaier, M. (2024). Ai psychometrics: Assessing the psychological profiles of large language models through psychometric inventories. *Perspectives on Psychological Science*, 19(5):808–826. PMID: 38165766.

Perez, J., Léger, C., Ovando-Tellez, M., Foulon, C., Dussauld, J., Oudeyer, P.-Y., and Moulin-Frier, C. (2024). Cultural evolution in populations of large language models.

Perlman, M., Dale, R., and Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, 2(8):150152.

Perniss, P., Thompson, R., and Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, 1.

Perry, L. K., Perlman, M., and Lupyan, G. (2015). Iconicity in english and spanish and its relation to lexical category and age of acquisition. *PLOS ONE*, 10(9):1–17.

Piantadosi, S. T. (2024). Modern language models refute chomsky's approach to language. In *From fieldwork to linguistic theory: A tribute to Dan Everett*, pages 353–414. Language Science Press.

Piantadosi, S. T., Tily, H., and Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9):3526–3529.

Pickering, M. J. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–190.

Pimentel, T., McCarthy, A. D., Blasi, D., Roark, B., and Cotterell, R. (2019). Meaning to form: Measuring systematicity as information. In Korhonen, A., Traum, D., and Màrquez, L., editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1751–1764, Florence, Italy. Association for Computational Linguistics.

Poletiek, F. H., Conway, C. M., Ellefson, M. R., Lai, J., Bocanegra, B. R., and Christiansen, M. H. (2018). Under what conditions can recursion be learned? effects of starting small in artificial grammar learning of center-embedded structure. *Cognitive Science*, 42(8):2855–2889.

Quinn, M. (2001). Evolving communication without dedicated communication channels. In Kelemen, J. and Sosík, P., editors, *Advances in Artificial Life*, pages 357–366, Berlin, Heidelberg. Springer Berlin Heidelberg.

Quinn, M., Smith, L., Mayley, G., and Husbands, P. (2003). Evolving controllers for a homogeneous system of physical robots: structured cooperation with minimal sensors. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 361(1811):2321–2343.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In Meila, M. and Zhang, T., editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR.

Ramachandran, V. S. and Hubbard, E. M. (2001). Synaesthesia–a window into perception, thought and language. *Journal of Consciousness Studies*, 8(12):3–34.

Raviv, L., de Heer Kloots, M., and Meyer, A. (2021). What makes a language easy to learn? a preregistered study on how systematic structure and community size affect language learnability. *Cognition*, 210:104620.

Raviv, L., Meyer, A., and Lev-Ari, S. (2019a). Compositional structure can emerge without generational transmission. *Cognition*, 182:151–164.

Raviv, L., Meyer, A., and Lev-Ari, S. (2019b). Larger communities create more systematic languages. *Proceedings of the Royal Society B: Biological Sciences*, 286(1907):20191262.

Ray, A., Radenovic, F., Dubey, A., Plummer, B., Krishna, R., and Saenko, K. (2023). Cola: A benchmark for compositional text-to-image retrieval. In Oh, A., Neumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S., editors, *Advances in Neural Information Processing Systems*, volume 36, pages 46433–46445. Curran Associates, Inc.

Ren, Y., Guo, S., Labeau, M., Cohen, S. B., and Kirby, S. (2020). Compositional languages emerge in a neural iterated learning model. In *International Conference on Learning Representations*.

Ren, Y., Guo, S., Qiu, L., Wang, B., and Sutherland, D. J. (2024). Bias amplification in language model evolution: An iterated learning perspective. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Ren, Y., Jin, R., Zhang, T., and Xiong, D. (2025). Do large language models mirror cognitive language processing? In Rambow, O., Wanner, L., Apidianaki, M., Al-Khalifa, H., Eugenio, B. D., and Schockaert, S., editors, *Proceedings of the 31st International Conference on Computational Linguistics*, pages 2988–3001, Abu Dhabi, UAE. Association for Computational Linguistics.

Rende, R., Gerace, F., Laio, A., and Goldt, S. (2024). A distributional simplicity bias in the learning dynamics of transformers. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Rita, M., Chaabouni, R., and Dupoux, E. (2020). "LazImpa": Lazy and impatient neural agents learn to communicate efficiently. In Fernández, R. and Linzen, T., editors, *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 335–343, Online. Association for Computational Linguistics.

Rita, M., Michel, P., Chaabouni, R., Pietquin, O., Dupoux, E., and Strub, F. (2024). Language evolution with deep learning. *arXiv preprint arXiv:2403.11958*.

Rita, M., Strub, F., Grill, J.-B., Pietquin, O., and Dupoux, E. (2022a). On the role of population heterogeneity in emergent communication. In *International Conference on Learning Representations*.

Rita, M., Tallec, C., Michel, P., Grill, J.-B., Pietquin, O., Dupoux, E., and Strub, F. (2022b). Emergent communication: Generalization and overfitting in lewis games. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A., editors, *Advances in Neural Information Processing Systems*, volume 35, pages 1389–1404. Curran Associates, Inc.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695.

Saito, K., Wachi, A., Wataoka, K., and Akimoto, Y. (2023). Verbosity bias in preference labeling by large language models. In *NeurIPS 2023 Workshop on Instruction Tuning and Instruction Following*.

Schaeffer, R., Miranda, B., and Koyejo, S. (2023). Are emergent abilities of large language models a mirage? In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S., editors, *Advances in Neural Information Processing Systems*, volume 36, pages 55565–55581. Curran Associates, Inc.

Schmidtke, D., Conrad, M., and Jacobs, A. M. (2014). Phonological iconicity. *Frontiers in Psychology*, 5.

Schulz, E. and Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Current Opinion in Neurobiology*, 55:7–14. Machine Learning, Big Data, and Neuroscience.

Schuster, M. and Paliwal, K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681.

Schwartz, R. and Stanovsky, G. (2022). On the limitations of dataset balancing: The lost battle against spurious correlations. In Carpuat, M., de Marneffe, M.-C., and Meza Ruiz, I. V., editors, *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 2182–2194, Seattle, United States. Association for Computational Linguistics.

Scott-Phillips, T. C. and Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, 14(9):411–417.

Scott-Phillips, T. C., Kirby, S., and Ritchie, G. R. (2009). Signalling signalhood and the emergence of communication. *Cognition*, 113(2):226–233.

Selten, R. and Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences*, 104(18):7361–7366.

Shapira, N., Levy, M., Alavi, S. H., Zhou, X., Choi, Y., Goldberg, Y., Sap, M., and Shwartz, V. (2024). Clever hans or neural theory of mind? stress testing social reasoning in large language models. In Graham, Y. and Purver, M., editors, *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2257–2273, St. Julian's, Malta. Association for Computational Linguistics.

Shiono, D., Brassard, A., Ishizuki, Y., and Suzuki, J. (2025). Evaluating model alignment with human perception: A study on shitsukan in LLMs and LVLMs. In Rambow, O., Wanner, L., Apidianaki, M., Al-Khalifa, H., Eugenio, B. D., and Schockaert, S., editors, *Proceedings of the 31st International Conference on Computational Linguistics*, pages 11428–11444, Abu Dhabi, UAE. Association for Computational Linguistics.

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., and Gal, Y. (2024). Ai models collapse when trained on recursively generated data. *Nature*, 631(8022):755–759.

Simner, J., Cuskley, C., and Kirby, S. (2010). What sound does that taste? cross-modal mappings across gustation and audition. *Perception*, 39(4):553–569. PMID: 20515002.

Skean, O., Arefin, M. R., Zhao, D., Patel, N., Naghiyev, J., LeCun, Y., and Shwartz-Ziv, R. (2025). Layer by layer: Uncovering hidden representations in language models.

Smith, K. (2011). Learning Bias, Cultural Evolution of Language, and theBiological Evolution of the Language Faculty. *Human Biology*, 83(2):261 – 278.

Smith, K. (2022). How language learning and language use create linguistic structure. *Current Directions in Psychological Science*, 31(2):177–186.

Smith, K. and Culbertson, J. (2020). Communicative pressures shape language during communication (not learning): Evidence from casemarking in artificial languages.

Smith, K., Kirby, S., Guo, S., and Griffiths, T. L. (2024). Ai model collapse might be prevented by studying human language transmission. *Nature*, 633(8030):525.

Smith, K., Tamariz, M., and Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In *Proceedings of the annual meeting of the cognitive science society*, volume 35.

Steels, L. (1999). *The Talking Heads Experiment.* Laboratorium, Antwerpen, Antwerpen.

Steels, L. (2006). Experiments on the emergence of human communication. *Trends in Cognitive Sciences*, 10(8):347–349.

Steels, L. (2012a). *Grounding Language through Evolutionary Language Games*, pages 1–22. Springer US, Boston, MA.

Steels, L. (2012b). Self-organization and selection in cultural language evolution. In *Experiments in Cultural Language Evolution*, pages 1–37. John Benjamins, Amsterdam.

Steels, L. and Loetzsch, M. (2012). The grounded naming game. In *Experiments in Cultural Language Evolution*, volume 3, pages 41–59. John Benjamins.

Steinert-Threlkeld, S., Zhou, X., Liu, Z., and Downey, C. M. (2022). Emergent communication fine-tuning (EC-FT) for pretrained language models. In *Emergent Communication Workshop at ICLR 2022*.

Steinmetz, J.-P., Loare, E., and Houssemand, C. (2011). Rigidity of attitudes and behaviors: A study on the validity of the concept. *Individual Differences Research*, 9(2):84 – 106.

Sucholutsky, I. and Griffiths, T. (2023). Alignment with human representations supports robust few-shot learning. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S., editors, *Advances in Neural Information Processing Systems*, volume 36, pages 73464–73479. Curran Associates, Inc.

Sucholutsky, I., Muttenthaler, L., Weller, A., Peng, A., Bobu, A., Kim, B., Love, B. C., Grant, E., Groen, I., Achterberg, J., Tenenbaum, J. B., Collins, K. M., Hermann, K. L., Oktar, K., Greff, K., Hebart, M. N., Jacoby, N., Zhang, Q., Marjieh, R., Geirhos, R., Chen, S., Kornblith, S., Rane, S., Konkle, T., O'Connell, T. P., Unterthiner, T., Lampinen, A. K., Müller, K.-R., Toneva, M., and Griffiths, T. L. (2023). Getting aligned on representational alignment.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. A Bradford Book, Cambridge, MA, USA.

Svecova, V. and Pavlovicova, G. (2016). Screening the personal need for the structure and solving word problems with fractions. *SpringerPlus*, 5(1):1–9.

Tamariz, M. and Kirby, S. (2015). Culture: Copying, compression, and conventionality. *Cognitive Science*, 39(1):171–183.

Tamariz, M., Roberts, S. G., Martínez, J. I., and Santiago, J. (2018). The interactive origin of iconicity. *Cognitive Science*, 42(1):334–349.

ter Hoeve, M., Kharitonov, E., Hupkes, D., and Dupoux, E. (2022). Towards interactive language modeling.

Theisen-White, C., Kirby, S., and Oberlander, J. (2011). Integrating the horizontal and vertical cultural transmission of novel communication systems. In *Proceedings of the annual meeting of the cognitive science society*, volume 33, pages 956–961, Boston, MA. Cognitive Science Society.

Thompson, M. M., Naccarato, M. E., and Parker, K. E. (1989). Assessing cognitive need: The development of the personal need for structure and personal fear of invalidity scales. In *annual meeting of the Canadian Psychological Association, Halifax, Nova Scotia, Canada*.

Thrush, T., Jiang, R., Bartolo, M., Singh, A., Williams, A., Kiela, D., and Ross, C. (2022). Winoground: Probing vision and language models for visio-linguistic compositionality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5238–5248.

Tieleman, O., Lazaridou, A., Mourad, S., Blundell, C., and Precup, D. (2019). Shaping representations through communication: community size effect in artificial learning systems. *arXiv preprint arXiv:1912.06208*.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press.

Tomasello, M. (2008). *Origins of human communication*. MIT Press, Cambridge, Mass.; London.

Turoman, N. and Styles, S. J. (2017). Glyph guessing for 'oo'and 'ee': Spatial frequency information in sound symbolic matching for ancient and unfamiliar scripts. *Royal Society open science*, 4(9):170882.

Tylén, K., Fusaroli, R., Bundgaard, P. F., and Østergaard, S. (2013). Making sense together: A dynamical account of linguistic meaning-making. *Semiotica*, 2013(194):39–62.

van Dijk, B., Kouwenhoven, T., Spruit, M., and van Duijn, M. J. (2023a). Large language models: The need for nuance in current debates and a pragmatic perspective on understanding. In Bouamor, H., Pino, J., and Bali, K., editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12641–12654, Singapore. Association for Computational Linguistics.

van Dijk, B., van Duijn, M., Verberne, S., and Spruit, M. (2023b). ChiSCor: A corpus of freely-told fantasy stories by Dutch children for computational linguistics and cognitive science. In Jiang, J., Reitter, D., and Deng, S., editors, *Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL)*, pages 352–363, Singapore. Association for Computational Linguistics.

van Dijk, B., van Duijn, M. J., Kloostra, L., Spruit, M., and Beekhuizen, B. (2024). Using a language model to unravel semantic development in children's use of a dutch perception verb. In Zock, M., Chersoni, E., Hsu, Y.-Y., and de Deyne, S., editors, *Proceedings of the Workshop on Cognitive Aspects of the Lexicon @ LREC-COLING 2024*, pages 98–106, Torino, Italia. ELRA and ICCL.

van Duijn, M., van Dijk, B., Kouwenhoven, T., de Valk, W., Spruit, M., and van der Putten, P. (2023). Theory of mind in large language models: Examining performance of 11 state-of-the-art models vs. children aged 7-10 on advanced tests. In Jiang, J., Reitter, D., and Deng, S., editors, *Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL)*, pages 389–402, Singapore. Association for Computational Linguistics.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and cognition*, 4(4):357–380.

Verhoef, T., de Boer, B., et al. (2011). Language acquisition age effects and their role in the preservation and change of communication systems. *Linguistics in Amsterdam*, 4(1).

Verhoef, T., Kirby, S., and De Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, 43:57–68.

Verhoef, T., Kirby, S., and De Boer, B. (2016a). Iconicity and the emergence of combinatorial structure in language. *Cognitive science*, 40(8):1969–1994.

Verhoef, T., Roberts, S. G., and Dingemanse, M. (2015). Emergence of systematic iconicity: Transmission, interaction and analogy. In *37th Annual Meeting of the Cognitive Science Society (CogSci 2015)*, pages 2481–2486. Cognitive Science Society.

Verhoef, T., Walker, E., and Marghetis, T. (2016b). Cognitive biases and social coordination in the emergence of temporal language. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, pages 2615–2620, Austin, TX. Cognitive Science Society.

Verhoef, T., Walker, E., and Marghetis, T. (2022). Interaction dynamics affect the emergence of compositional structure in cultural transmission of space-time mappings. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 44, pages 2133–2139.

Vital, F., Sardinha, A., and Melo, F. S. (2025). Implicit repair with reinforcement learning in emergent communication. In *24th International Conference on Autonomous Agents and Mutliagent Systems (AAMAS 2025)*.

Walsh, P., Soldaini, L., Groeneveld, D., Lo, K., Arora, S., Bhagia, A., Gu, Y., Huang, S., Jordan, M., Lambert, N., Schwenk, D., Tafjord, O., Anderson, T., Atkinson, D., Brahman, F., Clark, C., Dasigi, P., Dziri, N., Guerquin, M., Ivison, H., Koh, P. W., Liu, J., Malik, S., Merrill, W., Miranda, L. J. V., Morrison, J., Murray, T., Nam, C., Pyatkin, V., Rangapur, A., Schmitz, M., Skjonsberg, S., Wadden, D., Wilhelm, C., Wilson, M., Zettlemoyer, L., Farhadi, A., Smith, N. A., and Hajishirzi, H. (2025). 2 olmo 2 furious.

Wang, X., Ma, B., Hu, C., Weber-Genzel, L., Röttger, P., Kreuter, F., Hovy, D., and Plank, B. (2024). "my answer is C": First-token probabilities do not match text answers in instruction-tuned language models. In Ku, L.-W., Martins, A., and Srikumar, V., editors, *Findings of the Association for Computational Linguistics: ACL 2024*, pages 7407–7416, Bangkok, Thailand. Association for Computational Linguistics.

Ward, J., Huckstep, B., and Tsakanikos, E. (2006). Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all? *Cortex*, 42(2):264–280.

Warstadt, A. (2022). *Artificial Neural Networks as Models of Human Language Acquisition*. New York University.

Warstadt, A. and Bowman, S. R. (2022). What artificial neural networks can tell us about human language acquisition. *Algebraic Structures in Natural Language*, pages 17–60.

Warstadt, A., Mueller, A., Choshen, L., Wilcox, E., Zhuang, C., Ciro, J., Mosquera, R., Paranjabe, B., Williams, A., Linzen, T., and Cotterell, R. (2023). Findings of the BabyLM challenge: Sample-efficient pretraining on developmentally plausible corpora. In Warstadt, A., Mueller, A., Choshen, L., Wilcox, E., Zhuang, C., Ciro, J., Mosquera, R., Paranjabe, B., Williams, A., Linzen, T., and Cotterell, R., editors, *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*, pages 1–34, Singapore. Association for Computational Linguistics.

Washburn, M. F. (1916). *Movement and mental imagery: Outlines of a motor theory of the complexer mental processes*. Houghton Mifflin.

Weber, L., Bruni, E., and Hupkes, D. (2023). Mind the instructions: a holistic evaluation of consistency and interactions in prompt-based learning. In Jiang, J., Reitter, D., and Deng, S., editors, *Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL)*, pages 294–313, Singapore. Association for Computational Linguistics.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., and Fedus, W. (2022). Emergent abilities of large language models. *Transactions on Machine Learning Research*.

Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and language*, 93(1):10–19.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wilcox, E. G., Futrell, R., and Levy, R. (2023). Using Computational Models to Test Syntactic Learnability. *Linguistic Inquiry*, pages 1–44.

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8:229–256.

Winter, B., Perlman, M., Perry, L. K., and Lupyan, G. (2017). Which words are most iconic?: Iconicity in english sensory words. *Interaction Studies*, 18:443–464.

Wittgenstein, L. (1953). *Philosophical Investigations*. Basil Blackwell, Oxford.

Xu, T., Kuribayashi, T., Oseki, Y., Cotterell, R., and Warstadt, A. (2025). Can language models learn typologically implausible languages?

Xu, Z., Niethammer, M., and Raffel, C. A. (2022). Compositional generalization in unsupervised compositional representation learning: A study on disentanglement and emergent language. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A., editors, *Advances in Neural Information Processing Systems*, volume 35, pages 25074–25087. Curran Associates, Inc.

Yiu, E., Kosoy, E., and Gopnik, A. (2024). Transmission versus truth, imitation versus innovation: What children can do that large language and language-and-vision models cannot (yet). *Perspectives on Psychological Science*, 19(5):874–883. PMID: 37883796.

Zhang, J., Huang, J., Jin, S., and Lu, S. (2024a). Vision-language models for vision tasks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–20.

Zhang, Y., Verhoef, T., van Noord, G., and Bisazza, A. (2024b). Endowing neural language learners with human-like biases: A case study on dependency length minimization. In Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N., editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5819–5832, Torino, Italia. ELRA and ICCL.

Zhang, Y., Zhang, C., Tang, Y., and He, Z. (2024c). Cross-modal concept learning and inference for vision-language models. *Neurocomputing*, 583:127530.

Zheng, C., Zhang, J., Kembhavi, A., and Krishna, R. (2024). Iterated learning improves compositionality in large vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13785–13795.

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E., Zhang, H., Gonzalez, J. E., and Stoica, I. (2023). Judging llm-as-a-judge with mt-bench and chatbot arena. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S., editors, *Advances in Neural Information Processing Systems*, volume 36, pages 46595–46623. Curran Associates, Inc.

Zhou, C., Liu, P., Xu, P., Iyer, S., Sun, J., Mao, Y., Ma, X., Efrat, A., Yu, P., Yu, L., Zhang, S., Ghosh, G., Lewis, M., Zettlemoyer, L., and Levy, O. (2023). Lima: less is more for alignment. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA. Curran Associates Inc.