Emergence of communication and inductive biases towards compositionality

CaféTAL

May 23, 2022

In collaboration with Timothée Bernard

Outline

Intro

Exp. 1: Doing the math

Exp. 2: Signaling games & pretraining

Exp. 3: Signaling game & GANs (Teaser)

Conclusions

Outline

Intro

Exp. 1: Doing the math

Exp. 2: Signaling games & pretraining

Exp. 3: Signaling game & GANs (Teaser)

Conclusions

Emergence of Communication

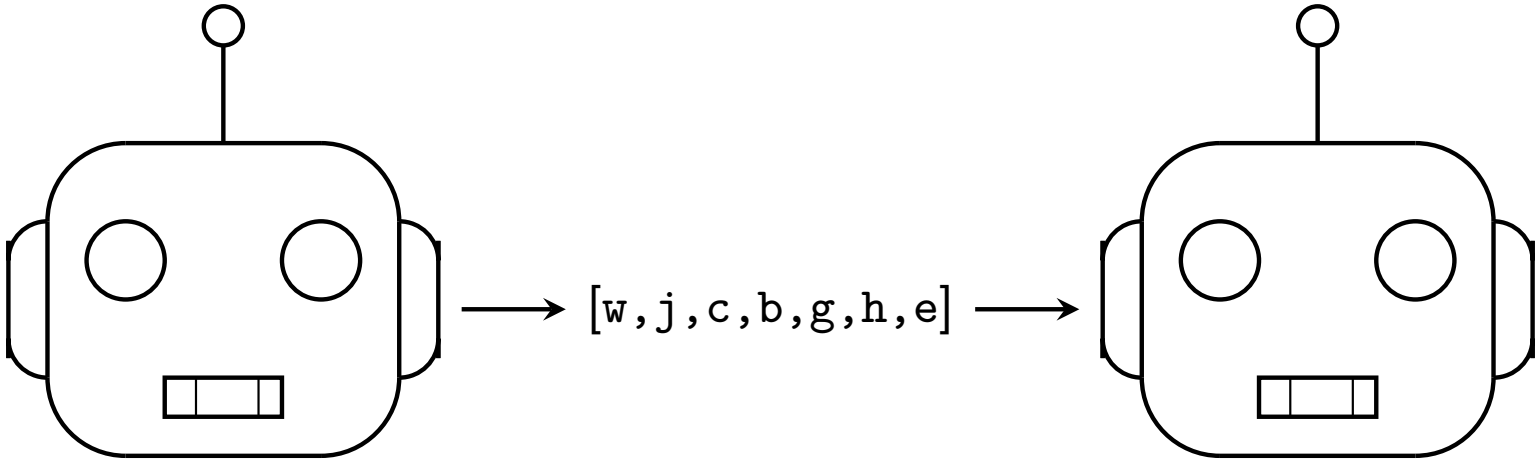
- How did language evolve?

Emergence of Communication

- How did language evolve?
- Can we replicate that?

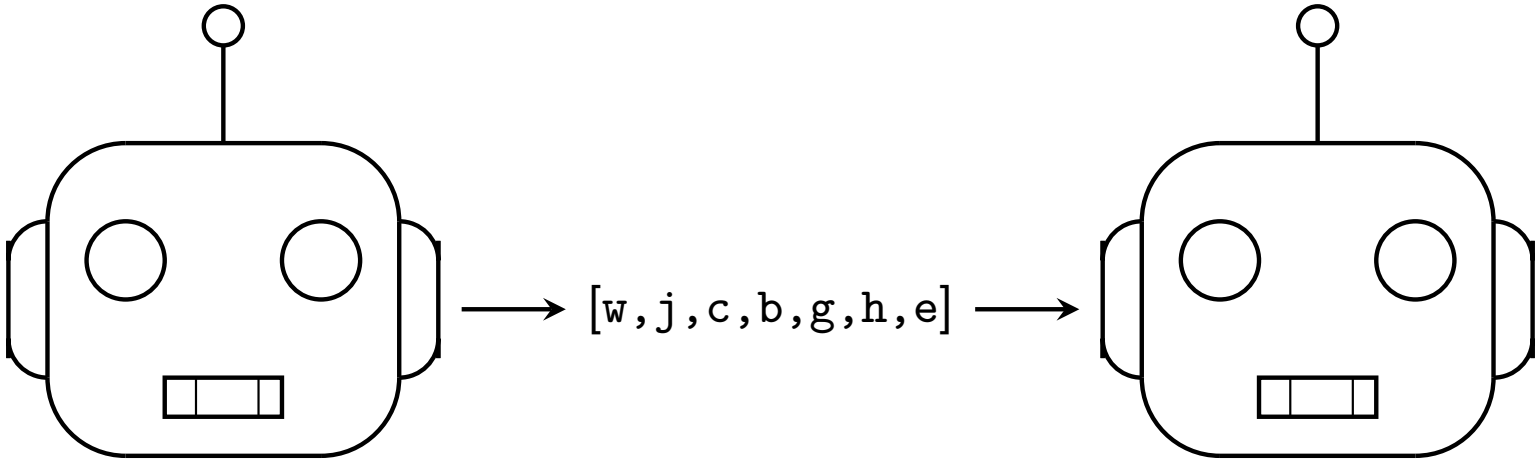
Emergence of Communication

- How did language evolve?
- Can we replicate that?



Emergence of Communication

- How did language evolve?
- Can we replicate that?



- Can we get specific characteristics, like compositionality?

Measuring compositionality

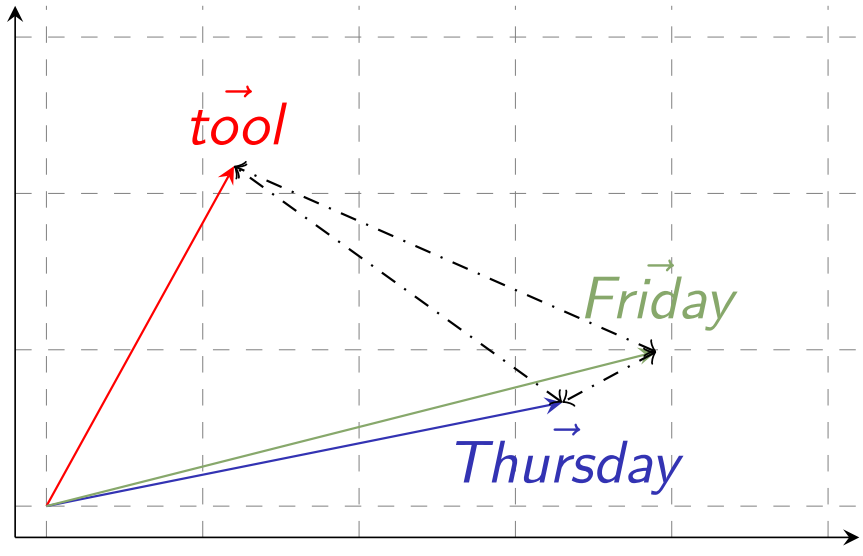
- One way to measure compositionality is topographic similarity (Brighton et al., 2006)

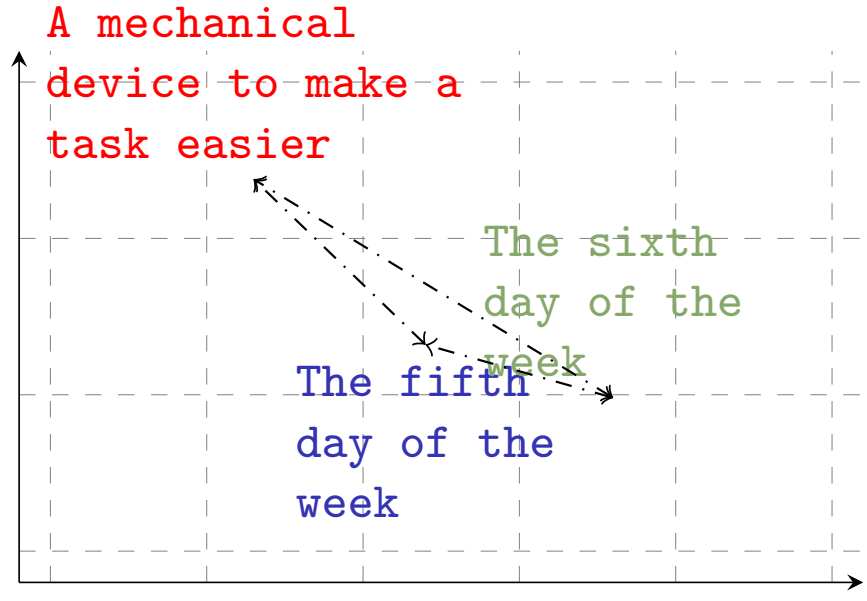
Measuring compositionality

- One way to measure compositionality is topographic similarity (Brighton et al., 2006)
- Compositional $\Longleftrightarrow$ gradual changes in form entail gradual changes in meaning

Measuring compositionality

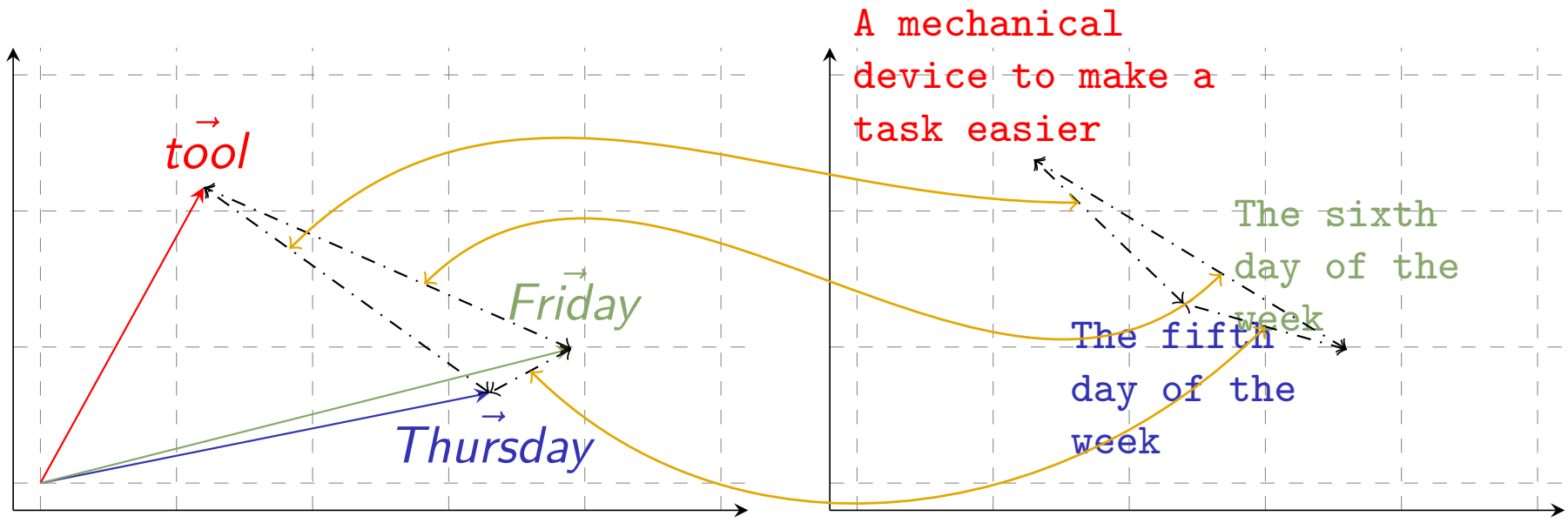
- One way to measure compositionality is topographic similarity (Brighton et al., 2006)
- Compositional $\iff$ gradual changes in form entail gradual changes in meaning





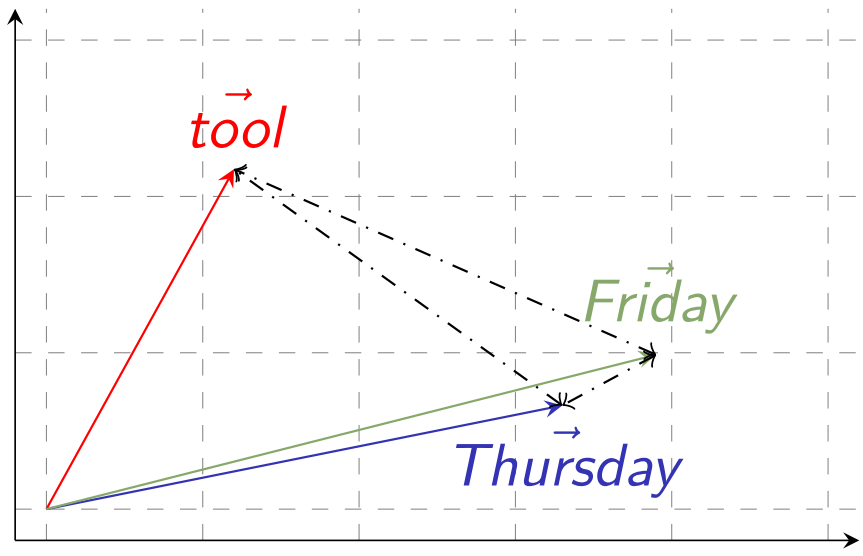
Measuring compositionality

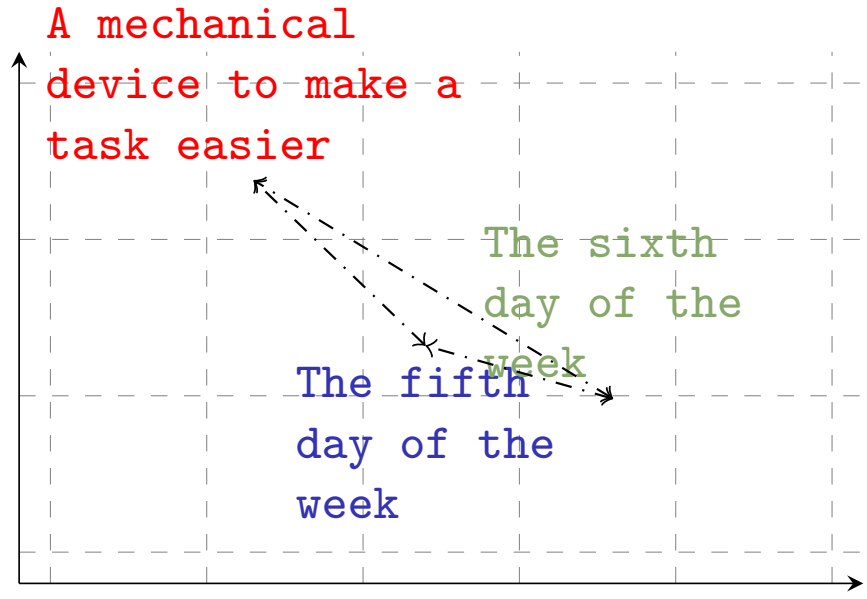
- One way to measure compositionality is topographic similarity (Brighton et al., 2006)
- Compositional $\iff$ gradual changes in form entail gradual changes in meaning



Measuring compositionality

- One way to measure compositionality is topographic similarity (Brighton et al., 2006)
- Compositional $\iff$ gradual changes in form entail gradual changes in meaning





- Still an open question

What we'll see today:

1. some language games designed for the emergence of compositionality
2. some tweaks and tricks to get the model to produce more reliable outputs

Outline

Intro

Exp. 1: Doing the math

Exp. 2: Signaling games & pretraining

Exp. 3: Signaling game & GANs (Teaser)

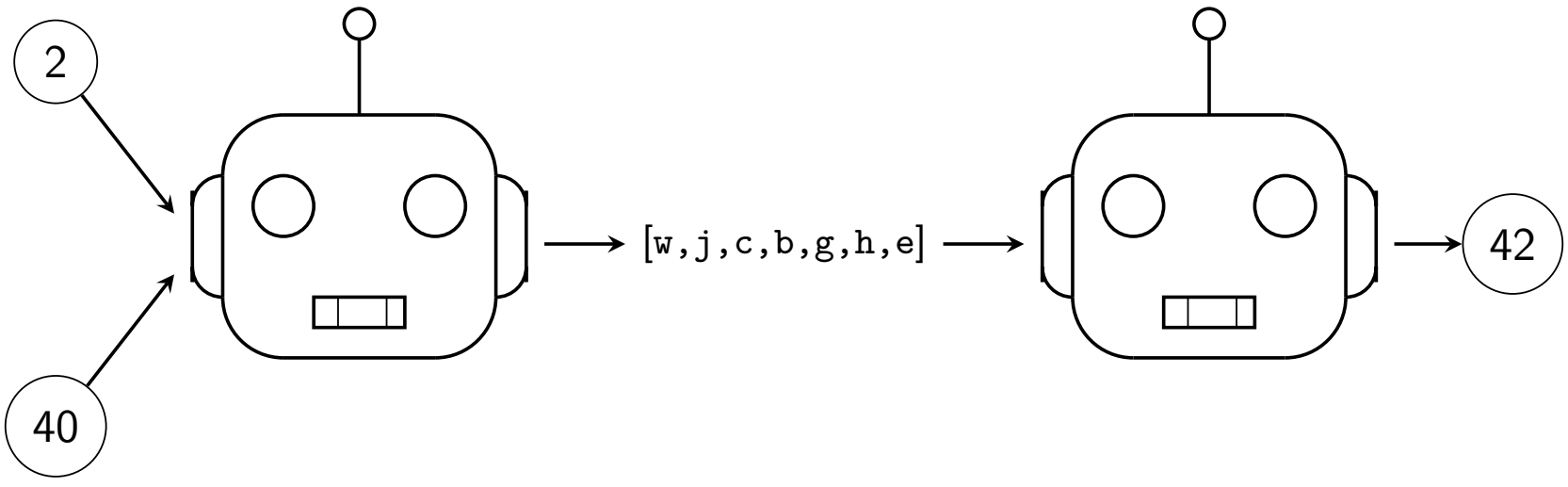
Conclusions

The sum game

- What's a good task for testing compositionality?

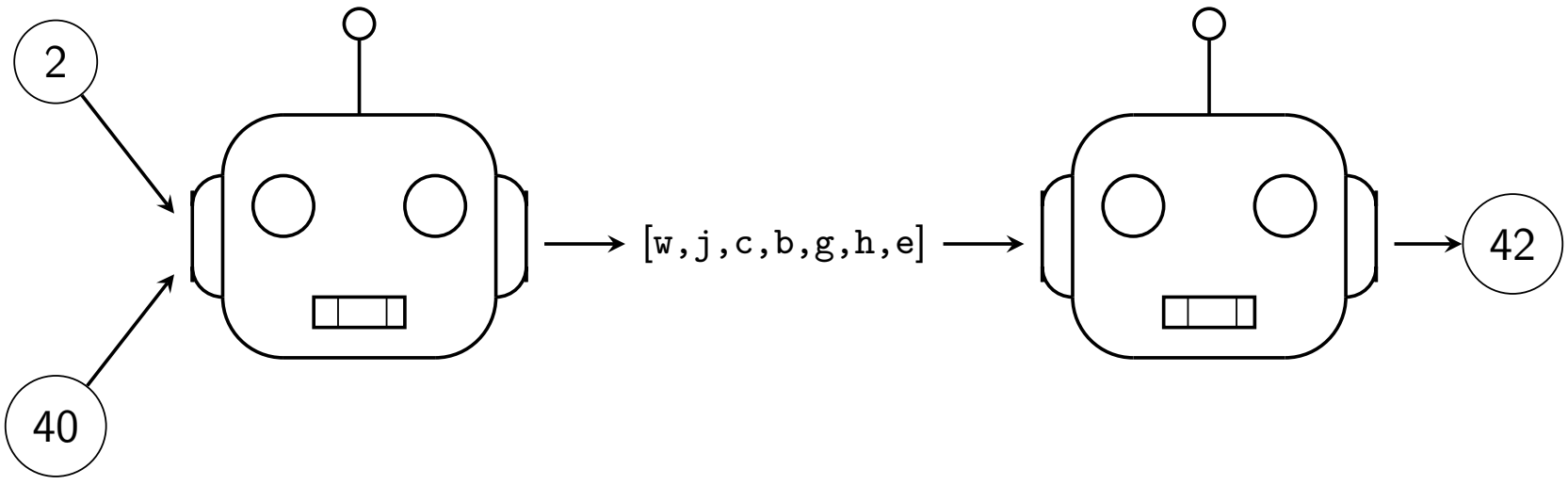
The sum game

- What's a good task for testing compositionality?



The sum game

- What's a good task for testing compositionality?



- We can use the known additive structure to see what's encoded: pairs of integers, sum of integers, other?

Classification vs. Regression

How should we train such a setup?

Classification vs. Regression

How should we train such a setup?

- as a classification problem: the correct label is the matching sum

Classification vs. Regression

How should we train such a setup?

- as a classification problem: the correct label is the matching sum
- as a regression problem: the receiver has to output the matching sum

Classification vs. Regression

How should we train such a setup?

- as a classification problem: the correct label is the matching sum
- as a regression problem: the receiver has to output the matching sum

Computational details:

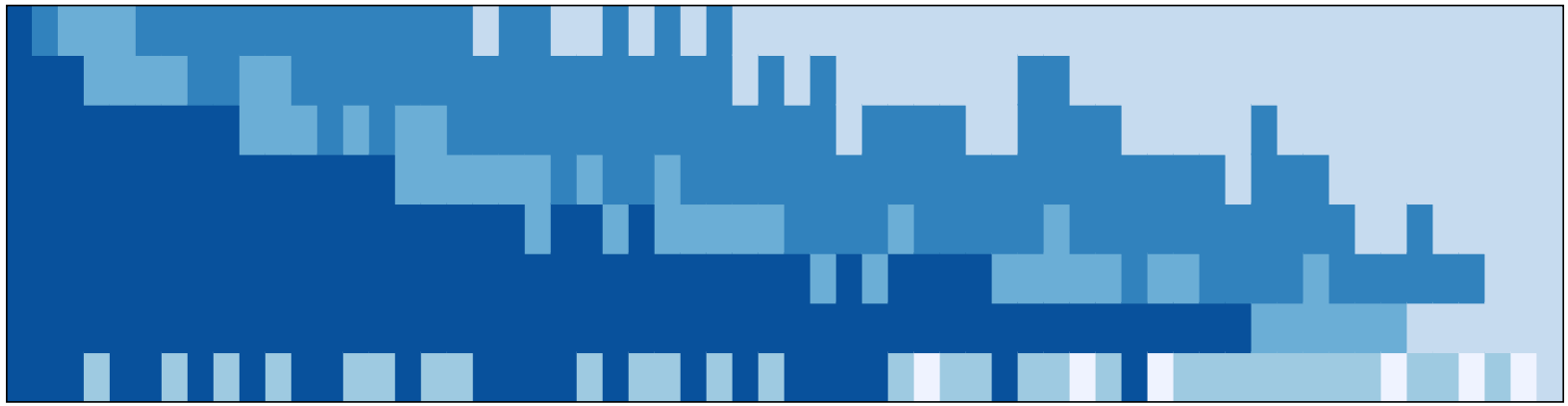
- LSTM-based agents
- inptus ar represented using a concatenation of learned embeddings: $e(40) \oplus e(2)$
- Exploring hyperparameters with Bayesian Optimization.

Classification-based results

	Train	**Dev**	**Test**
XENT	2.718	2.751	2.723
Acc.	0.191	0.195	0.163
$\rho_{\vec{e}}$	0.204	0.210	0.192
$\rho_{\langle a,b \rangle}$	0.545	0.573	0.538
ρ_{a+b}	0.846	0.862	0.831

Classification-based results

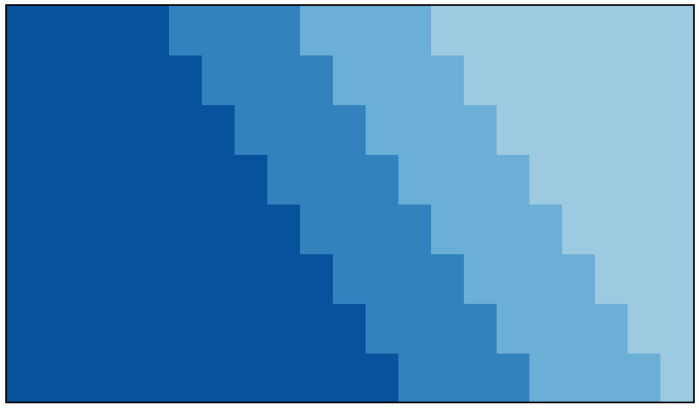
	Train	Dev	Test
XENT	2.718	2.751	2.723
Acc.	0.191	0.195	0.163
$\rho_{\vec{e}}$	0.204	0.210	0.192
$\rho_{\langle a,b \rangle}$	0.545	0.573	0.538
ρ_{a+b}	0.846	0.862	0.831



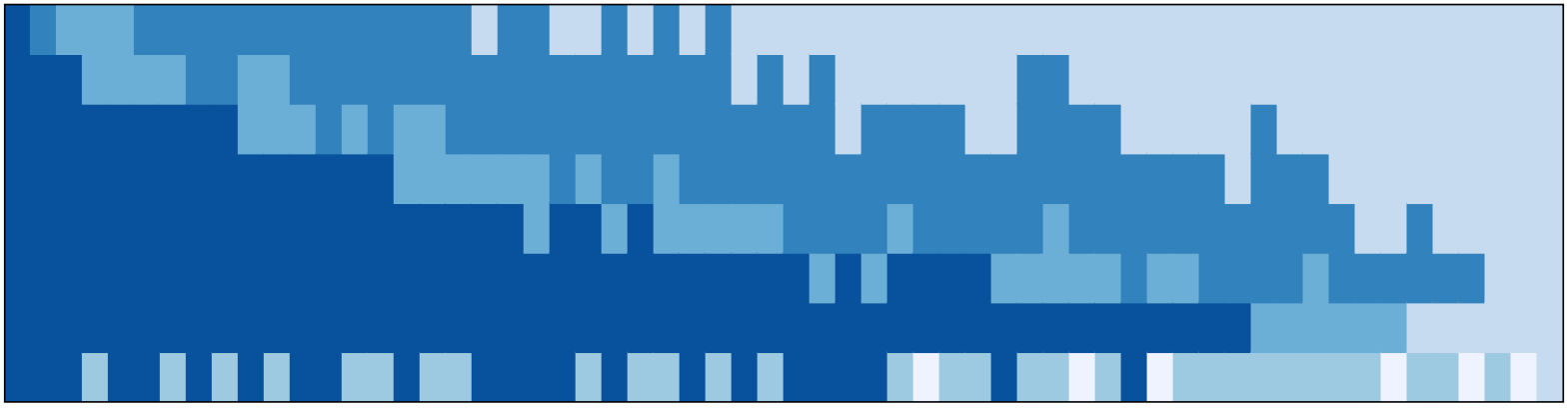
messages produced

Classification-based results

	Train	Dev	Test
XENT	2.718	2.751	2.723
Acc.	0.191	0.195	0.163
$\rho_{\vec{e}}$	0.204	0.210	0.192
$\rho_{\langle a,b \rangle}$	0.545	0.573	0.538
ρ_{a+b}	0.846	0.862	0.831



FILO structure



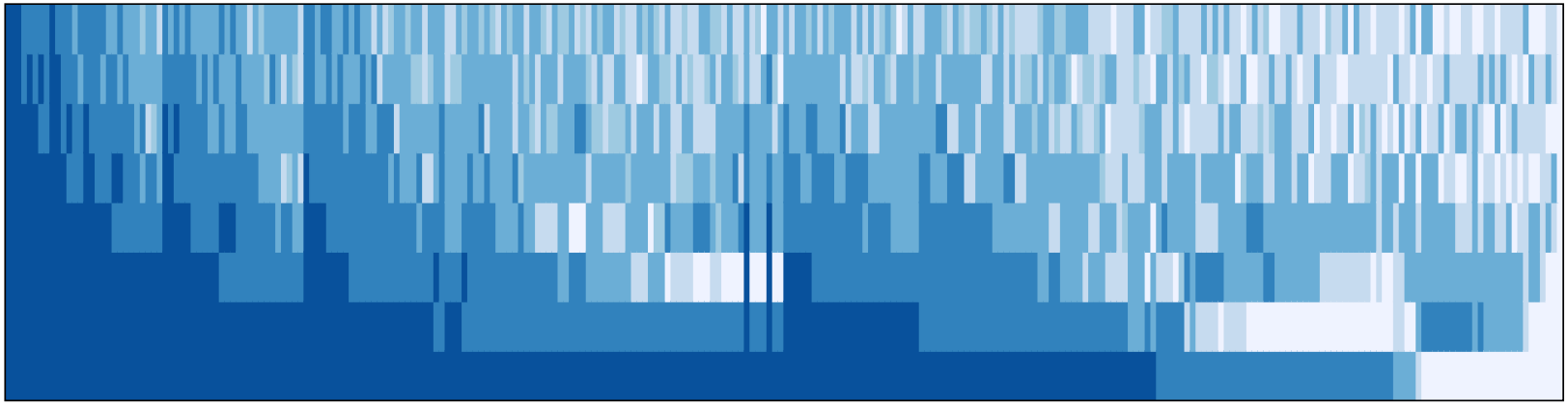
messages produced

Regression-based results

	Train	**Dev**	**Test**
MSE	0.260	0.259	0.269
Acc.	0.712	0.741	0.734
$\rho_{\vec{e}}$	0.140	0.152	0.131
$\rho_{\langle a,b \rangle}$	0.459	0.487	0.454
ρ_{a+b}	0.722	0.722	0.704

Regression-based results

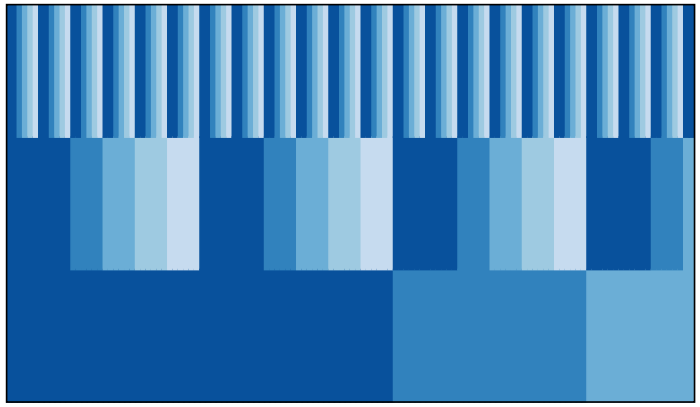
	Train	Dev	Test
MSE	0.260	0.259	0.269
Acc.	0.712	0.741	0.734
$\rho_{\vec{e}}$	0.140	0.152	0.131
$\rho_{\langle a,b \rangle}$	0.459	0.487	0.454
ρ_{a+b}	0.722	0.722	0.704



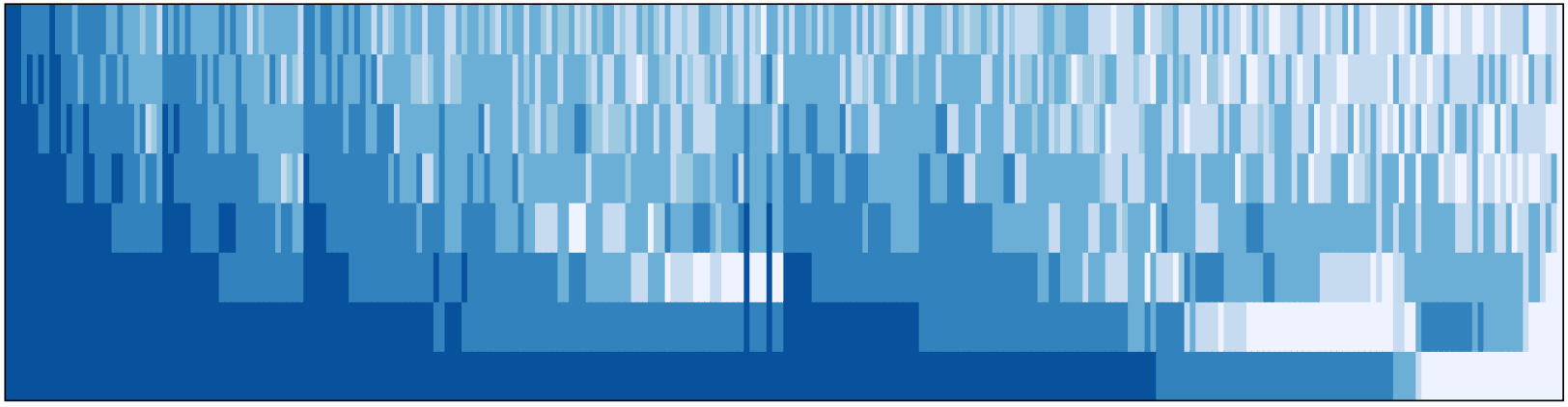
messages produced

Regression-based results

	Train	Dev	Test
MSE	0.260	0.259	0.269
Acc.	0.712	0.741	0.734
$\rho_{\vec{e}}$	0.140	0.152	0.131
$\rho_{\langle a,b \rangle}$	0.459	0.487	0.454
ρ_{a+b}	0.722	0.722	0.704



$a + b$, expressed in base 6



messages produced

In short

- How to best train a model is dependent on the exact language game

In short

- How to best train a model is dependent on the exact language game
- More effective training doesn't necessary entail more compositional outputs

In short

- How to best train a model is dependent on the exact language game
- More effective training doesn't necessary entail more compositional outputs

- This task is fairly limited

Outline

Intro

Exp. 1: Doing the math

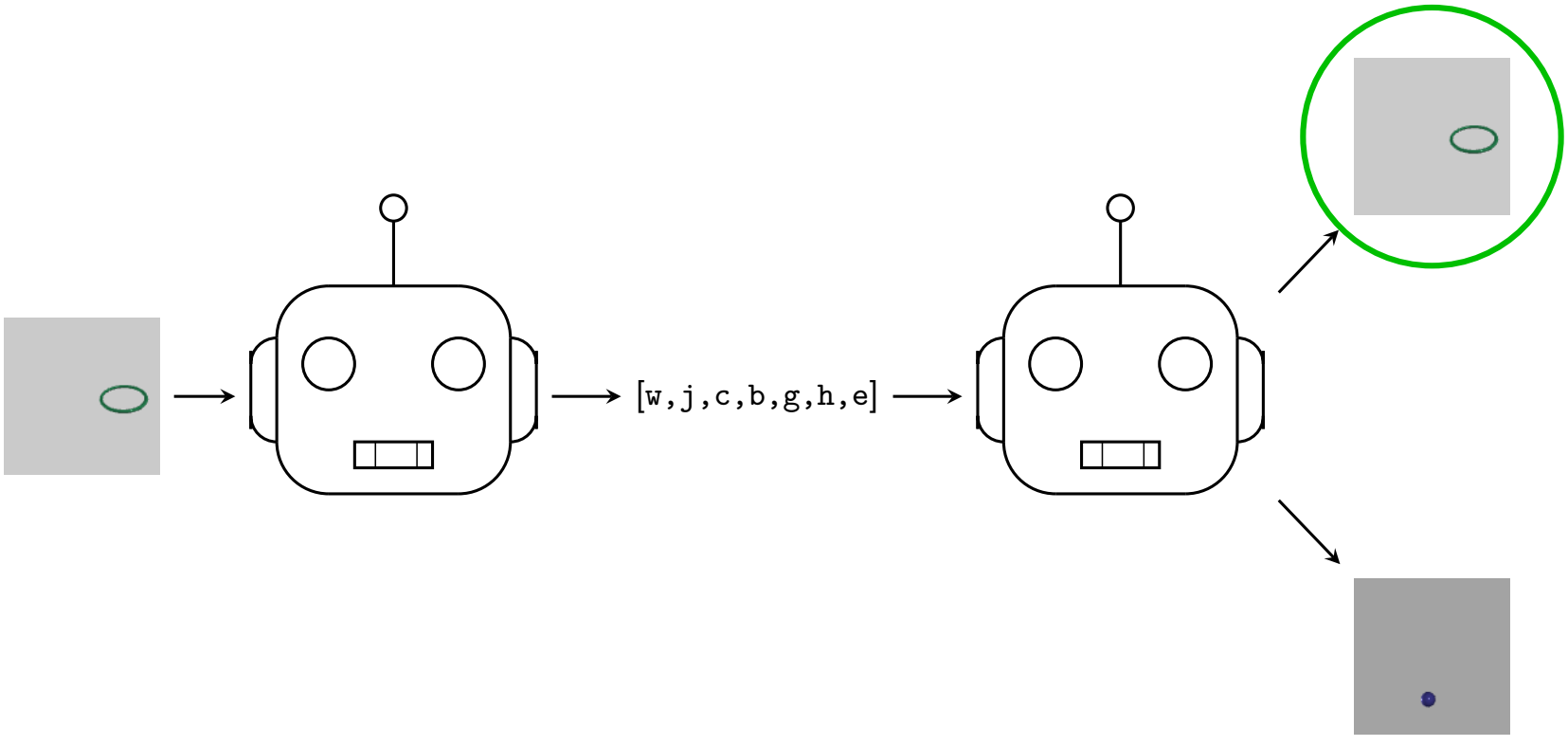
Exp. 2: Signaling games & pretraining

Exp. 3: Signaling game & GANs (Teaser)

Conclusions

Signaling Game

- Have the receiver select the image shown to the sender



Synthetic Dataset

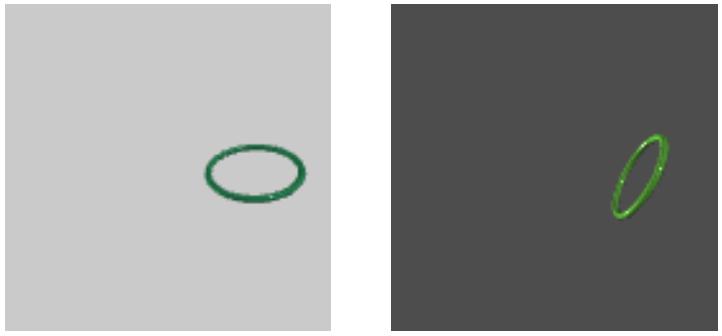
Images each containing one object, based on five *features*: **color**, **shape**, **size**, **vertical position** and **horizontal position**

Synthetic Dataset

Images each containing one object, based on five *features*: **color**, **shape**, **size**, **vertical position** and **horizontal position**

Some examples:

Images of different categories



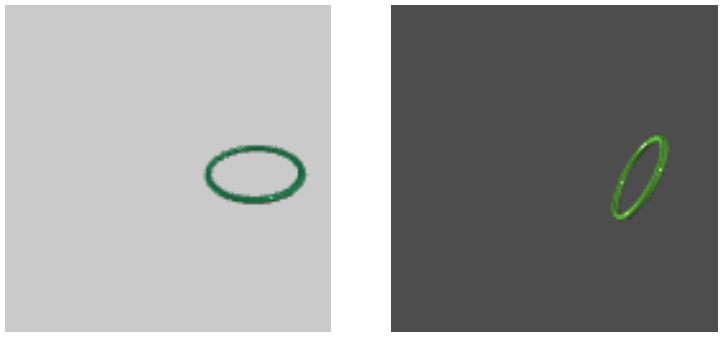
Images of the same category

Synthetic Dataset

Images each containing one object, based on five *features*: **color**, **shape**, **size**, **vertical position** and **horizontal position**

Some examples:

Images of different categories



Images of the same category

- The sender has to convey the values of the 5 features

Baking in inductive bias

LSTMs are not biased towards compositionality (Liška et al., 2018)
We compare four pretraining regimens:

- no pretraining
- category-wise classification: predict the category of the presented image
- feature-wise classification: predict the value of each feature independently (one classifier per feature)
- auto-encoding: learn to reconstruct the full image.

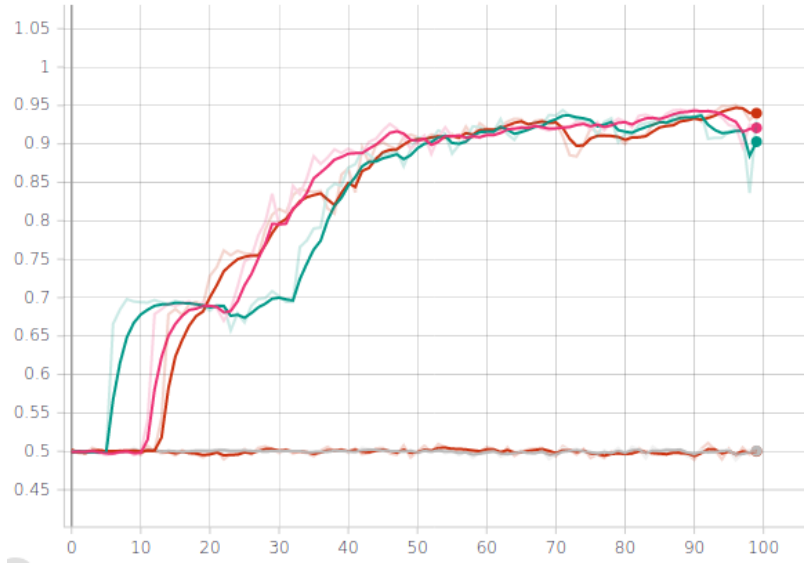
Baking in inductive bias

LSTMs are not biased towards compositionality (Liška et al., 2018)
We compare four pretraining regimens:

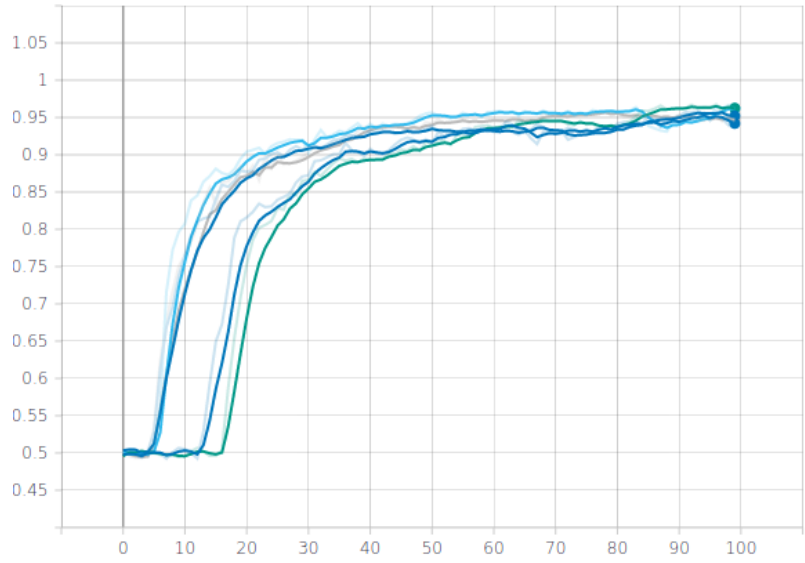
- no pretraining
- category-wise classification: predict the category of the presented image
- feature-wise classification: predict the value of each feature independently (one classifier per feature)
- auto-encoding: learn to reconstruct the full image.

Computational details: CNN + LSTM agents. Hyperparameters are explored by grid. We also study whether to freeze CNN weights after pretraining, or whether further adaptation is required.

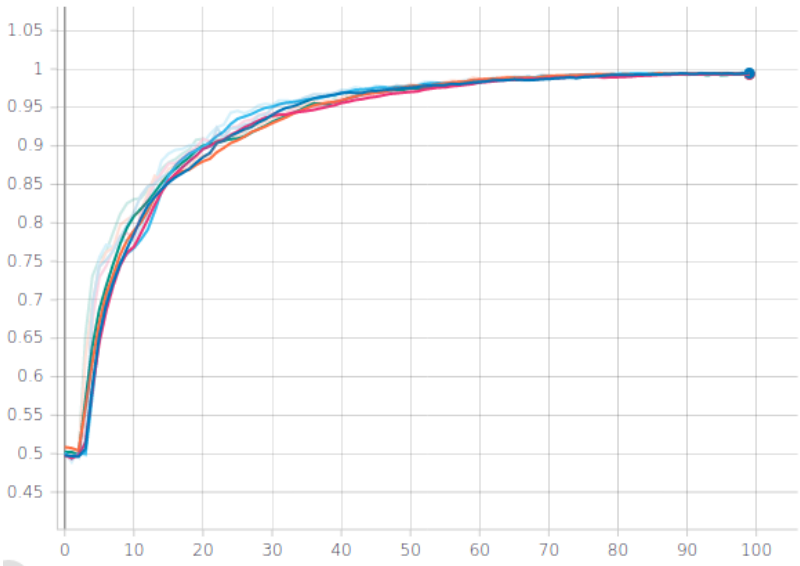
Results: Accuracy



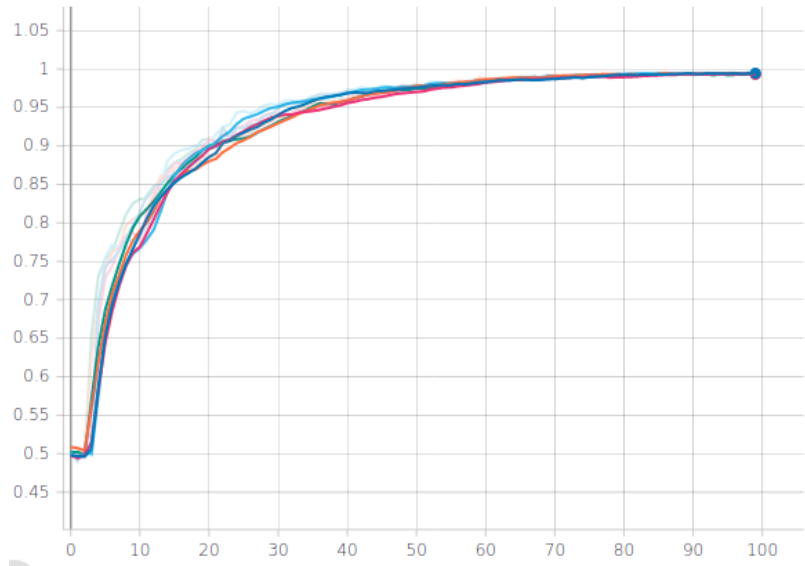
No pretraining



Auto-encoder

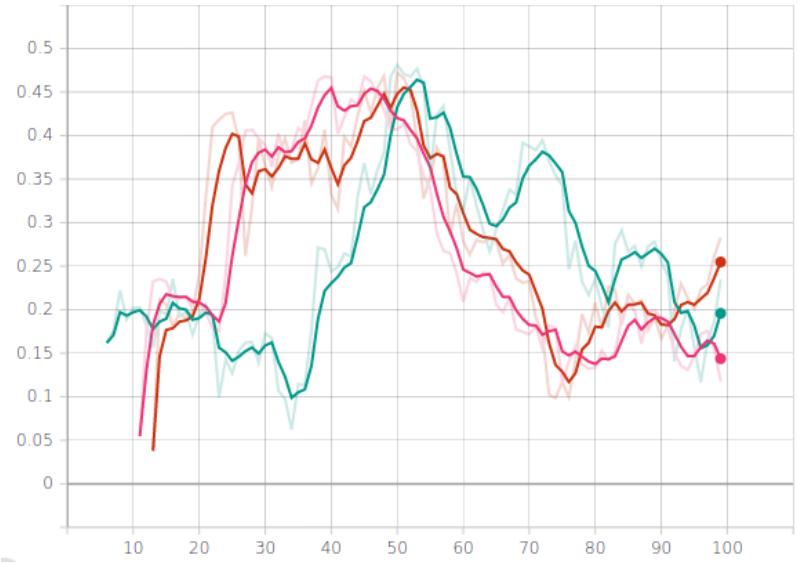


Feature-wise (Frozen)

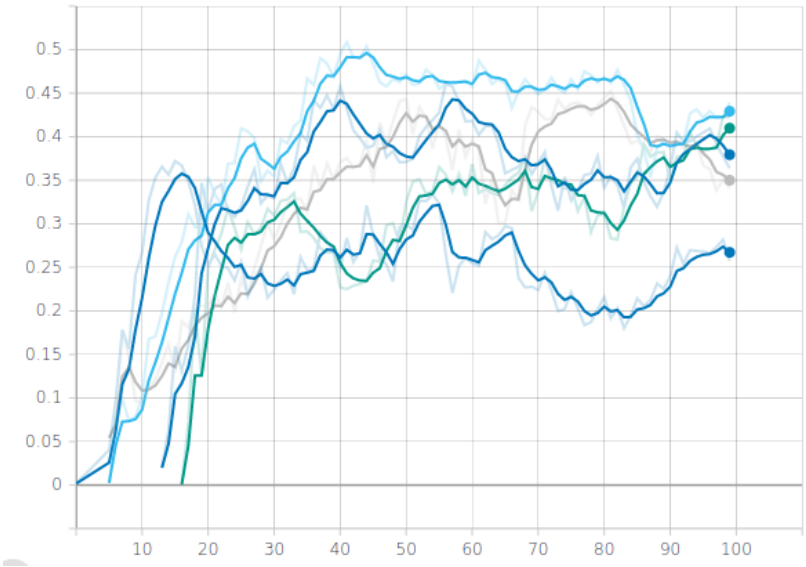


Category-wise (Frozen)

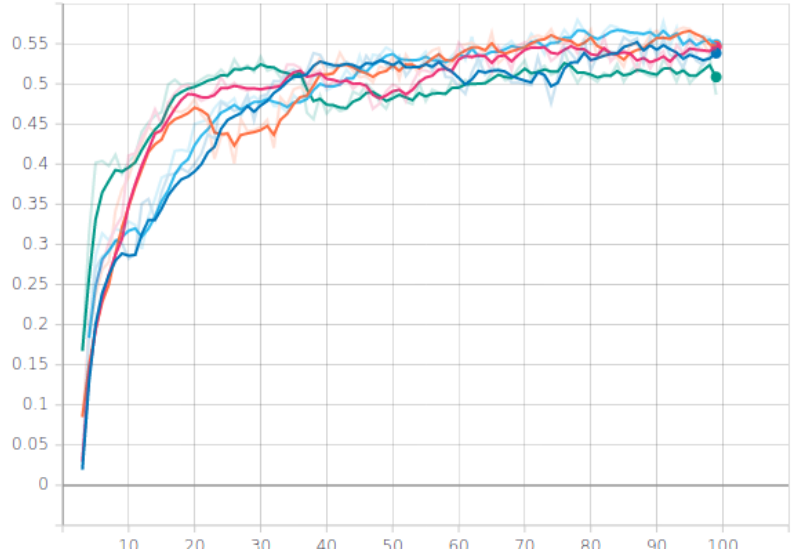
Results: Compositionality



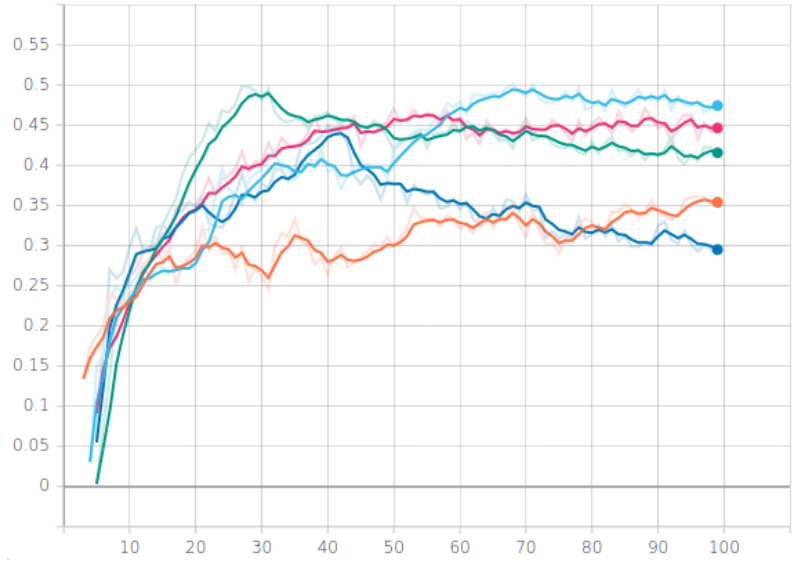
No pretraining



Auto-encoder



Feature-wise (Frozen)



Category-wise (Frozen)

In short

- Any pretraining helps

In short

- Any pretraining helps
- More direct supervision helps more

In short

- Any pretraining helps
- More direct supervision helps more

- Can we induce compositionality less directly?

Outline

Intro

Exp. 1: Doing the math

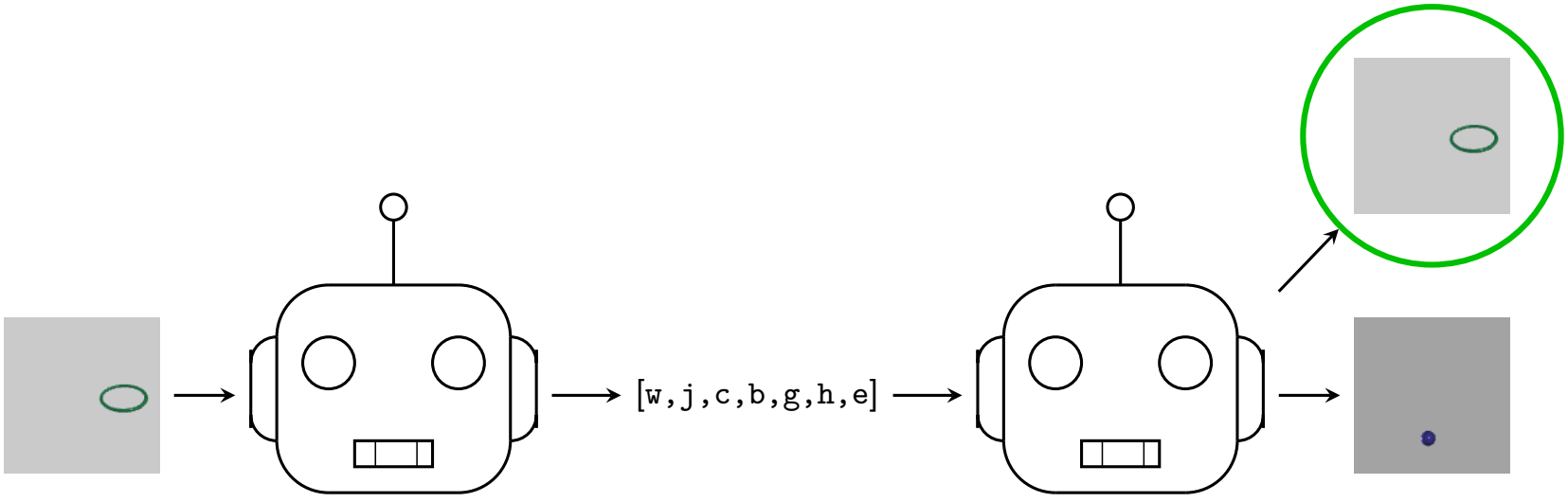
Exp. 2: Signaling games & pretraining

Exp. 3: Signaling game & GANs (Teaser)

Conclusions

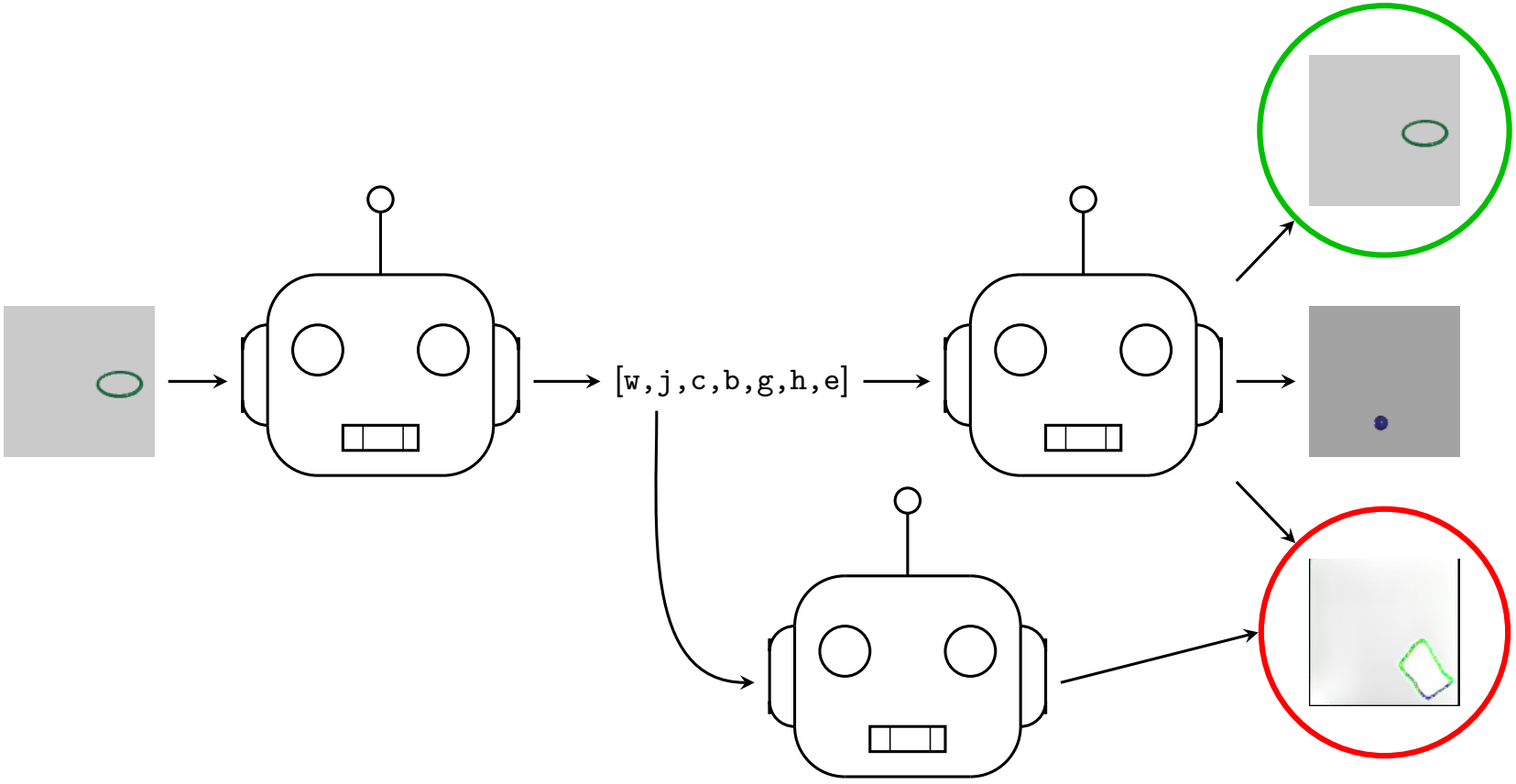
Meet Charlie

- Signaling game with a tweak



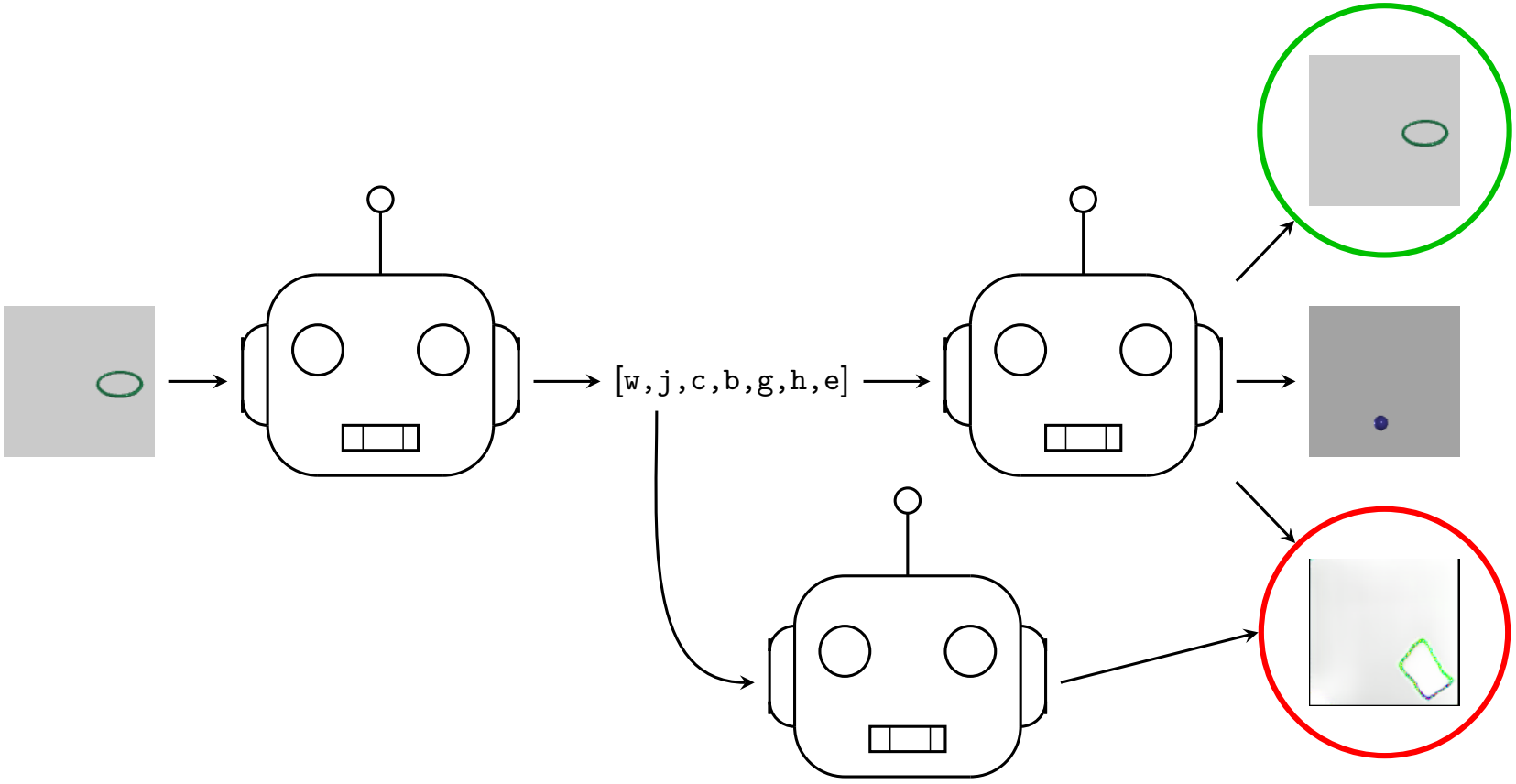
Meet Charlie

- Signaling game with a tweak



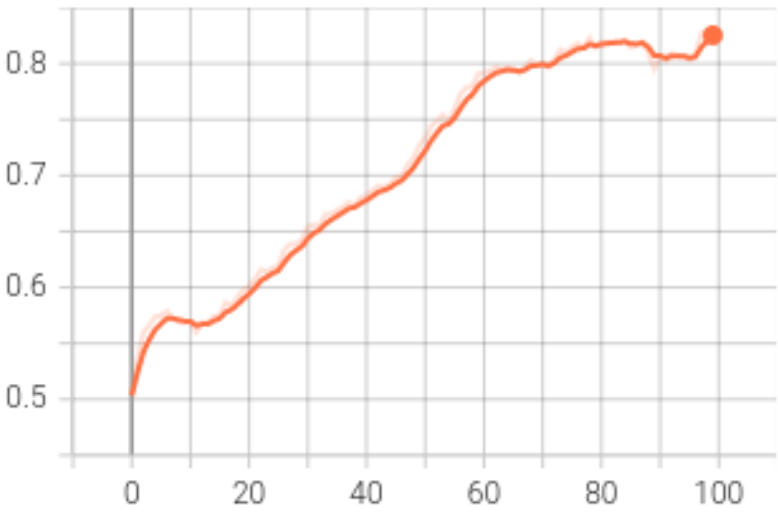
Meet Charlie

- Signaling game with a tweak

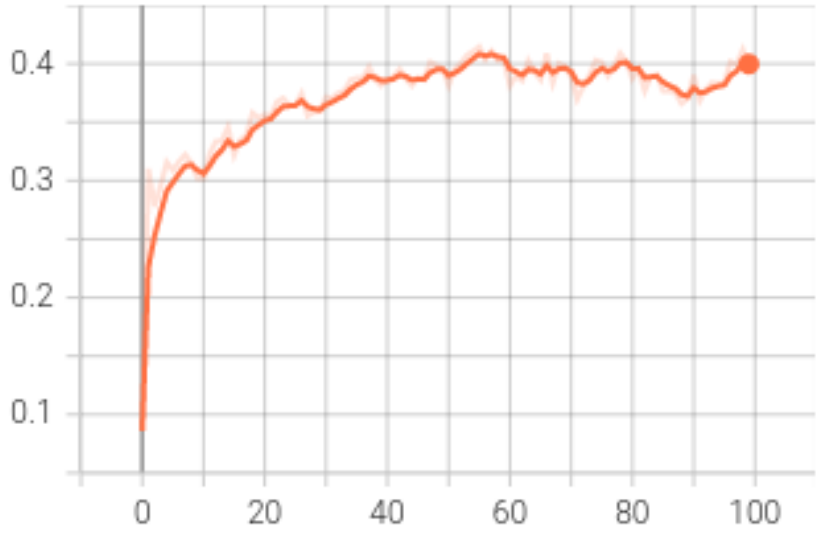


- Feature values are no longer the sole plausible meaning

Early results

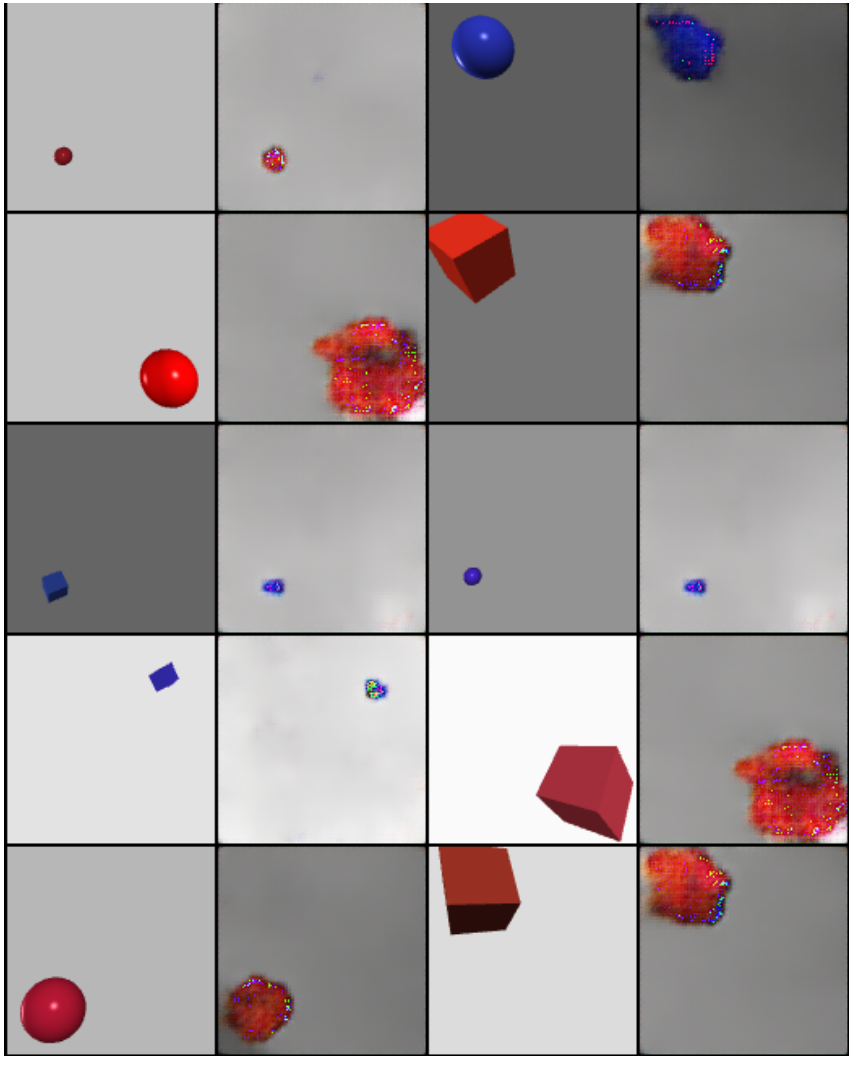


Accuracy

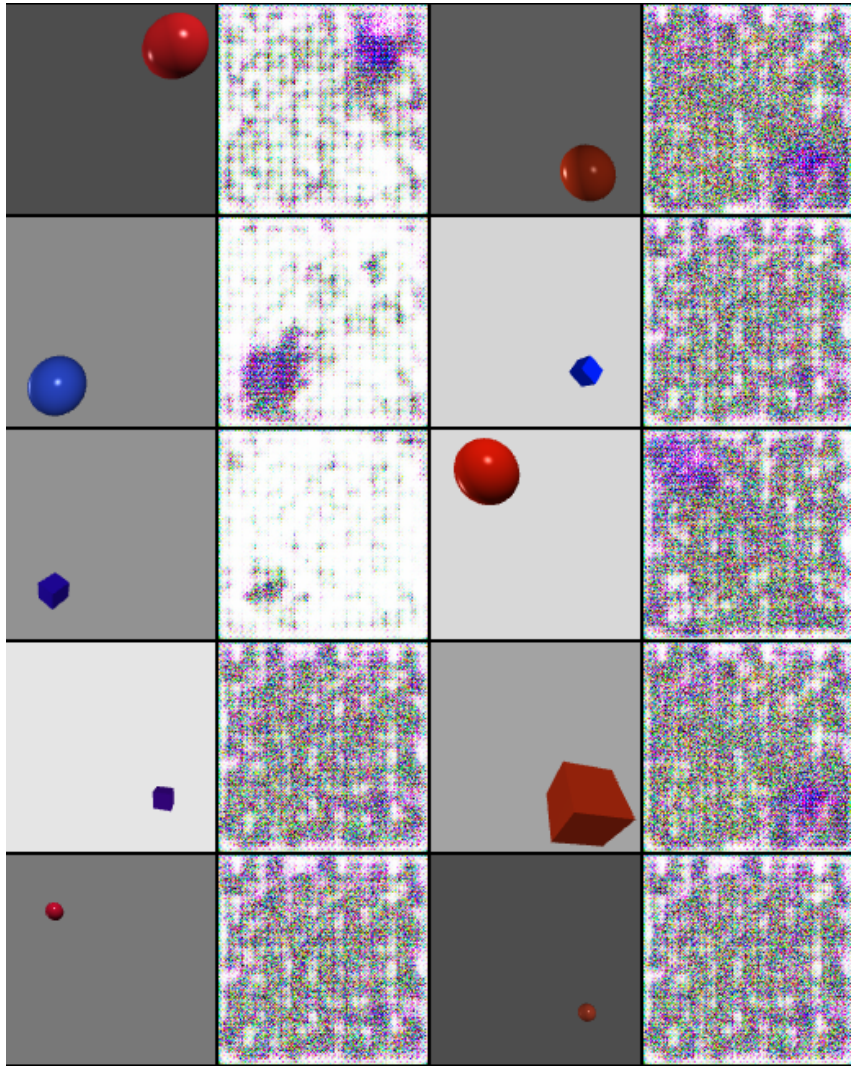


Compositionality

Having a look at the images



Pretraining



No pretraining

In short

- Still possible to get some interesting results

In short

- Still possible to get some interesting results
- Lower accuracy, slower training, CNN pretraining seems necessary

In short

- Still possible to get some interesting results
- Lower accuracy, slower training, CNN pretraining seems necessary

- More work to be done!

Outline

Intro

Exp. 1: Doing the math

Exp. 2: Signaling games & pretraining

Exp. 3: Signaling game & GANs (Teaser)

Conclusions

Conclusions

Today's overview:

Conclusions

Today's overview:

- Not all tasks are equally complex, nor equally likely to yield compositional languages

Conclusions

Today's overview:

- Not all tasks are equally complex, nor equally likely to yield compositional languages
- How to (pre-)train a model is crucial

Conclusions

Today's overview:

- Not all tasks are equally complex, nor equally likely to yield compositional languages
- How to (pre-)train a model is crucial

- Metrics for compositionality are unsatisfactory

Conclusions

Today's overview:

- Not all tasks are equally complex, nor equally likely to yield compositional languages
- How to (pre-)train a model is crucial

- Metrics for compositionality are unsatisfactory
- Many other fators to consider: dataset? reward function?