

# Semantics and Spatiality of Emergent Communication

**Rotem Ben Zion**, Boaz Carmeli, Orr Paradise, Yonatan Belinkov

December 2024



**Berkeley**  
UNIVERSITY OF CALIFORNIA

# Emergent Communication

Language as a **side product** of collaboration



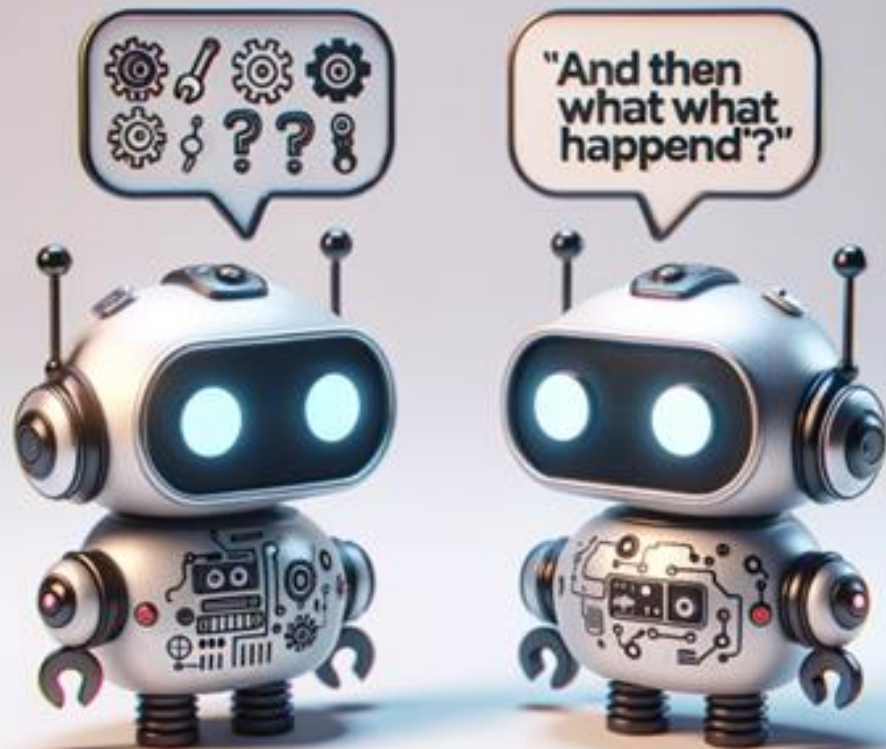
Active communication



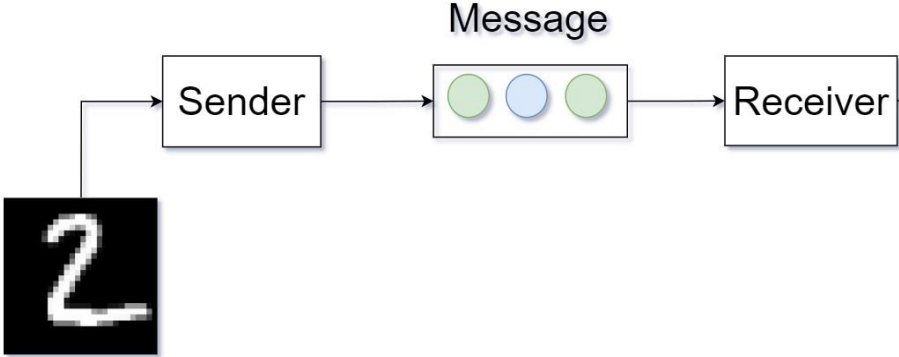
Multi-modal LM training



Insights into language evolution

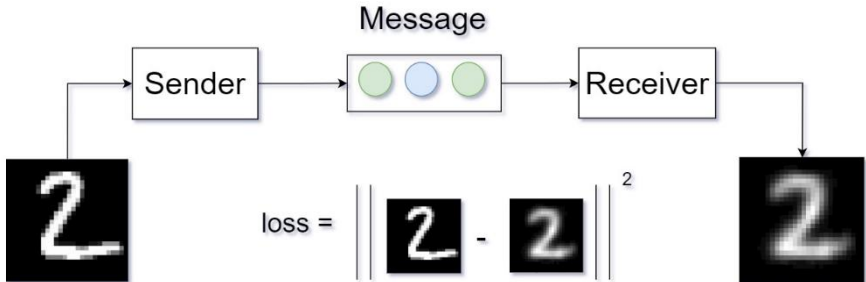


# Lewis games

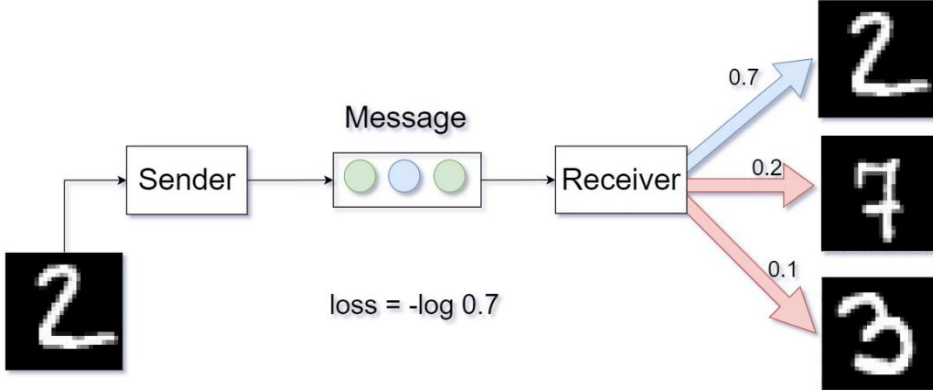


# Lewis games

## Reconstruction



## Discrimination



# A central goal of EC

Developing emergent  
communication protocols with  
properties of natural language

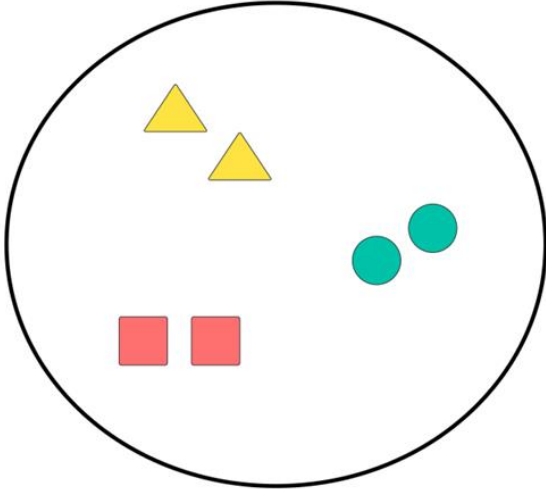
# Emergent Communication is counter-intuitive

**“Not compositional”** ~Kottur, Moura, Lee, Batra (2017)

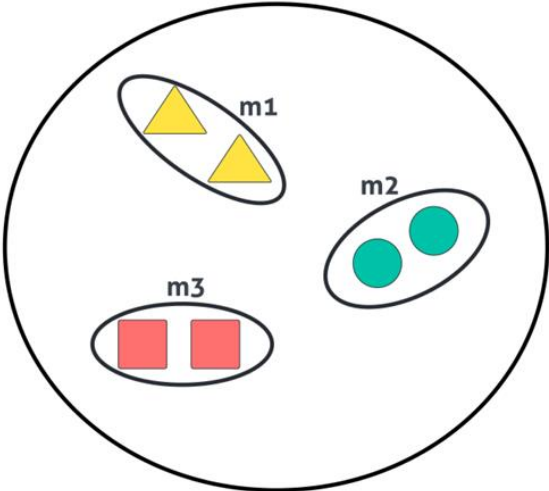
**“Anti-efficient”** ~Chaabouni, Kharitonov, Dupoux, Baroni (2019)

**“Semantic consistency”** ~This paper

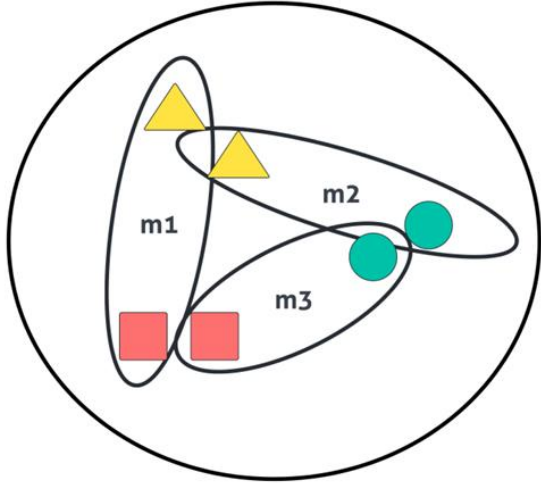
# Semantic consistency



Input space



Semantically consistent mapping



Semantically inconsistent mapping

# Semantic consistency definition

- ▶ On average, inputs mapped to the same message are more similar than random inputs.

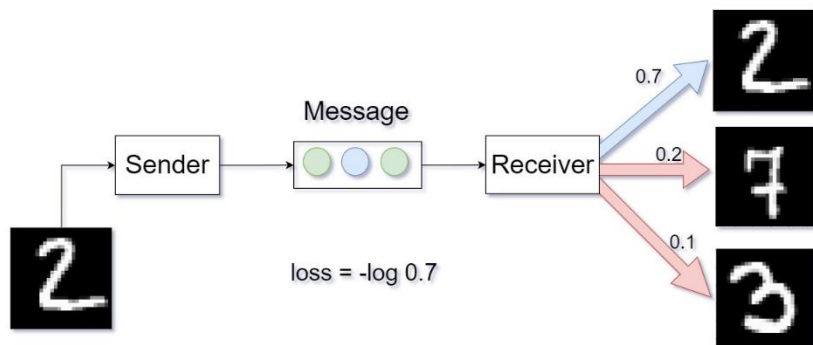
$$\mathbb{E}_{x_1, x_2 \sim X} \left[ \|x_1 - x_2\|^2 \mid S_\theta(x_1) = S_\theta(x_2) \right] < \mathbb{E}_{x_1, x_2 \sim X} \left[ \|x_1 - x_2\|^2 \right]$$

This can be simplified into:

$$\mathbb{E}_{m \sim S_\theta(X)} [\text{Var} [X \mid S_\theta(X) = m]] < \text{Var} [X]$$

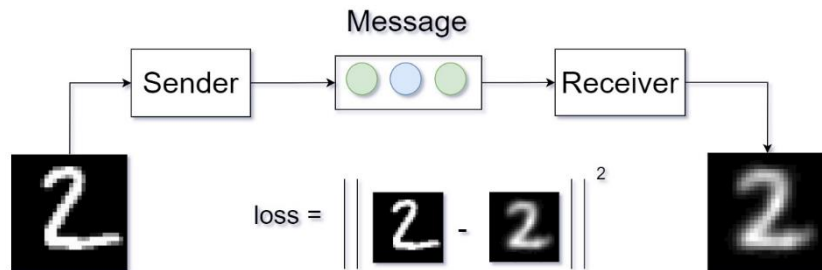


# Formal results



optimality  $\longrightarrow$

**Not necessarily**  
semantically consistent



optimality  $\longrightarrow$

**Necessarily**  
semantically consistent

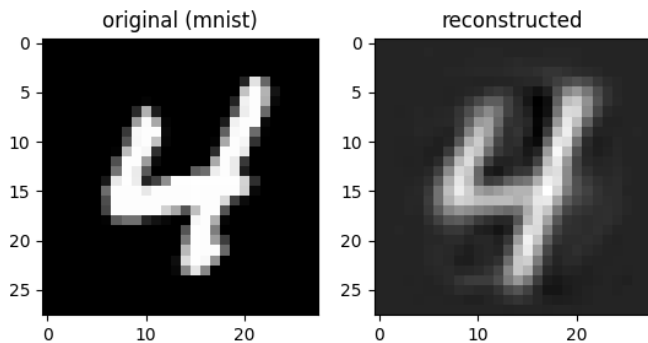
# Other formal results

- ▶ Interpretable equivalent objectives.
- ▶ Variations of both games.
- ▶ Spatial meaningfulness analysis.

# Experiments - MNIST

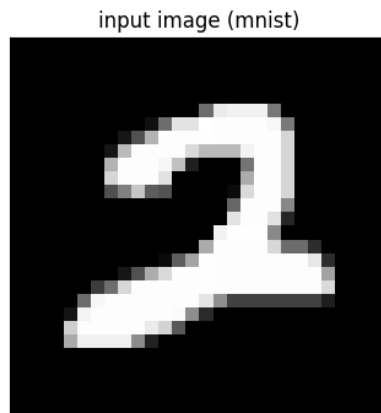
## Reconstruction

Channel message [6, 9, 5, 6]



## Discrimination


Channel message [1, 1, 8, 4]



- candidates
- score: 0.929  

  - score: 0.211  

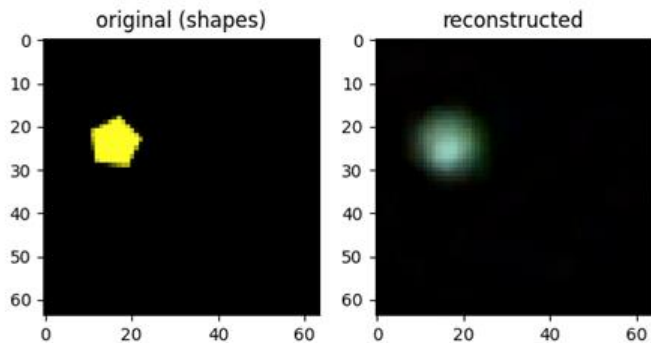
  - score: -0.080  

  - score: -0.098  


# Experiments - Shapes

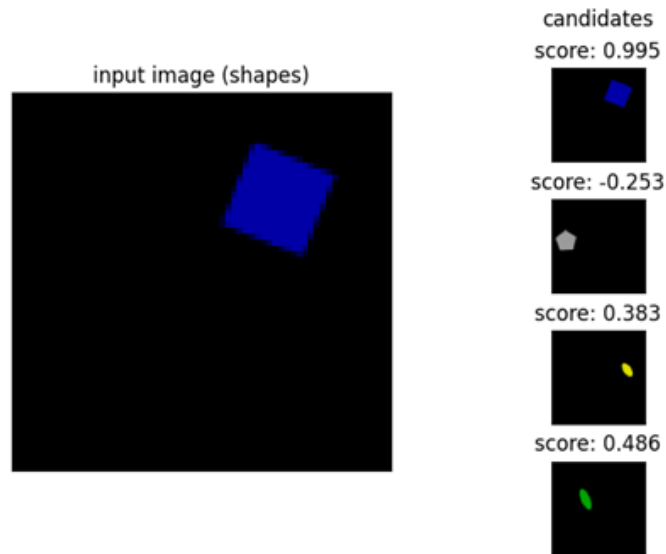
## Reconstruction

Channel message [8, 8, 3, 3]



## Discrimination

Channel message [4, 5, 5, 4]



# Message variance

Table 1: Empirical results on Shapes, averaged over five randomly initialized training runs.

| EC setup       | Unique Msgs        | Disc. accuracy $\uparrow$ | TopSim $\uparrow$ | Msg Var $\downarrow$ |                      |
|----------------|--------------------|---------------------------|-------------------|----------------------|----------------------|
|                |                    |                           |                   | Trained              | Rand                 |
| Reconstruction | 306.60 $\pm$ 28.52 | 31.64 $\pm$ 2.51          | 0.34 $\pm$ 0.02   | 1334.38 $\pm$ 78.05  | 2554.77 $\pm$ 108.19 |
| Discrimination | 251.60 $\pm$ 29.53 | 61.96 $\pm$ 4.78          | 0.09 $\pm$ 0.01   | 2280.24 $\pm$ 157.38 | 2793.65 $\pm$ 115.45 |

# Other empirical results

- ▶ A supervised setting: message purity.
- ▶ Spatial meaningfulness experiments.
- ▶ Message variance vs. compositionality measures.

# Summary: our contributions

- ▶ **Semantic consistency**: a prerequisite to meaningful communication.
- ▶ **Reconstruction** and **discrimination**: formal and empirical analysis.
- ▶ Findings:
  - The discrimination objective can lead to counter-intuitive solutions.
  - The reconstruction objective induces consistent messages.