

Rebuttal file

Table 1: Empirical results on Shapes (5 runs for each task).

Setup	Disc. accuracy (21 candidates)	Num. unique Messages	BosDis	PosDis	S-PosDis	TopSim	Message variance
Reconstruction	36.89 $\pm$ 1.8	326.80 $\pm$ 33.6	0.11 $\pm$ 0.0	0.04 $\pm$ 0.0	0.01 $\pm$ 0.0	0.37 $\pm$ 0.1	1282.42 $\pm$ 81.7
Discrimination	60.83 $\pm$ 1.7	177.20 $\pm$ 19.5	0.23 $\pm$ 0.2	0.13 $\pm$ 0.1	0.01 $\pm$ 0.0	0.08 $\pm$ 0.0	2724.39 $\pm$ 48.0

Table 2: Correlation matrix (5 runs for each task).

	TopSim	Unique Msgs	BosDis	PosDis	S-PosDis	Disc. accuracy	Msg Var
TopSim	1.000	0.800	-0.399	-0.404	0.045	-0.893	-0.904
Unique Msgs	0.800	1.000	-0.333	-0.374	0.089	-0.949	-0.965
BosDis	-0.399	-0.333	1.000	0.982	0.128	0.485	0.451
PosDis	-0.404	-0.374	0.982	1.000	0.083	0.525	0.486
S-PosDis	0.045	0.089	0.128	0.083	1.000	-0.083	-0.131
Disc. accuracy	-0.893	-0.949	0.485	0.525	-0.083	1.000	0.986
Msg Var	-0.904	-0.965	0.451	0.486	-0.131	0.986	1.000

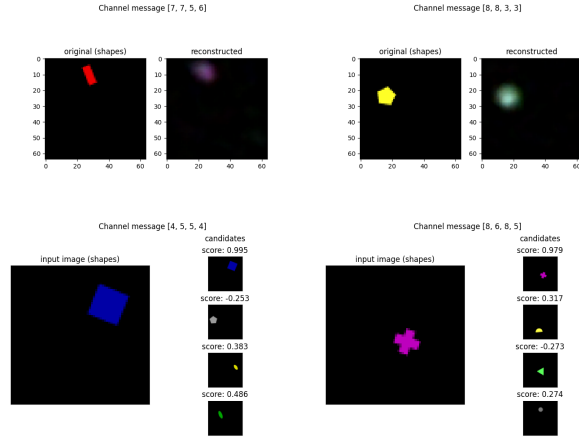


Figure 1: Reconstruction (top) and discrimination (bottom) examples on Shapes