

S8 Fig. Logical alignment of pretrained models. Logical alignment of pretrained models (step=0) as a function of background knowledge ratio in the training corpus (rows) and number of distractors (columns). For each model, 1000 inferences (premises, conclusion, distractors) are sampled, and for each inference, the unconditional degree of belief in the conclusion (x-axis) as well as the conditional degree of belief given premises and distractors are elicited. The y-axis show the absolute difference between conditional and unconditional beliefs.

