verbal - KL Divergence vs Number of Scenarios (Zoomed to 0.5) (Maieutic Prompting Results) 0.5 Before all B sampled Ater all B ampled → Average across runs (±5E) 0.4 0.3 KL Divergence 0.2 -0.1 -0.0 10 20 30 40 50

Number of Scenarios