Naive 0.99 0.61 0.52 0.18 0.89 0.78 0.46 0.64 0.67 0.92 0 0.98 0.71 0.37 0.18 0.56 0.58 0.19 8.0 0.89 0.56  $\vdash$ 0.97 0.54 0.82 0.55 0.57 0.14 0.71 0.8 0.35 0.13  $\sim$ 0.98 0.85 0.94 0.61 0.23 0.74 0.85 0.36 0.79 0.51  $_{\odot}$ Training task 0.85 0.94 0.61 0.98 0.23 0.74 0.85 0.36 0.79 0.51 0.86 0.03 0.88 0.76 0.22 0.26 0.97 0.36 0.8 0.54 0.86 0.03 0.88 0.76 0.22 0.26 0.97 0.36 0.8 0.54 9 0.8 0.46 0.97 0.39 0.55 0.04 0.83 0.27 0.78 0.64 / 0.89 0.74 0.35 0.29 0.58 0.45 0.54 0.98 0.22 0.31  $\infty$ 0.54 0.84 0.67 0.01 0.36 0.09 0.54 0.85 0.43 0.98 0 0 5 1 2 3 4 6 7 8 9 **Evaluation task** 

- 0.8

- 0.6

- 0.4

- 0.2