Naive 0.98 0.64 0.56 0.18 0.87 0.8 0.43 0.54 0.63 0.91 0 0.97 0.31 0.07 0.48 0.53 0.18 0.75 0.86 0.82 0.63  $\vdash$ 0.86 0.07 0.97 0.53 0.18 0.75 0.31 0.82 0.63 0.48  $\sim$ 0.26 0.77 0.97 0.32 0.42 0.75 0.85 0.5 0.87 0.35  $_{\odot}$ Training task 0.26 0.77 0.97 0.32 0.42 0.75 0.85 0.5 0.87 0.35 0.85 0.26 0.77 0.97 0.32 0.42 0.75 0.5 0.87 0.35 0.03 0.83 0.76 0.34 0.4 0.79 0.97 0.33 0.52 0.66 9 0.83 0.97 0.03 0.76 0.34 0.4 0.79 0.33 0.52 0.66 / 0.83 0.76 0.34 0.79 0.97 0.33 0.66 0.03 0.4 0.52  $\infty$ 0.65 0.89 0.65 0.73 0.88 0.02 0.71 0.14 0.93 0.83 0 5 0 1 2 3 4 6 7 8 9 **Evaluation task** 

- 0.8

- 0.6

- 0.4

- 0.2