Naive 0.98 0.64 0.87 0.8 0.43 0.56 0.54 0.63 0.91 0.18 0 0.98 0.33 0.11 0.47 0.57 0.18 0.73 0.83 0.86 0.6  $\vdash$ 0.98 0.83 0.57 0.18 0.73 0.86 0.11 0.47 0.33 0.6  $\sim$ 0.98 0.47 0.57 0.18 0.73 0.83 0.33 0.11 0.86 0.6  $_{\odot}$ Training task 0.08 0.21 0.53 0.02 0.97 0.44 0.48 0.88 0.37 0.75 0.08 0.21 0.53 0.02 0.97 0.44 0.48 0.88 0.37 0.75 0.08 0.21 0.53 0.02 0.97 0.44 0.48 0.88 0.37 0.75 9 0.89 0.35 0.91 0.96 0.42 0.02 8.0 0.21 0.72 0.62 / 0.92 0.75 0.59 0.54 0.42 0.98 0.07 0.81 0.14 0.5  $\infty$ 0.84 0.67 0.85 0.01 0.54 0.36 0.09 0.54 0.43 0.98 6 5 0 1 2 3 4 6 7 8 9 **Evaluation task** 

- 0.8

- 0.6

- 0.4

- 0.2