Naive 0.99 0.61 0.89 0.78 0.46 0.52 0.67 0.92 0.18 0.64 0 0.99 0.61 0.89 0.78 0.46 0.52 0.67 0.64 0.92 0.18  $\vdash$ 0.94 0.91 0.39 0.31 0.74 0.86 0.44 0.16 0.93 0.44  $\sim$ 0.16 0.94 0.91 0.39 0.31 0.74 0.86 0.44 0.93 0.44  $_{\odot}$ Training task 0.16 0.94 0.91 0.39 0.31 0.74 0.86 0.44 0.93 0.44 0.86 0.16 0.94 0.91 0.39 0.31 0.74 0.44 0.93 0.44 0.94 0.16 0.91 0.39 0.31 0.74 0.86 0.44 0.93 0.44 9 0.94 0.91 0.31 0.44 0.16 0.39 0.74 0.86 0.93 0.44 / 0.86 0.94 0.91 0.39 0.31 0.74 0.44 0.93 0.16 0.44  $\infty$ 0.61 0.44 0.42 0.73 0.58 0.78 0.76 0.03 0.46 0.85 0 5 0 1 2 4 6 7 8 9 3 **Evaluation task** 

- 0.8 - 0.6 - 0.4 - 0.2