Naive 1 0.010.830.950.020.80.86 1 1 0.730.030.08 1 0.980.06 0 1 0.95 1 0.7 0.030.08 1 0.980.06 0 1 <mark>0.01</mark>0.830.99<mark>0.02</mark>0.80.86 1 1 0.95 \vdash 1 0.730.030.08 1 0.980.06 0 1 0.010.830.990.020.80.86 1 1 0.95 2 -0.81 1 0.010.830.990.020.80.86 1 1 0.750.030.08 1 0.980.06 0 1 0.95 1 0.010.830.990.020.80.86 1 1 0.750.030.08 1 0.980.06 0 1 1 0.95 4 1 1 0.010.830.990.020.80.86 1 1 0.750.030.08 1 0.980.06 0 1 0.95 2 1 0.010.830.990.020.80.86 1 1 0.750.030.08 1 0.980.06 0 1 0.95 1 9 1 0.750.030.08 1 0.980.06 0 1 0.010.830.990.020.80.86 1 1 0.95 / 1 0.6 **Training task** 1 0.730.030.08 1 0.980.06 0 1 0.010.830.990.020.80.86 1 1 0.95 1 0.010.830.990.020.80.86 1 1 0.730.030.08 1 0.980.06 0 1 1 0.95 1 0.010.830.990.020.80.86 1 1 0.730.030.08 1 0.980.06 0 1 1 0.95 1 0.730.030.08 1 0.980.06 0 1 0.010.830.990.020.80.86 1 1 0.95 1 0.010.830.990.020.80.86 1 1 0.750.030.08 1 0.980.06 0 1 1 0.95 0.4 13 1 0.010.830.990.020.80.86 1 1 0.750.030.08 1 0.980.06 0 1 0.95 1 14 1 0.730.030.08 1 0.980.06 0 1 0.95 1 0.010.830.990.020.80.86 1 1 15 1 0.730.030.08 1 0.980.06 0 1 0.010.830.990.020.80.86 1 1 0.95 1 16 1 0.730.030.08 1 0.980.06 0 1 0.010.830.990.020.80.86 1 1 0.95 1 1 0.010.830.990.020.80.86 1 1 0.750.030.08 1 0.980.06 0 1 0.95 -0.218 1 0.010.830.990.020.80.86 1 1 0.750.030.08 1 0.980.06 0 1 0.95 19 1 0.90.91 1 0.720.99 1 0.980.86 1 0.990.940.910.99 1 0.980.99

Evaluation task

9 10 11 12 13 14 15 16 17 18 19

1

0

3

-0.0