Assessing the Impact of Context Inference Error and Partial Observability on RL Methods for Just-In-Time Adaptive Interventions (Supplementary Material)

Karine Karine[1] **Predrag Klasnja**[2] **Susan A. Murphy**[3] **Benjamin M. Marlin**[1]

[1]University of Massachusetts Amherst
[2]University of Michigan
[3]Harvard University

A ADDITIONAL RESULTS

A.1 ACTION SELECTION ANALYSIS FOR REINFORCE AND DQN

Figures 1 and 2 show the distribution of actions taken by the REINFORCE agent and the DQN agent. The top row of plots shows the distribution of actions selected by the agent when given access to context probabilities. The bottom row of plots shows the distribution of actions selected by the agent when given access only to the inferred most likely context. Each plot in each row corresponds to the distribution of actions in a specific range of context inference probabilities. All results are for a context inference error rate of 18%.

A.2 STATISTICAL SIGNIFICANCE OF PERFORMANCE DIFFERENCES FOR SCENARIOS P-H-D VS. L-H-D

To formally assess the differences between agents with access to $\mathbf{p}_t$ and l_t, we perform unpaired t-tests over the ten repetitions for each context inference error rate. A p-value < 0.05 indicates a statistically significant difference. The unpaired t-tests confirm that up to a context error rate of approximately 30%, access to $\mathbf{p}_t$ results in statistically significant improvements in total reward compared to access to l_t. These results are presented in Tables 1 and 2.

A.3 STATISTICAL SIGNIFICANCE OF PERFORMANCE DIFFERENCES UNDER PARTIAL OBSERVABILITY

We perform unpaired t-tests to formally contrast the DQN agent with the REINFORCE agent for each context inference error rate, under the partial observability condition. The performance differences are highly statistically significant with large differences in mean performance across all context inference error rates. These results are presented in Tables 3 and 4.

A.4 PERFORMANCE AS A FUNCTION OF DISENGAGEMENT DYNAMICS PARAMETERS

For both agents, we study how the performance of learned policies varies as a function of the disengagement increment parameter ϵ_d and disengagement decay parameter δ_d. The presented results correspond to various values of σ with habituation and disengagement observed. The results for REINFORCE are in Figure 3, and the results for DQN are in Figure 4. As we can see, these results show that the use of context inference probabilities improves on using most likely context inference over a wide range of settings of these variables. We note that he performance difference tends to be larger in cases that lead to a greater chance of disengagement events occurring. This corresponds to larger values of the disengagement risk increment parameter value ϵ_d and smaller values of the disengagement risk decay parameter value δ_d. For context inference error rates larger than 41% ($\sigma = 2$) the contrast is less apparent.

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

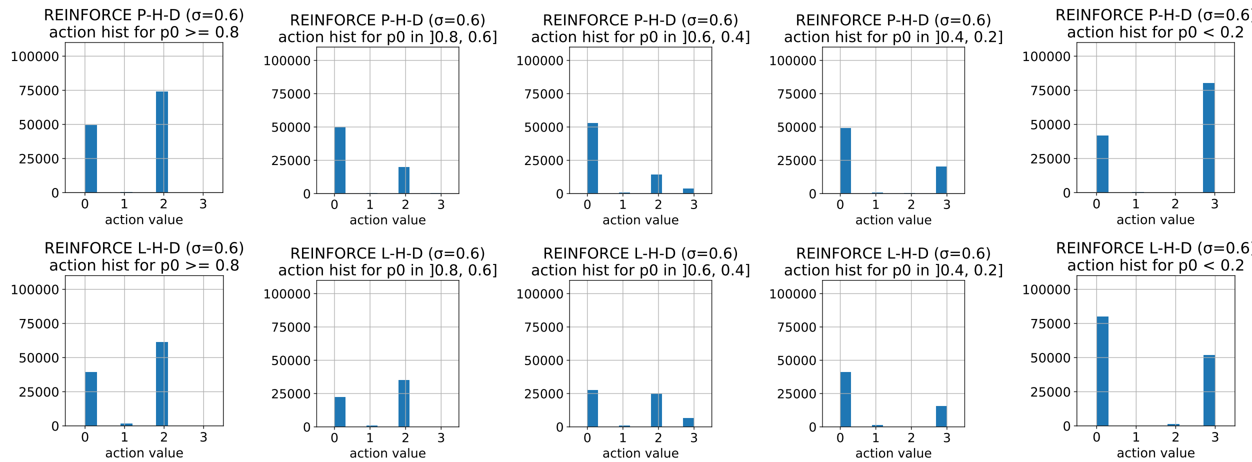

Figure 1: The top row of plots shows the distribution of actions selected by REINFORCE when given access to context probabilities. The bottom row of plots shows the distribution of actions selected by REINFORCE when given access only to the inferred most likely context.

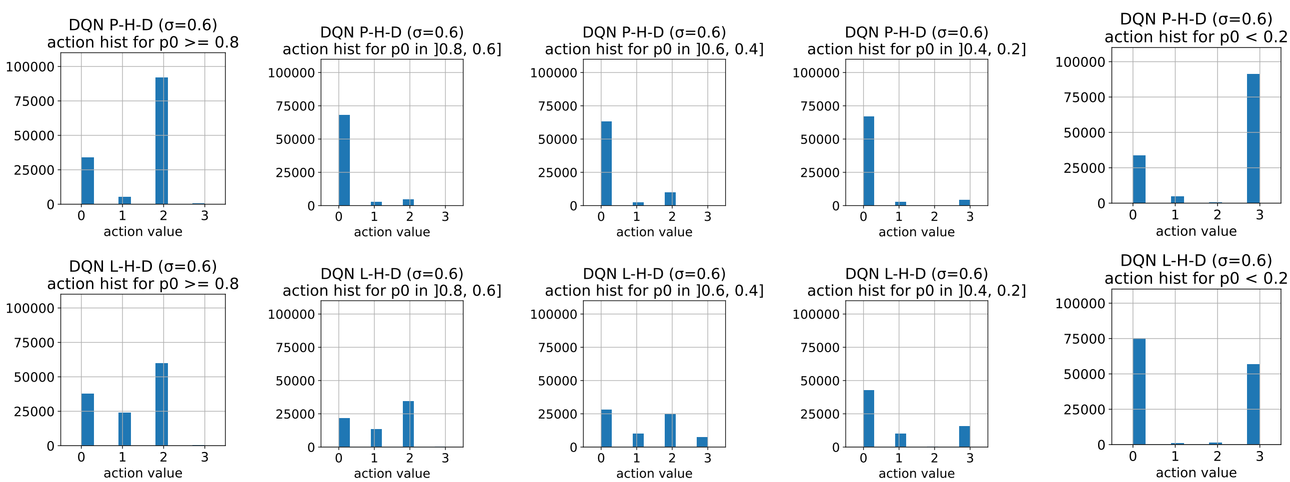

Figure 2: The top row of plots shows the distribution of actions selected by DQN when given access to context probabilities. The bottom row of plots shows the distribution of actions selected by DQN when given access only to the inferred most likely context.

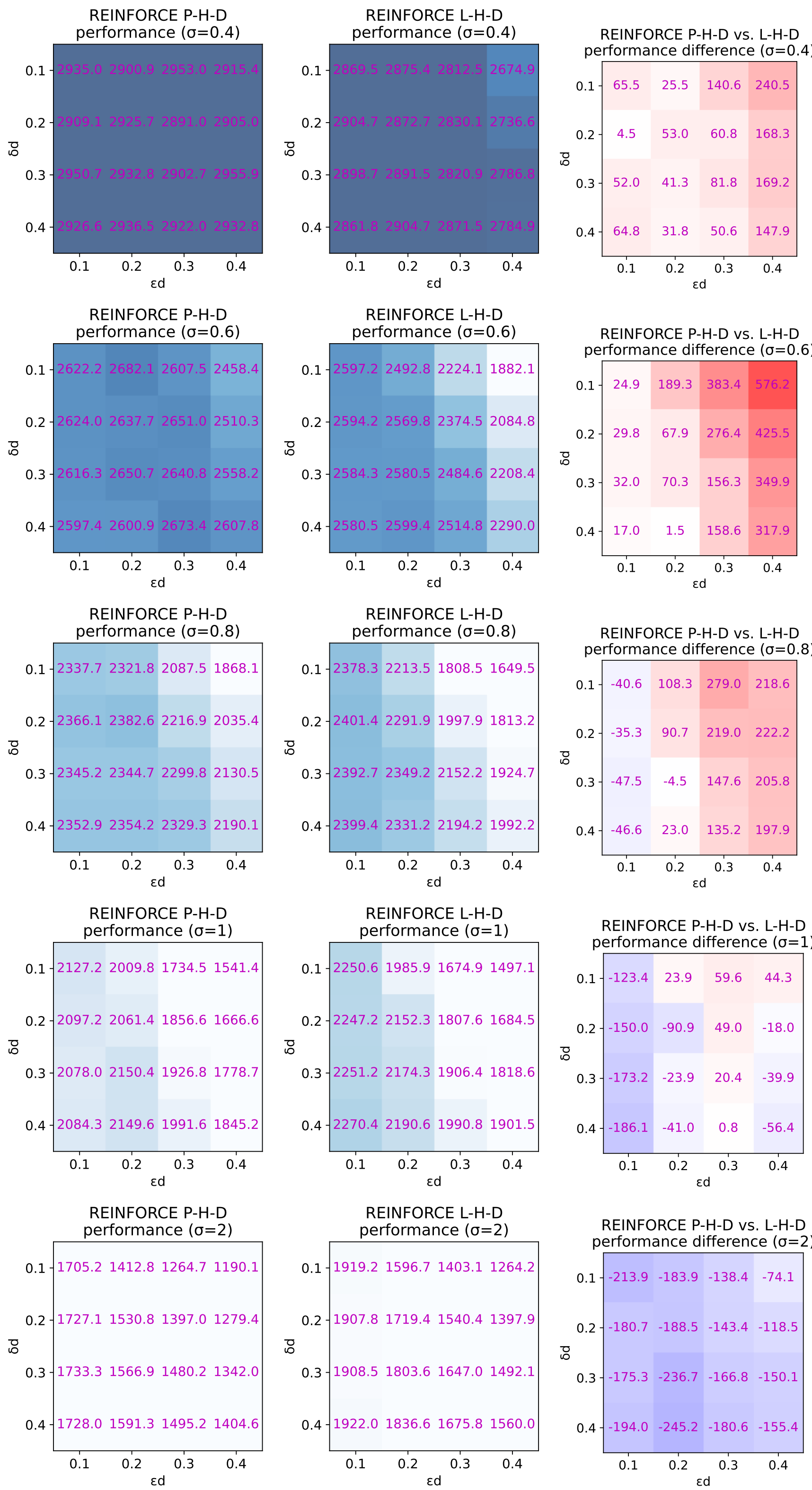

Figure 3: performance for REINFORCE, for $\sigma = 0.4, 0.6, 0.8, 1, 2$, with 10 repeats.

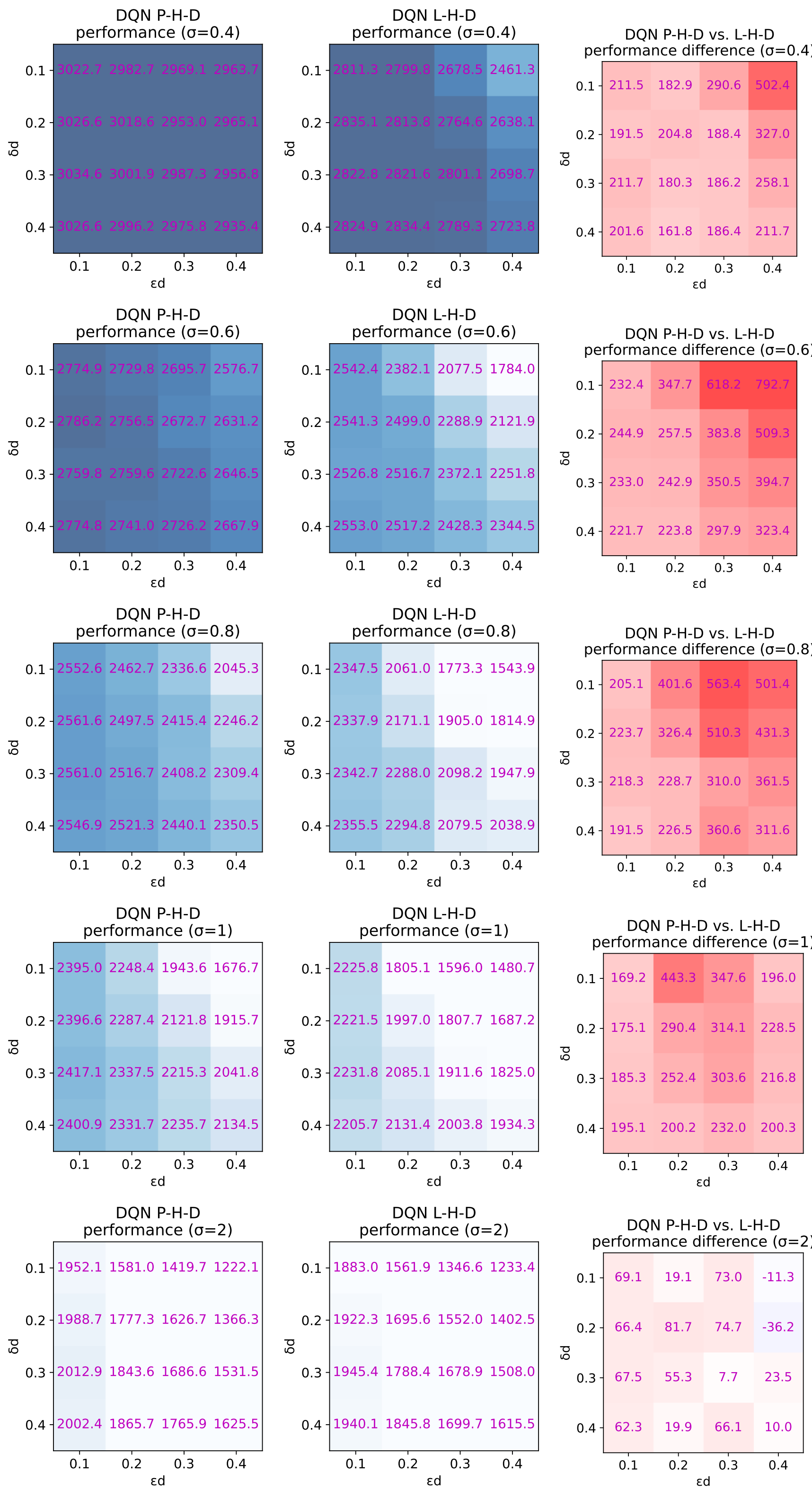

Figure 4: performance for DQN, for $\sigma = 0.4, 0.6, 0.8, 1, 2$, with 10 repeats.

Table 1: Unpaired t-tests on performance for scenarios P-H-D vs. L-H-D, for different error rates, for both agents, with $\delta_d = 0.1, \epsilon_d = 0.4$.

P-H-D vs. L-H-D	σ	**Error Rate**	**Effect**	**p-value**
DQN	0.4	10%	502.43	1.20e-19
DQN	0.6	18%	792.72	4.88e-14
DQN	0.8	27%	501.39	7.12e-08
DQN	1.0	31%	195.99	1.66e-03
DQN	2.0	41%	-11.32	7.22e-01
REINFORCE	0.4	10%	240.49	3.11e-08
REINFORCE	0.6	18%	576.24	2.37e-12
REINFORCE	0.8	27%	218.59	1.42e-07
REINFORCE	1.0	31%	44.27	2.96e-02
REINFORCE	2.0	41%	-74.13	2.91e-05

Table 2: Unpaired t-tests on performance for scenarios P-H-D vs. L-H-D, for different error rates, for both agents, with $\delta_d = 0.2, \epsilon_d = 0.3$.

P-H-D vs. L-H-D	σ	**Error Rate**	**Effect**	**p-value**
DQN	0.4	10%	188.41	3.62e-11
DQN	0.6	18%	383.83	8.77e-12
DQN	0.8	27%	510.30	1.84e-09
DQN	1.0	31%	314.10	2.16e-05
DQN	2.0	41%	74.71	1.32e-02
REINFORCE	0.4	10%	60.83	2.82e-01
REINFORCE	0.6	18%	276.41	2.73e-10
REINFORCE	0.8	27%	218.99	1.67e-06
REINFORCE	1.0	31%	48.97	2.87e-01
REINFORCE	2.0	41%	-143.41	1.75e-08

Table 3: Unpaired t-tests on performance for scenarios REINFORCE L-T vs. DQN L-T, and scenarios REINFORCE P-T vs. DQN P-T, with $\delta_d = 0.1, \epsilon_d = 0.4$.

REINF. vs. DQN	σ	**Error Rate**	**Effect**	**p-value**
L-T	0.4	10%	1261.05	2.53e-22
L-T	0.6	18%	735.05	1.66e-19
L-T	0.8	27%	495.52	3.67e-18
L-T	1.0	31%	439.07	4.39e-14
L-T	2.0	41%	481.95	6.48e-19
P-T	0.4	10%	1586.30	6.47e-25
P-T	0.6	18%	1354.90	4.44e-21
P-T	0.8	27%	887.54	2.37e-16
P-T	1.0	31%	559.32	3.73e-11
P-T	2.0	41%	378.99	3.06e-15

Table 4: Unpaired t-tests on performance for scenarios REINFORCE L-T vs. DQN L-T, and scenarios REINFORCE P-T vs. DQN P-T, with $\delta_d = 0.2, \epsilon_d = 0.3$.

REINF. vs. DQN	σ	**Error Rate**	**Effect**	**p-value**
L-T	0.4	10%	1336.32	6.87e-17
L-T	0.6	18%	1139.26	1.88e-26
L-T	0.8	27%	872.99	1.25e-23
L-T	1.0	31%	718.74	4.45e-18
L-T	2.0	41%	630.50	1.20e-18
P-T	0.4	10%	1420.69	3.82e-17
P-T	0.6	18%	1384.04	1.88e-19
P-T	0.8	27%	1145.62	7.01e-23
P-T	1.0	31%	881.51	5.58e-18
P-T	2.0	41%	463.40	2.10e-12