SUPPLEMENTARY MATERIALS

We have included several plots in these materials that while not critical to our work, provide a more complete picture of the data and analyses involved.

Histograms from Section III-A, Synthesizing from Small n

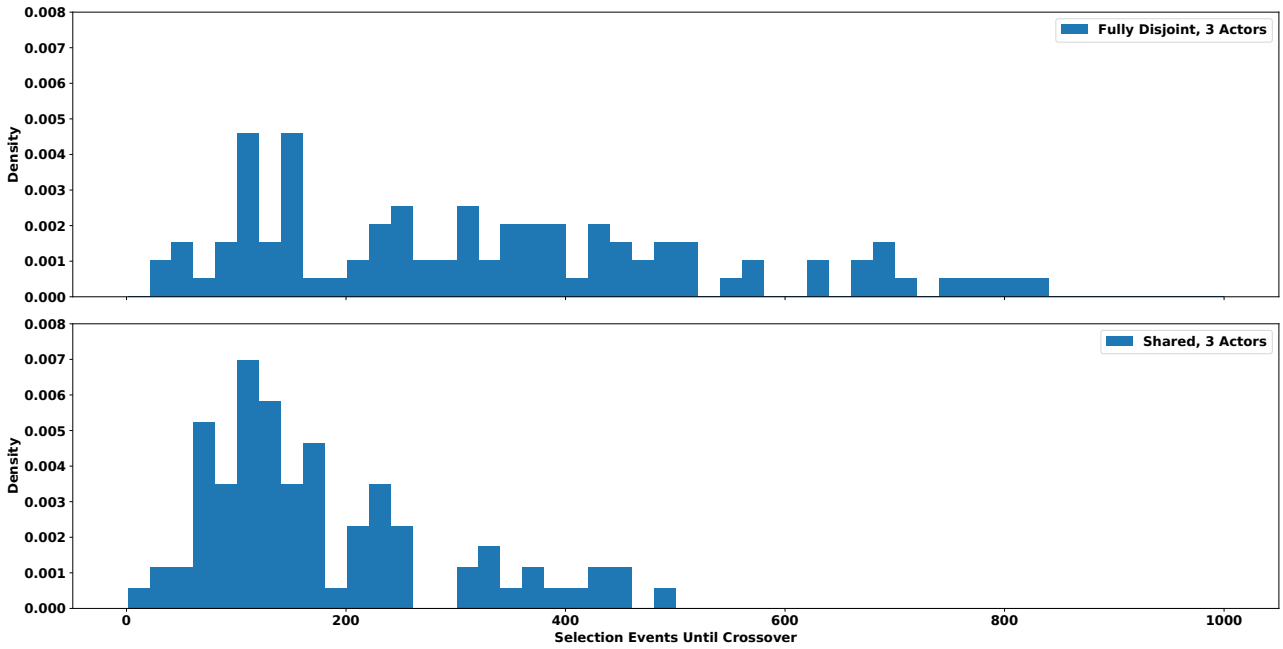

Fig. 7: Distribution of crossover points for 3 actors

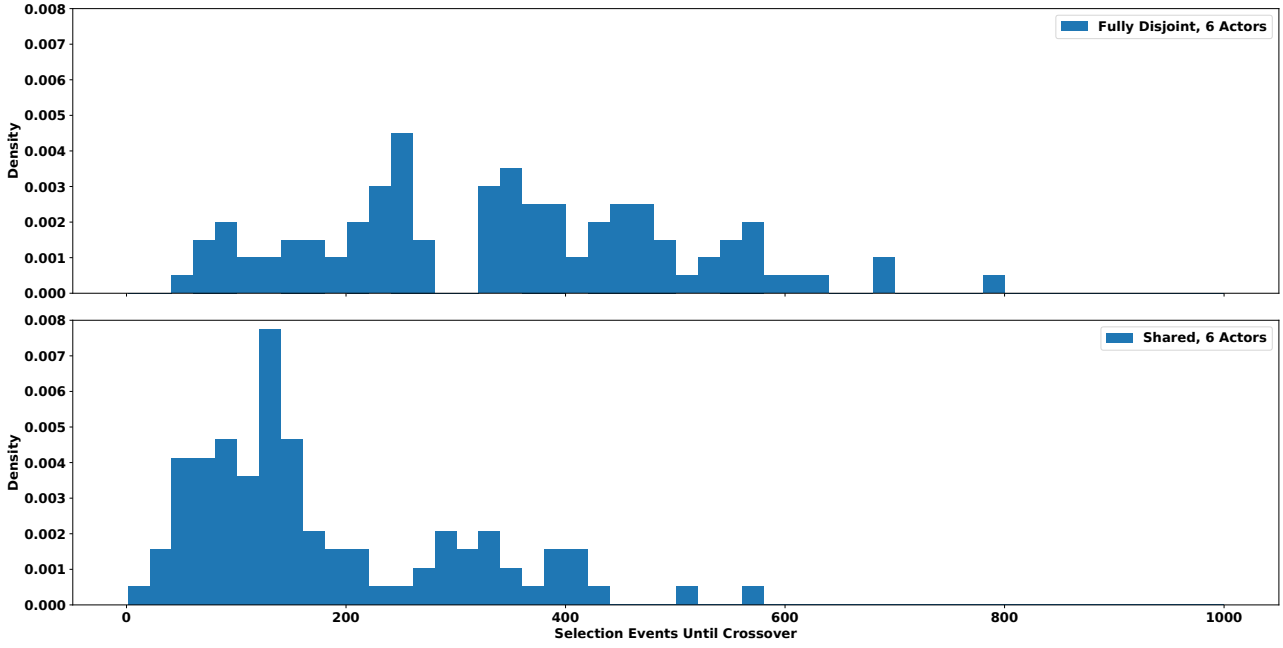

Fig. 8: Distribution of crossover points for 6 actors

Histograms from Section IV: Assessment of a reinforcement-learning adaptive policy using human-sourced data

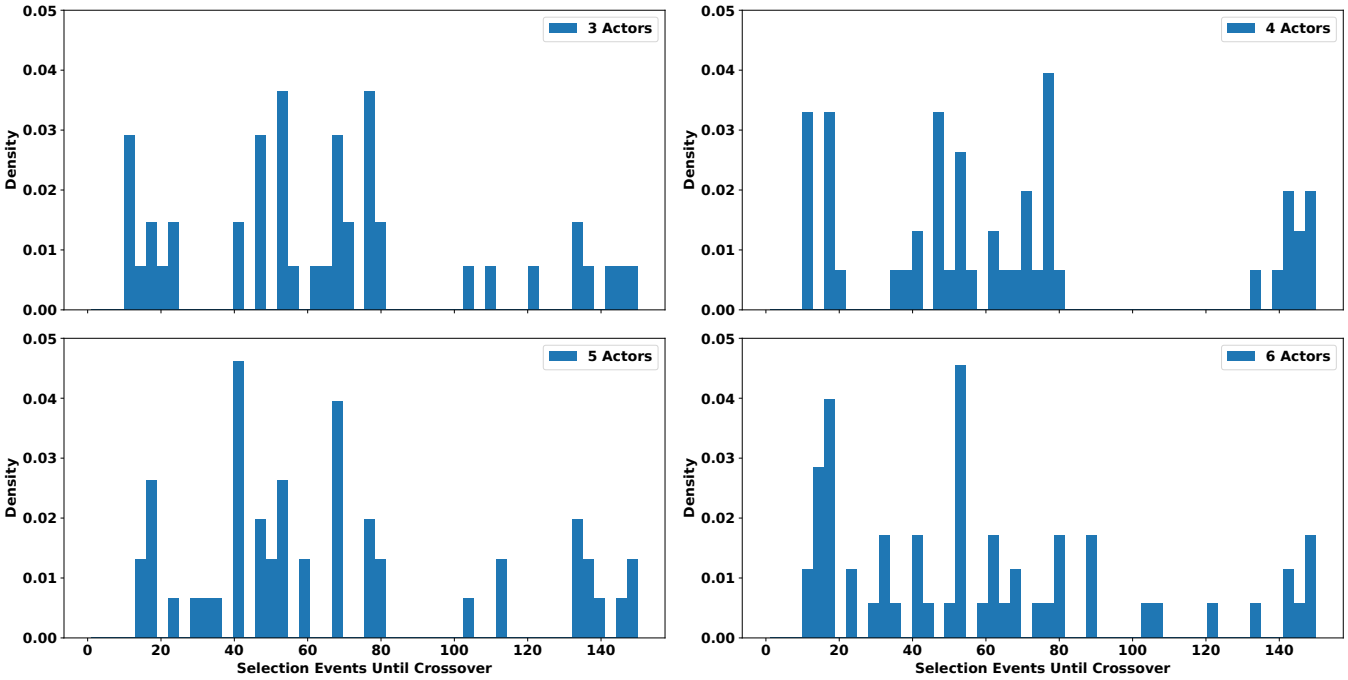

Fig. 9: Histograms of crossover points, showing the relative density of the number of selection events needed before a bandit policy outperforms an informed static policy

Cumulative Regret plots from Section IV: Assessment of a reinforcement-learning adaptive policy using human-sourced data

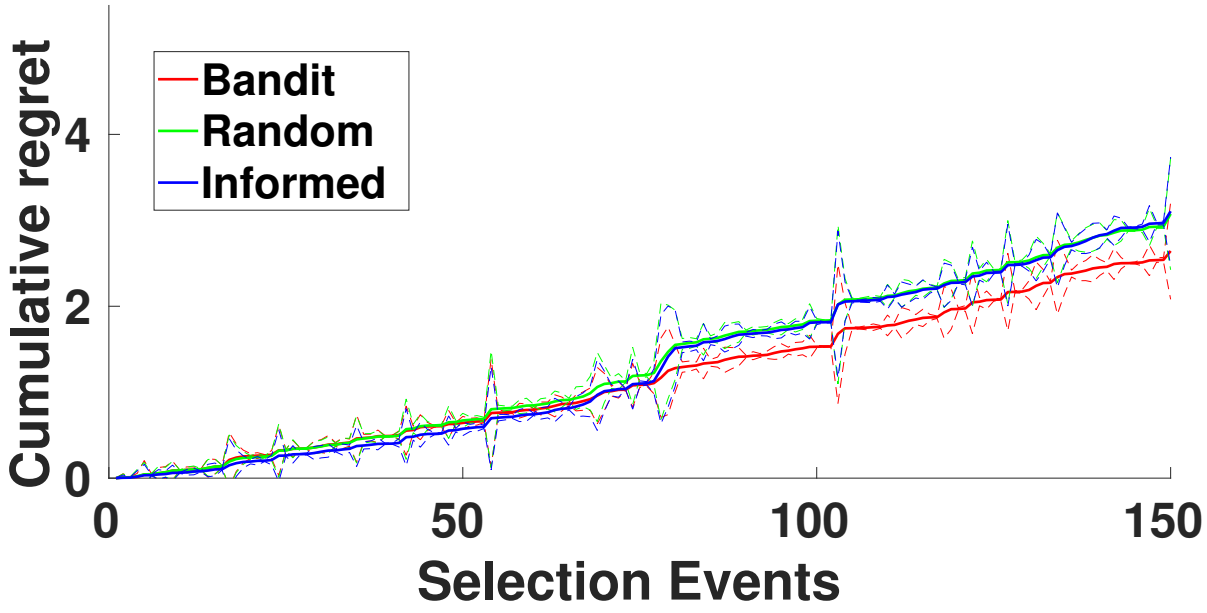

Fig. 10: Plots of cumulative regret for 3 actors; $\pm 5\sigma$ bounds are shown in dashed lines

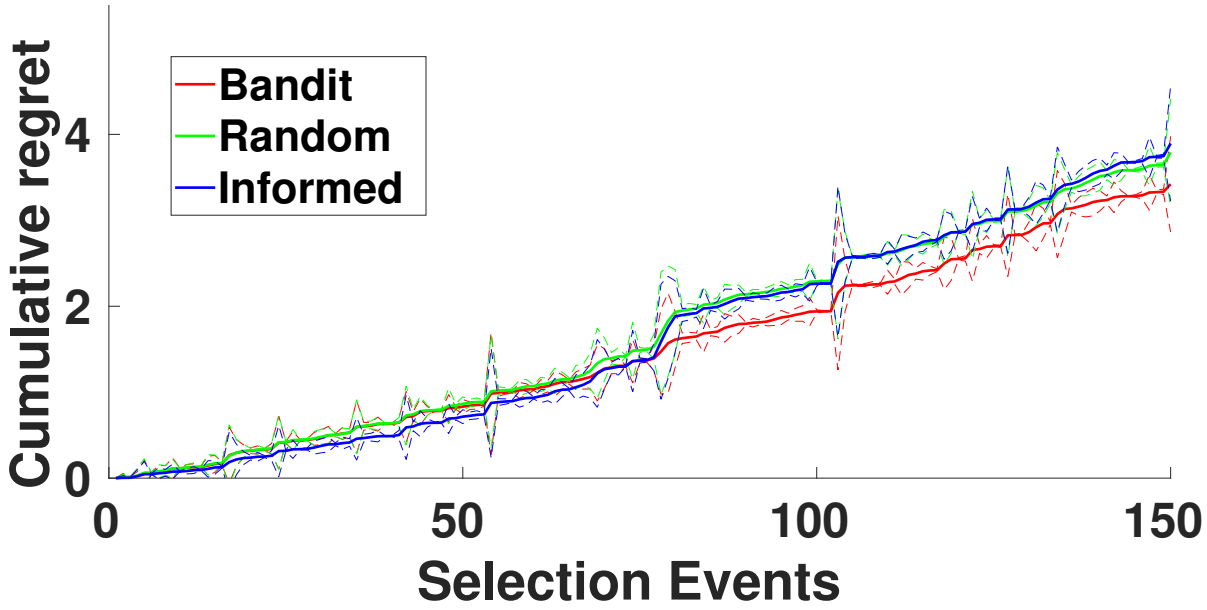

Fig. 11: Plots of cumulative regret for 4 actors; $\pm 5\sigma$ bounds are shown in dashed lines

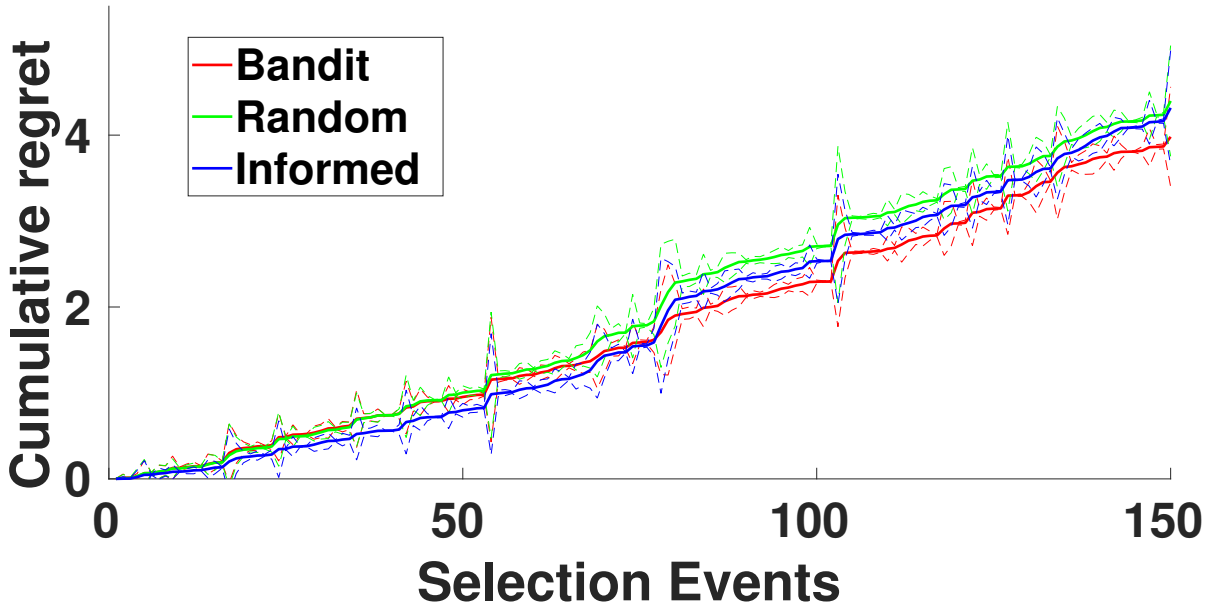

Fig. 12: Plots of cumulative regret for 5 actors; $\pm 5\sigma$ bounds are shown in dashed lines

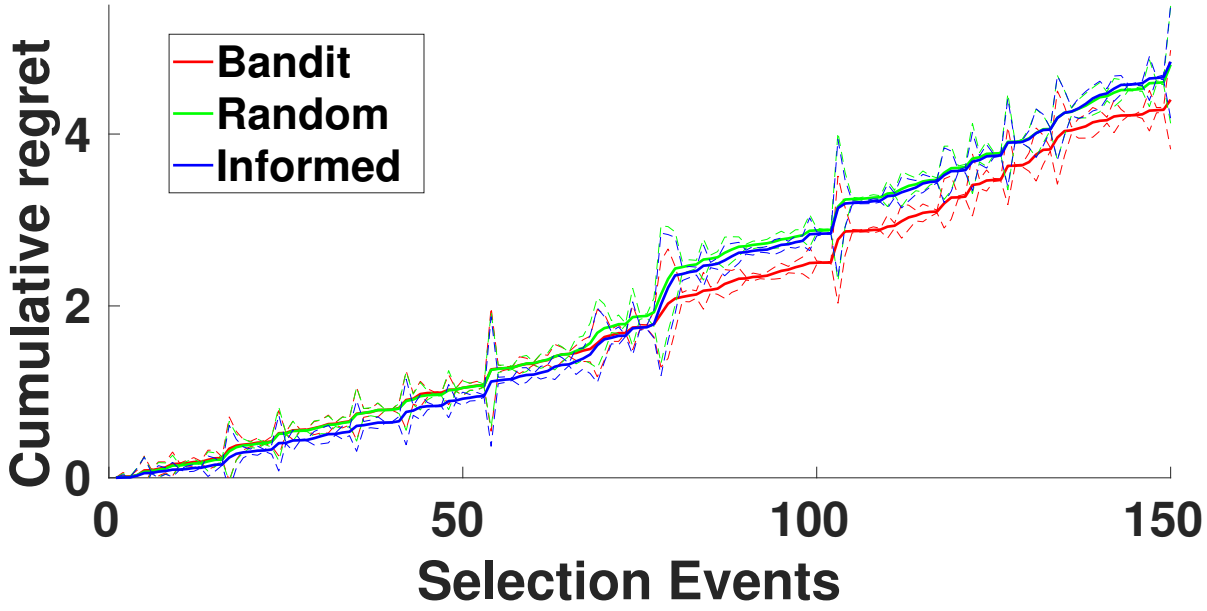

Fig. 13: Plots of cumulative regret for 6 actors; $\pm 5\sigma$ bounds are shown in dashed lines

Analysis of Deviance Table (Type II Wald χ^2 tests)
Model: $Productivity \sim Workload * Attention * GSR * HR + (1 + TaskType/TaskDifficulty) + (1|Actor)$

Response: Productivity	χ^2	Df	Pr($>\chi^2$)
Workload	4.9634	6	0.548510
Attention	15.8703	6	0.014468
GSR	0.0079	1	0.929395
HR	5.3067	1	0.021244
TaskType	1789.4987	2	$<$ 2.2e-16
Workload:Attention	14.7087	4	0.005345
Workload:GSR	2.7055	2	0.258528
Attention:GSR	5.3802	2	0.067874
Workload:HR	2.3334	2	0.311388
Attention:HR	1.0852	2	0.581230
GSR:HR	0.2499	1	0.617142
TaskType:TaskDifficulty	1037.6611	6	$<$ 2.2e-16
Workload:Attention:GSR	2.1026	4	0.716899
Workload:Attention:HR	4.6067	4	0.330079
Workload:GSR:HR	2.0478	2	0.359183
Attention:GSR:HR	3.2907	2	0.192945
Workload:Attention:GSR:HR	3.7095	4	0.446752

TABLE II: Type II Wald Test Results. Rows indicating observable signals of interest are bolded, while rows indicating experimental controls are italicized. Heartrate (HR) alone is a significant predictors of performance, while neither galvanic skin response (GSR) nor GSR with HR are significant predictors of performance. The prediction of the italicized control variables indicates that our experimental controls were effective in manipulating performance.