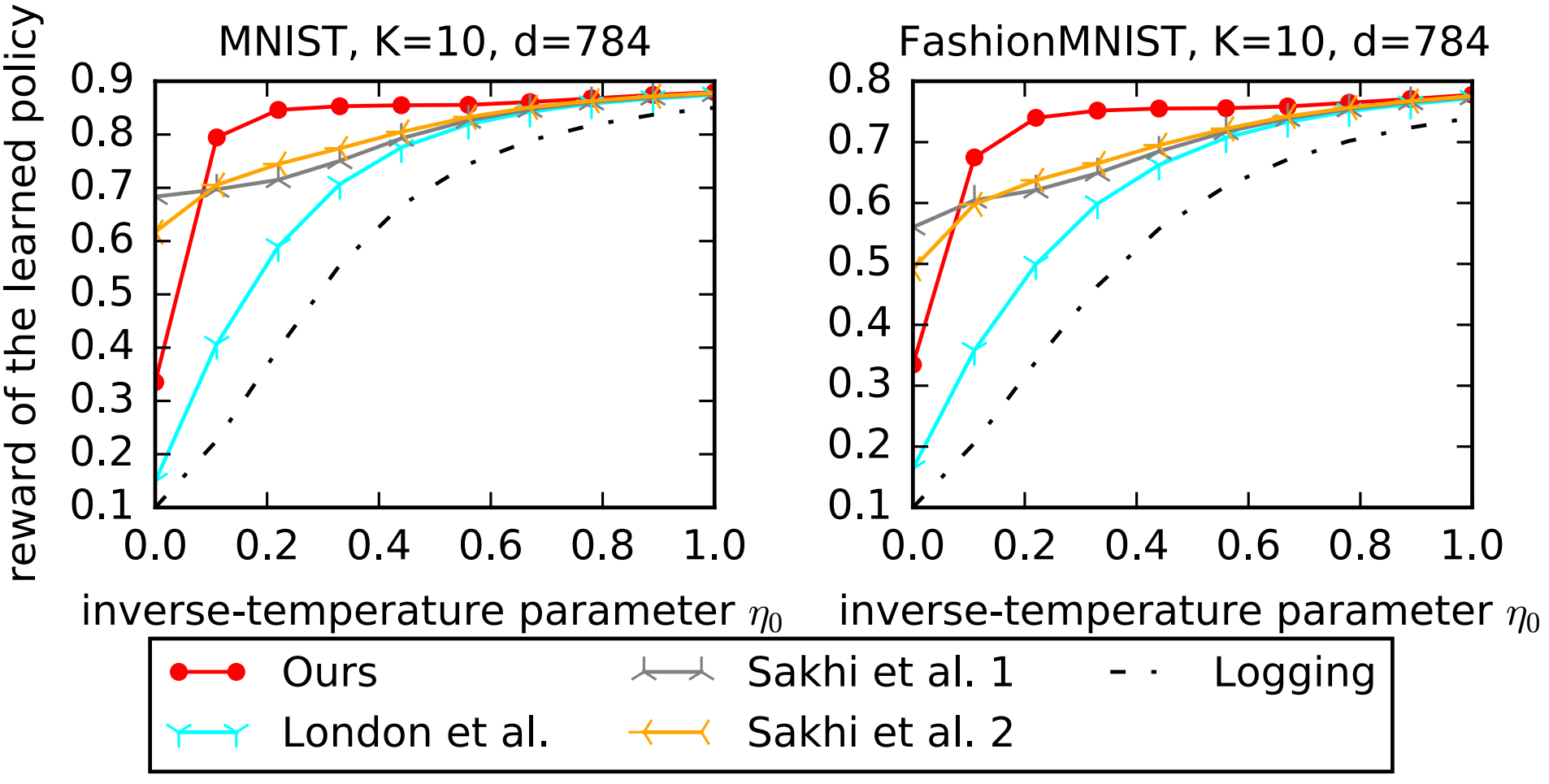
MNIST, K=10, d=784
FashionMNIST, K=10, d=784
reward of the learned policy
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.2
0.4
0.6
0.8
1.0
inverse-temperature parameter η_0
inverse-temperature parameter η_0
Ours
London et al.
Sakhi et al. 1
Sakhi et al. 2
Logging