## Supplementary Material for T. Osogami, *Robust partially observable Markov decision processes*

The robust HSVI finds the robust policy optimized for the worst case in $0 \leq p_e \leq \hat{p}_e/\kappa = 0.5$ (i.e., $\kappa = 2$). The HSVI finds the "optimal" policy with the assumption of $p_e = \hat{p}_e = 0.1$. These algorithms are implemented with Python and run on a single core of a 2.6 GHz processor. Because the purpose of our experiments is to study the general characteristics of the robust HSVI, neither the robust HSVI nor the HSVI is optimized for this particular instance.

Figure 2 (a) shows the discounted cumulative reward (here, also denoted as performance ) that is obtained by the robust policy (solid line) or the optimal policy (dashed line) against the true $p_e$. Each data point shows the average and the (small) standard deviation over 30 trials, each consisting of 1,000 runs of simulation. The optimal policy has higher performance (up to 3%) than the robust policy when the true $p_e$ is around $0.1$, for which the optimal policy is optimized. The performance of the optimal policy quickly diminishes as the $p_e$ deviates from $0.1$. For $p_e > 0.3$, the robust policy has higher performance than the optimal policy. Although the value that gives the worst case is generally not known in advance, we can infer, for this instance, that $p_e = 0.5$ gives the worst case in the uncertainty set, $[0, 0.5]$. When the true $p_e$ is $0.5$, the robust policy has 32% higher performance than the optimal policy.

Figure 2 (b) shows how the bounds on the (robust) value function at $\mathbf{b}_0$ are updated with the robust HSVI (solid line) or the HSVI (dashed line). The HSVI stops at 10.1 seconds when the lower bound coincides with the upper bound. The robust HSVI stops at 23.1 seconds when it completes the exhaustive search with a given error bound of $10^{-6}$. Recall from Section 3.3 that the robust HSVI uses the nominal values in constructing the upper bound, so that the upper bound is valid but, in general, not tight. Giving tight upper bounds is an important future work. Also, there exist many techniques (not used in our experiments) to improve the efficiency of the HSVI, and tailoring these techniques for the robust HSVI is an interesting direction of future work.

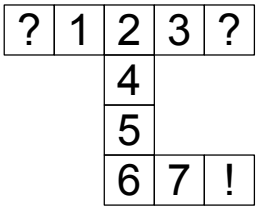
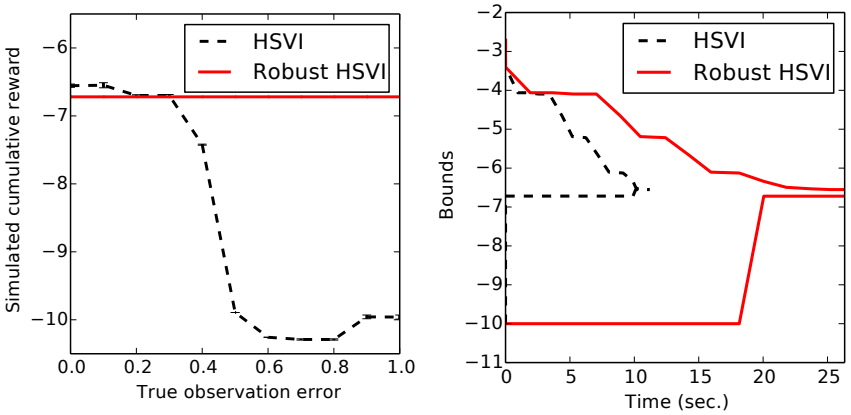



*Figure 1.* Heaven & Hell




(a) Performance (b) Bounds

*Figure 2.* Progress of the bounds updated with robust HSVI