APPENDIX

A EXPERIMENTAL DETAILS

In this appendix, we describe the technical details of our experiments (run using the code at https://github.com/praal/beconsiderate). All environments are deterministic, and models are trained using Q-Learning and the ϵ-greedy algorithm is used to balance between exploration and exploitation. Experiments are all done on an AMD Ryzen Threadripper 2990WX with 128 GB of RAM, and the training time is measured on the same machine. Each experiment is repeated 10 times. In all the experiments $\alpha_1 = 1$, $\gamma = 1$ and the learning rate is 1.

Table 1 provides details of our setup. The top two entries pertain to the quantitative experiments in the main body of the paper. The third and fourth entries refer to the qualitative experiments illustrating different notions of augmenting the reward function and considerations for different agents. Finally, the last entry refers to the options formulation in subsection 3.3. Our setup for all of our experiments assumes that agents, other than the acting agent, are executing fixed policies (resp. options). In the options case, the actual option policies did not need to be defined and we simply encoded the initiation sets for each of those options. In all other cases, the fixed policies of the "other agents" were learned. As such in Table 1, where relevant, the column describing Training Steps distinguishes between the training steps for "acting agent" and "others". The training steps for "others" (the other agents) is done in advance of training the acting agent and serves to establish the fixed policies of those agents and to populate our distribution of value functions. The training steps for the acting agent reflects the training steps for our approach. For the next experiment (Figure 2), the model is trained 700 times by changing α_2 from 0 to 7.0 with steps of 0.01. Similarly, in the experiments that follow, the models are trained by setting α_2 to three different values.

Table 1: Training steps, running time and hyperparameters of the experiments reported in the paper

Experiment	ϵ	Training Steps	Training Time (secs)
Table 1	0.2	$8 \times 2 \times 10^5$ (acting agent), 8×10^5 (others)	37.64 ± 0.06
Figure 2	0.5	$700 \times 7 \times 10^5$ (acting agent), 4×10^5 (others)	9332.86 ± 22.65
Figure 3	0.2	$4 \times 2 \times 10^5$ (acting agent), 10×10^5 (others)	29.72 ± 0.08
Figure 4	0.2	$3 \times 2 \times 10^5$ (acting agent), 3×10^5 (others)	14.02 ± 0.12
Figure 5	0.2	$3 \times 2 \times 10^5$ (acting agent)	9.30 ± 0.12