A Appendix

Table 1 summarizes all the hyper-parameters used in the implementation of the baselines as well as the Deep RePReL agent.

Hyperparameters	Values	Exceptions
Learning Rate	0.003	–
Batch Size	128	–
Input Size (State Dimension) in Taxi	91	69 for RePReL abstract state
Input Size (State Dimension) in Office	11	4 for option pickup and 5 for option drop in RePReL
Output Size (# of Actions) in Taxi	6	2 for HDQN meta-controller
Output Size (# of Actions) in Office	4	3 for HDQN meta-controller
Hidden layers	2	–
Hidden units	256	
Discount Rate	0.99	–
Max Episode Length	150 for *task* 1, 200 for *task* 2, and 300 for *task* 3 of Taxi Domain; 1000 for both tasks of office World	
Intrinsic reward on subgoals	30	no intrinsic reward in DQN
Max buffer size	$1e5$	–
Exploration Epsilon	Fixed 0.1 for Taxi; Decay from 1 to 0.1 with rate in Office World	
Max steps in training environment	$1e6$	–

Table 1: Summary of hyperparameters used