In the fully observed case, the difference between BC (~1430) and DAgger (~2267) rewards is due to the fact that during behavior cloning, sometimes the learner’s action takes it to a state where no expert action has been documented (aka the training observations are different from actual interactions). Meanwhile, DAgger by querying the expert in these “learner-reached” edge cases learns more about what the best action to take when reaching these states. This difference between BC (~100s) and DAgger (~50) rewards remains true, but to a smaller extent, in the partially observable environment as both have more trouble imitating the expert when missing an information dimension in the observation space. Again, even being able to observe this partial information in the “learner-reached” edge cases helps DAgger handle them slightly better even when it is now tough to distinguish between different states with missing information. Behavior Cloning meanwhile struggles with the dual problem of having less information andreaching states that don't appear in expert observation which *compounds* as these edge-case states are reached more often with the missing information (edge cases occur more frequently leading even worse performance).