



Figure R.1: MiniGrid reward performance for 2 agents over 1000 epochs, with a maximum step limit of 50.