

DETERMINING ACTION SELECTION POLICIES OF AN EXECUTION DEVICE**ABSTRACT**

Disclosed herein are methods, systems, and apparatus of an execution device for generating an action selection policy for completing a task in an environment that includes the execution device and one or more other devices. One method includes: in a current iteration, identifying an iterative action selection policy of an action in a state of the execution device in a previous iteration; computing a regret value in the previous iteration based on the iterative action selection policy in the previous iteration; computing an incremental action selection policy in the current iteration based on the regret value in the previous iteration but not any regret value in any iteration prior to the previous iteration; computing an iterative action selection policy in the current iteration based on the iterative action selection policy in the previous iteration and the incremental action selection policy in the current iteration.

FIG. 2B