[0004a] According to one embodiment, there is provided a computer-implemented method 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. The method includes: at each of a plurality of iterations and for each action among a plurality of possible actions in a state of the execution device in a current iteration, wherein the state of the execution device results from a history of actions taken by the execution device, obtaining a regret value of the action in the state of the execution device in a previous iteration, wherein the regret value of the action in the state of the execution device represents a difference between a gain of the execution device after taking the action in the state and a gain of the execution device in the state; and computing a parameterized regret value of the action in the state of the execution device in the previous iteration including: determining a maximum of a nonnegative flooring cutoff regret value and the regret value of the action in the state of the execution device in the previous iteration, and computing the parameterized regret value by raising the determined maximum to the power of β, where β is a fixed value that is larger than 1; determining a respective normalized regret value for each of the plurality of possible actions in the previous iteration from parameterized regret values for the plurality of possible actions in the state of the execution device in the previous iteration; determining, from the normalized regret values, a parameterized action selection policy of the action in the state of the execution device; determining, from the parameterized action selection policy of the action in the state of the execution device, an action selection policy of the action in the state of the execution device, wherein the action selection policy specifies a probability of selecting the state of the plurality of possible actions; and controlling operations of the execution device according to the action selection policy.

[0005] This specification also describes one or more non-transitory computer-readable storage media, coupled to one or more processors and having instructions stored thereon which, when executed by the one or more processors, cause the one or more processors to perform operations in accordance with embodiments of the methods provided herein.

[0006] This specification further describes a system for implementing the methods described herein. The system includes one or more processors, and a computer-readable storage medium coupled to the one or more processors having instructions stored thereon

which, when executed by the one or more processors, cause the one or more processors to perform operations in accordance with embodiments of the methods provided herein.

[0007] Methods, systems, and computer media in accordance with this specification may include any combination of the aspects and features described herein. That is, methods in accordance with this specification are not limited to the combinations of aspects and features specifically described herein, but also include any combination of the aspects and features described.

[0008] The details of one or more embodiments of this specification are set forth in the accompanying drawings and the description below. Other features and advantages of this specification will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. 1 is a diagram illustrating examples of partial game trees in one-card poker, in accordance with embodiments of this specification.

[0010] FIG. 2A is a diagram illustrating an example of a workflow of original CFR and streamline CFR, and FIG. 2B illustrates an example of a workflow of streamline CFR, in accordance with embodiments of this specification.

[0011] FIG. 3 is a pseudocode of an example of a streamline CFR algorithm, in accordance with embodiments of this specification.

[0012] FIG. 4 is a flowchart of an example of a process for performing a streamline CFR for determining action selection policies for software applications, in accordance with embodiments of this specification.