**INITIALIZE** for all values in Q-value table # Unbiased or biased w/ prior

**INITIALIZE** # Time and convergence counter

**SET** hyper-parameter # Exploration and convergence

**FOR** iterations in budget ***:*** # Number times we probe user

**INCREMENT**  # Increment time step counter

**CALCULATE** uncertainty for each action # See Equation 1

**SELECT** action with max value = # See Equation 2

**INCREMENT** # Increment action a counter

**EXECUTE** action # Present sound to user

**PROBE**  user for feedback and # Ask users two questions

**CALCULATE** reward signal # See Equation 3

**UPDATE**  # See Equation 4

**CALCULATE** # See Equation 5

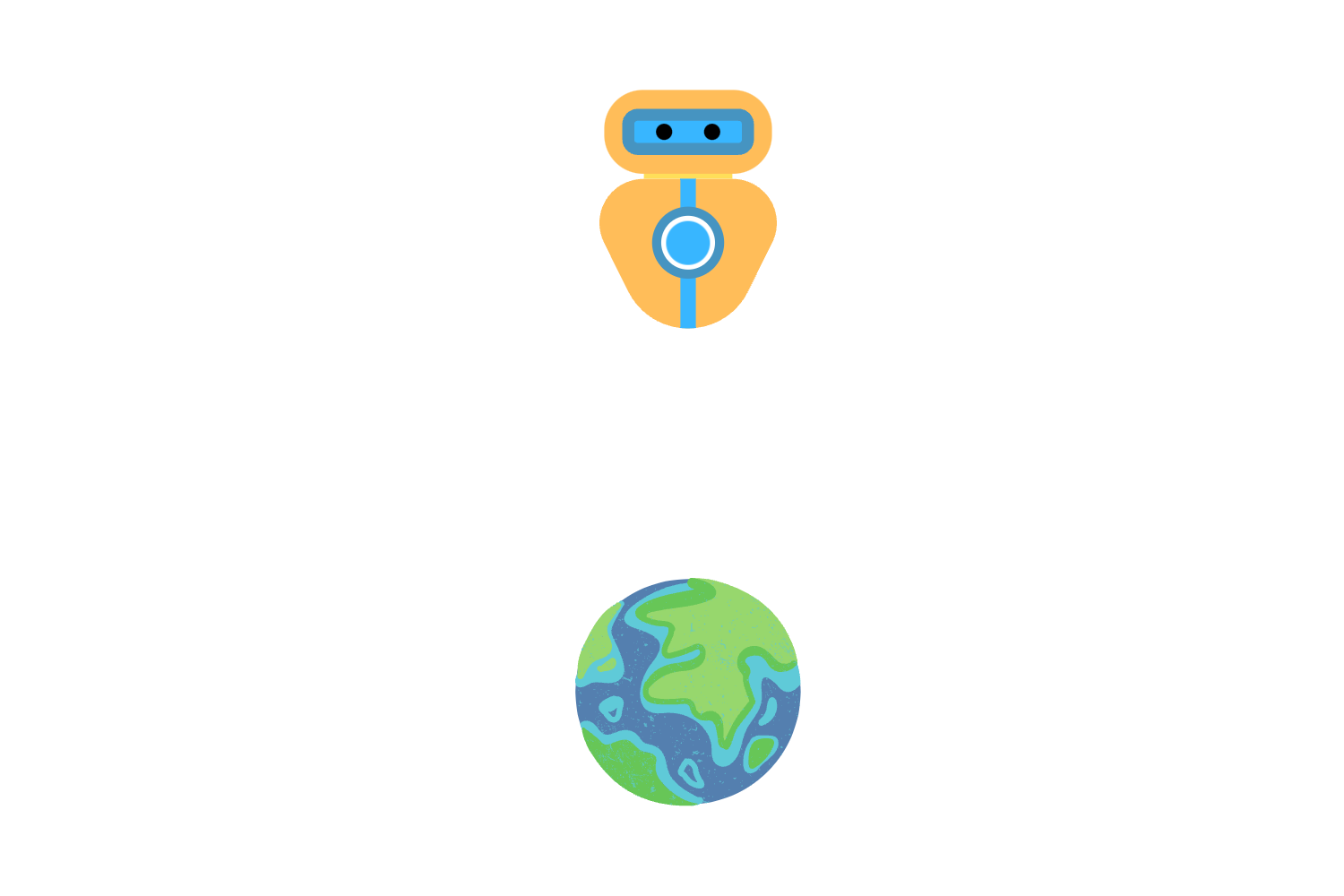
**IF**  # See Equation 6

**ELSE**  **:**

**IF :** # Check if the state has reached convergence

**BREAK**

**END** loop

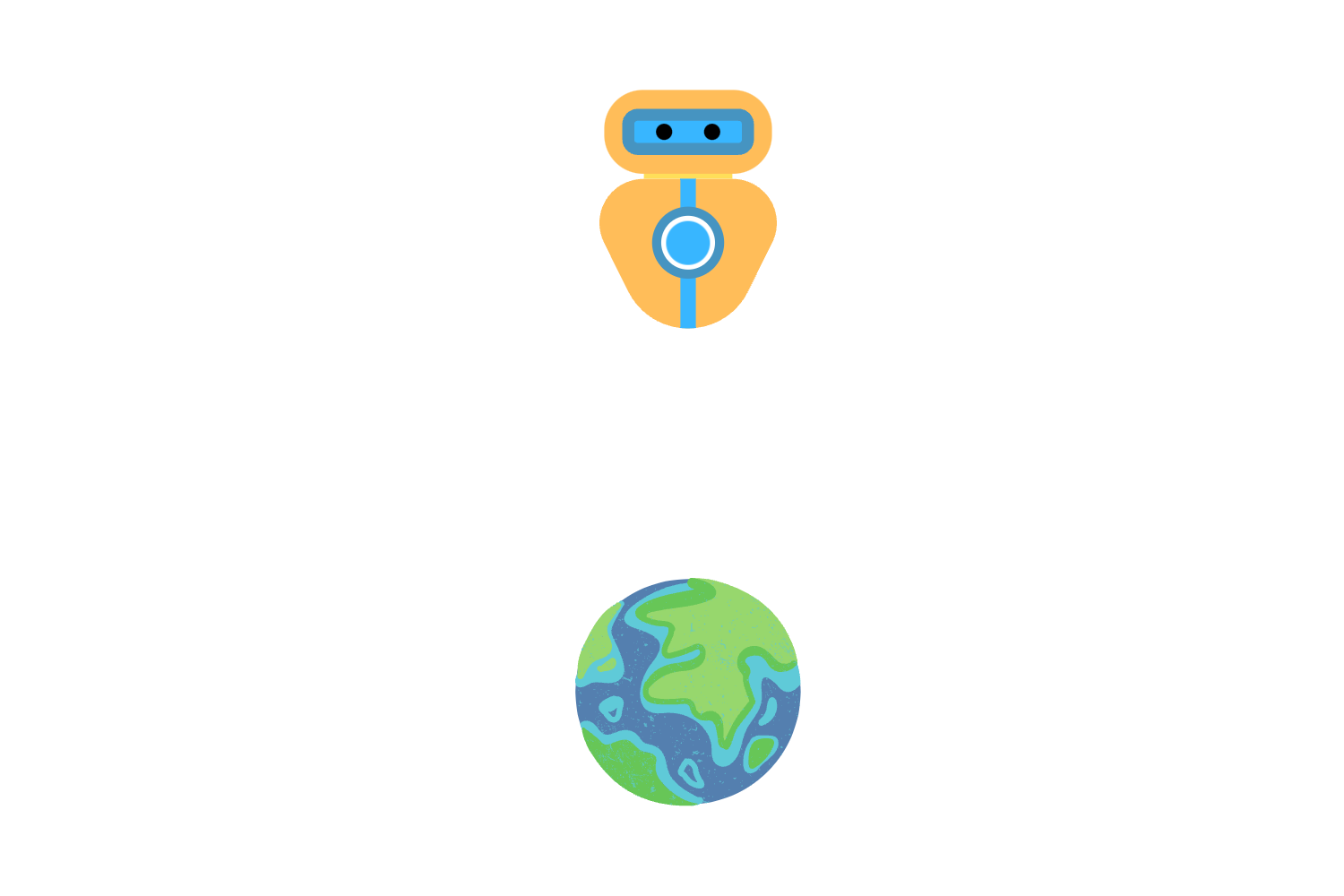
****

Agent

Observation of state

Reward

Action

****

Environment

Initialize Q-value table and hyper-parameters

Present sound to user and probe user for feedback and

Calculate uncertainty for each action

Select sound using

Terminate

Calculate change in

Calculate reward andupdate Q-value table with

Increment budget

Increment timestep

NO

YES

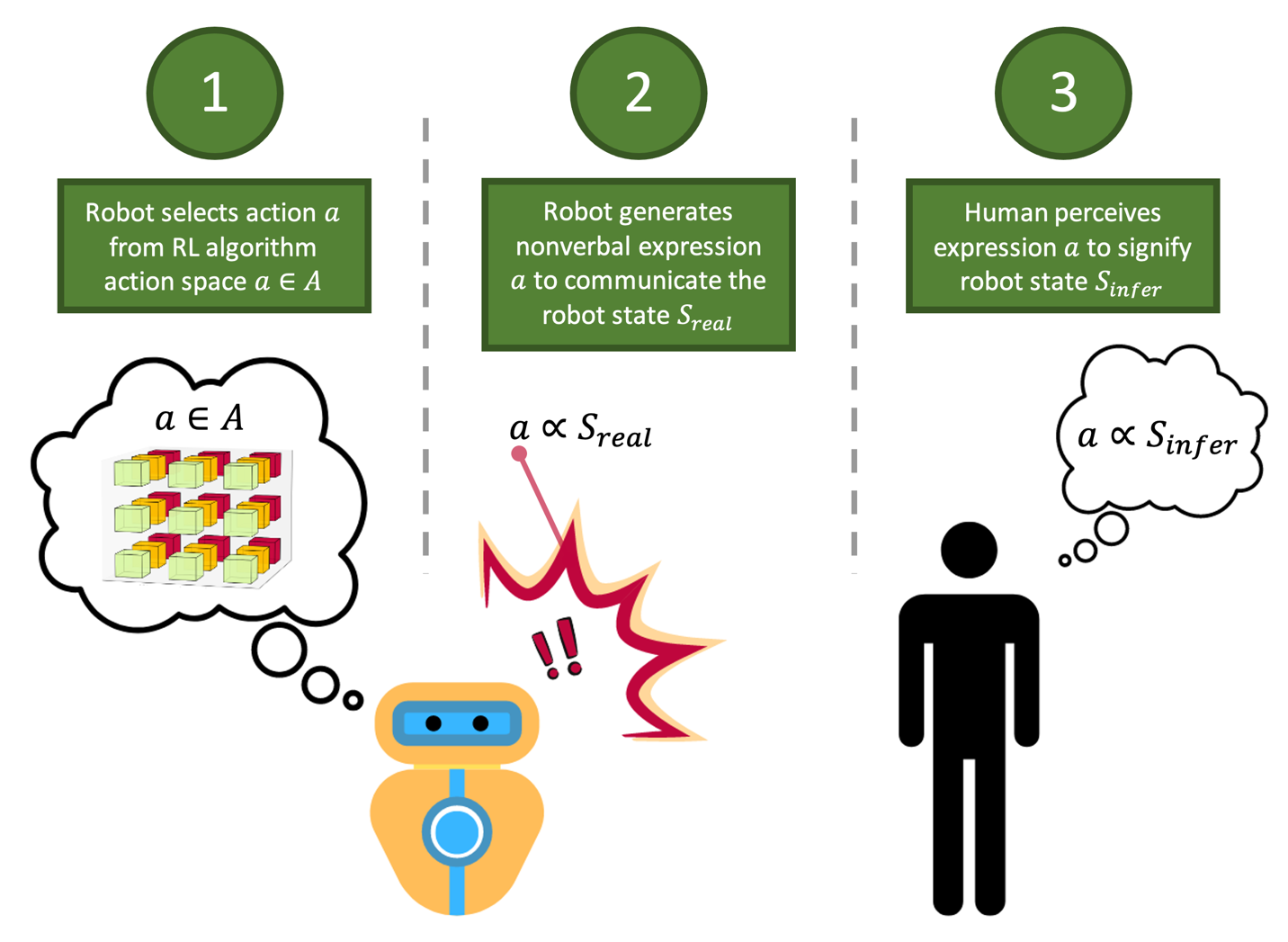
Remaining budget ?

Converged?

NO

YES

Check for convergence using

****

1

Robot selects action from RL algorithm action space

2

Robot generates nonverbal expression to communicate the robot state

3

Robot selects action from RL algorithm action space

3

Robot selects action from RL algorithm action space

**A person and a robot with speech bubbles

Description automatically generated­**