Stand
Clap
Wave
$\alpha = 0.5, \gamma = 1$

Diagram:

- Stand
- Clap
- Wave

Table:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>0.5</th>
<th>0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>7.5</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
$\alpha = .5, \gamma = 1$

Stand

Clap

Wave
$\chi = .5, \; Y = 1$

- Stand
- Clap
- Wave
\[ \alpha = .5, \quad \gamma = 1 \]

- **Stand**
- **Clap**
- **Wave**

\[ S_{i-1} \]

\[ c_{i,0} \]

\[ c_{i,10} \]

\[ w, 0 \]

\[ \text{Dyna-Q learns } Q \text{ values so } (\uparrow, 9) \text{ and } (\downarrow, 10) \]
$\alpha = .5, \gamma = 1$

Stand

Clap

Wave

\begin{tabular}{|c|c|c|c|}
\hline
0 & 0 & .5 & -7.5 \\
\hline
0 & 0 & 1 & 7.5 \\
\hline
0 & 0 & 0 & 0 \\
\hline
\end{tabular}
- What about unvisited states?
- What if transition function were stochastic?
- What about unvisited states?
- What if transition function were stochastic?
- Does the order of updates matter?
\( \alpha = .5, \beta = 1 \)

- What about unvisited states?
- What if transition function were stochastic?
- Does the order of updates matter?
- What about unvisited states?
- What if transition function were stochastic?
- Does the order of updates matter?
  - prioritized sweeping
  - could actually do update in step (e)
\( \alpha = .5, \gamma = 1 \)

Trajectory sampling: how differs from Dyna here?
$\alpha = .5, \gamma = 1$

Trajectory sampling: how differs from Dyna here?

How does Dyna-$Q^+$ differ from Dyna here?

- relationship to UCB?
- In what way does Dyna-$Q^+$ violate the author's principles?
- Could you accomplish something similar w/out changing updates?

(see ex. 8.4)
MCTS: Monte Carlo Tree Search - Planning at decision time
MCTS

Stand

Clap

Wave

+1

-3

-3

+2

+10

+11
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td></td>
<td>-3</td>
</tr>
<tr>
<td>-3</td>
<td></td>
<td>+10</td>
</tr>
<tr>
<td>+2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**MCTS**

- Stand
- Clap
- Wave
MCTS: Monte Carlo Tree Search - Planning at decision time

- Interleaving planning and acting: model known
- Focusses search on current state
- Can combine w/ learning a model
- Can combine w/ a learned value function
- Random rollouts especially useful in game playing
- Can use more informed rollouts
Approximation

Stand
Clap
Wave
Approximation

Stand
Clap
Wave

Stand
Clap
Wave