- Incorporate time to a decision network

  \[ \text{observe} \rightarrow \text{acts} \rightarrow \text{reward/punishment} \]

- Long but unknown amount of time.
- Indefinitely (infinite horizon process)

**Problem #1: Utility**

Consider:
- A: $1,000, $1,000, $1,000, \ldots$
- B: $1, $1, $2, $2, $3, $4, $5, $6, $7, \ldots$
- C: $1,000,000, $50, $50, $50, $50, $50, $50, \ldots$
- D: $0, $0, $0, $0, \ldots$ $1,000,000, $0, $0, $0, $0, \ldots$

“Value”

\[ V = \sum_{i=0}^{\infty} r_i \]

\[ V = \lim_{n \to \infty} \frac{r_1 + \cdots + r_n}{n} \]

\[ V = r_1 + \gamma r_2 + \gamma^2 r_3 + \gamma^3 r_4 + \gamma^4 r_5 + \cdots \]

\[ \gamma: \text{discount rate} \quad 0 \leq \gamma \leq 1 \]

\[ = r_1 + \gamma (r_2 + \gamma (r_3 + \gamma (r_4 + \cdots ))) \]

\[ V_t = r_1 + \gamma V_{t+1} \]

If we have the following properties:
- The next state depends only on the current action.
- The effects of actions are stochastic, but do not change over time.

**Belief Network: Markov Decision Process (MDP)**
An MDP consists of:
- a set $S$ of states
- a set $A$ of actions
- the dynamics $P(S_{i+1} | S_i, A_i)$
- Reward table $R(S, A, S')$
  
  Sometimes expected value of $R,$ we use $\gamma$ the discount factor

Example:
Agent Bob

$S = \{\text{healthy, sick}\}$

$A = \{\text{relax, party}\}$

What should Bob do each weekend?

\[
P(S' | S, A)\]

<table>
<thead>
<tr>
<th>$S$</th>
<th>$A$</th>
<th>$P(S' = \text{healthy})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>healthy</td>
<td>relax</td>
<td>0.95</td>
</tr>
<tr>
<td>healthy</td>
<td>party</td>
<td>0.70</td>
</tr>
<tr>
<td>sick</td>
<td>relax</td>
<td>0.58</td>
</tr>
<tr>
<td>sick</td>
<td>party</td>
<td>0.18</td>
</tr>
</tbody>
</table>

\[
R(S, A)\]

<table>
<thead>
<tr>
<th>$S$</th>
<th>$A$</th>
<th>$R(S, A)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>healthy</td>
<td>relax</td>
<td>70</td>
</tr>
<tr>
<td>healthy</td>
<td>party</td>
<td>100</td>
</tr>
<tr>
<td>sick</td>
<td>relax</td>
<td>80</td>
</tr>
<tr>
<td>sick</td>
<td>party</td>
<td>20</td>
</tr>
</tbody>
</table>

A Policy:

A policy $\pi : S \rightarrow A$ what to do at each state.

Among policies, $\pi^*$ Optimal policy is the one with maximum expected reward.

An MDP with stationary dynamics always has an optimal policy.
An MDP with stationary dynamics always has an optimal policy.

- Partially Observable Markov Decision Processes (POMDP)