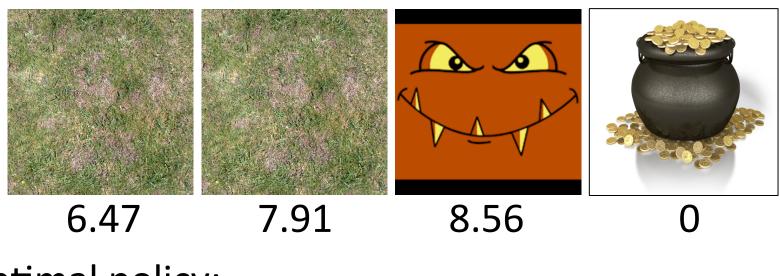
V[s] values converge to:



Optimal policy:

A B ---

- You are a doctor performing a clinical trial for new medications (called A and B) to treat Ebola.
- It turns out the medications interact with each other.
 - Giving A then B makes you healthier 90% of the time,
 but sicker 10% of the time.
 - Giving B then A makes you sicker 90% of the time, but healthier 10% of the time.
 - Giving A then A makes you healthier 50% of the time,
 but does nothing 50% of the time.
 - Giving B then B makes you sicker 50% of the time, but does nothing 50% of the time.

Review

- Value iteration requires a perfect model of the environment.
 - You need to know P(s' | s, a) and R(s, a, s') ahead
 of time for all combinations of s, a, and s'.
 - Optimal V or Q values are computed directly from the environment using the Bellman equations.
- Often impossible or impractical.