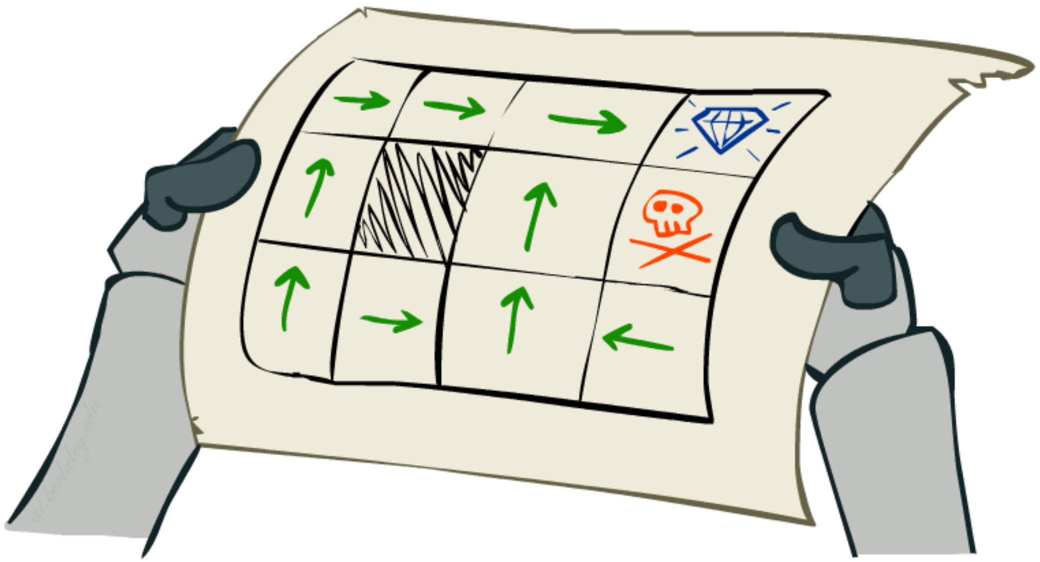
Advanced Topics in AI

Solving MDPs



Instructor: Prof. Dr. techn. Wolfgang Nejdl

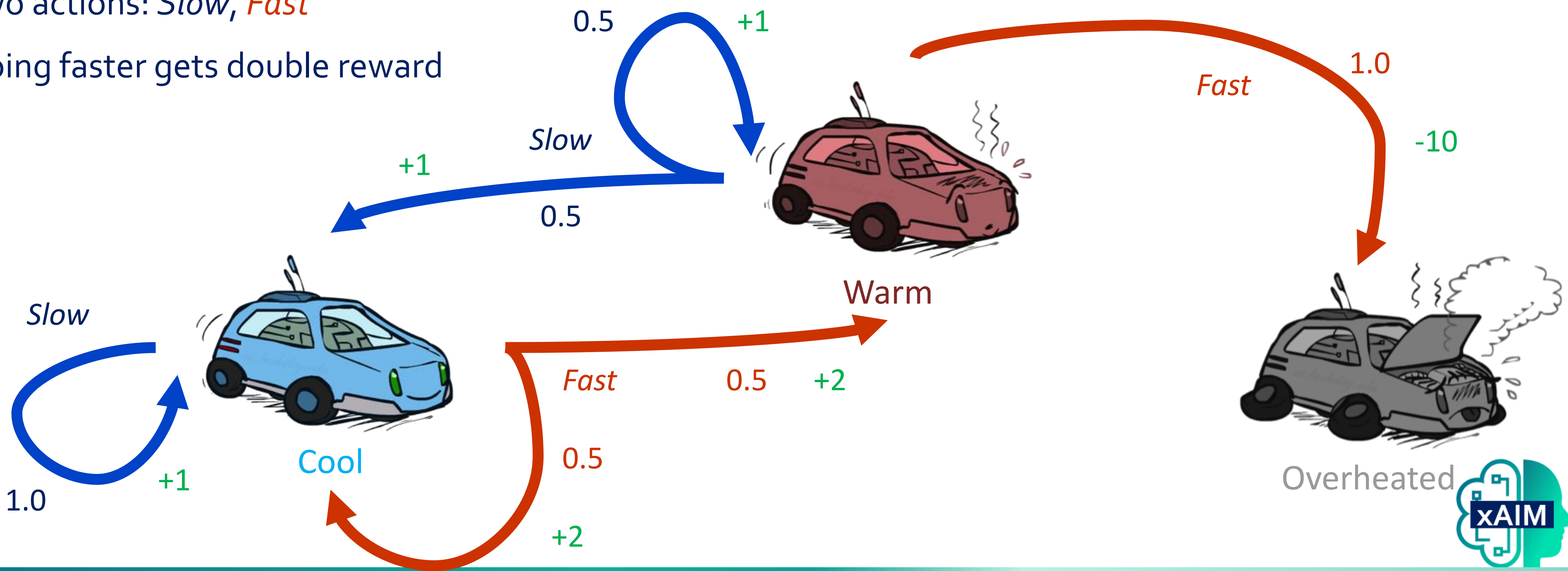
Leibniz University Hannover

[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All materials are available at http://ai.berkeley.edu.]

Co-financed by the Connecting Europe Facility of the European Union

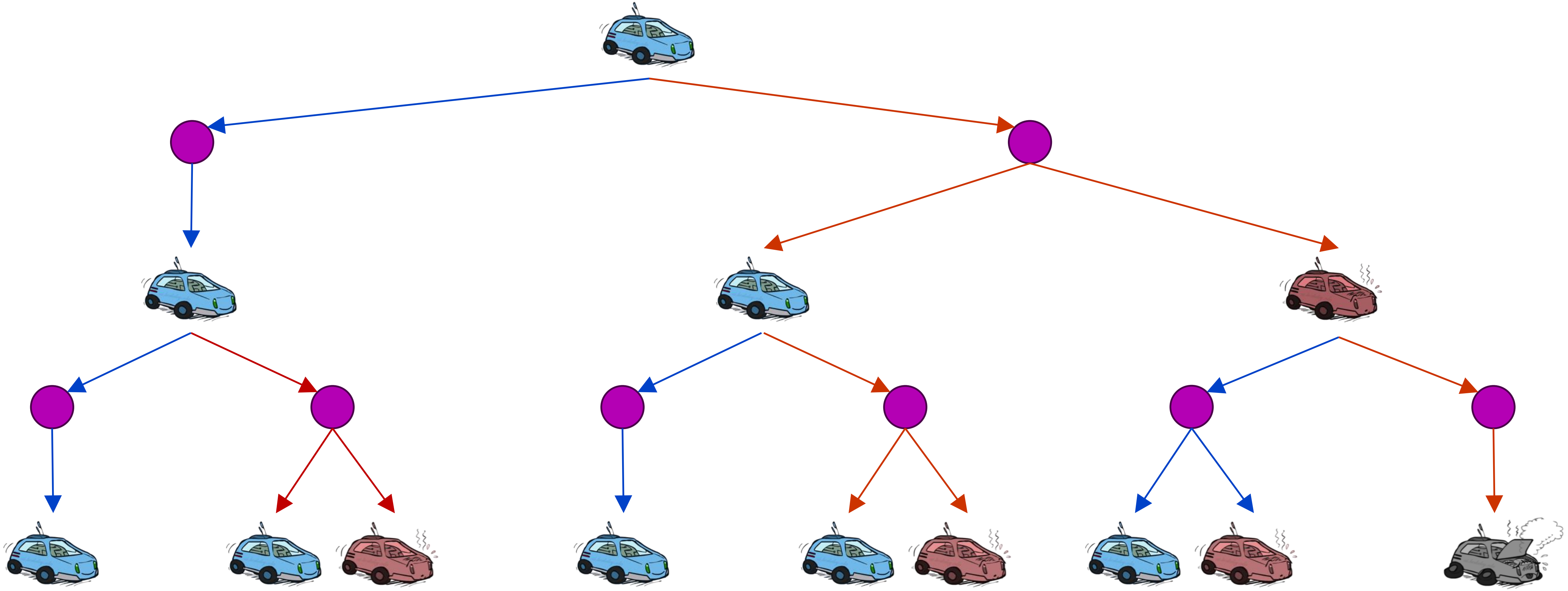
Recall: Racing MDP

- A robot car wants to travel far, quickly
- Three states: Cool, Warm, Overheated
- Two actions: *Slow*, *Fast*
- Going faster gets double reward



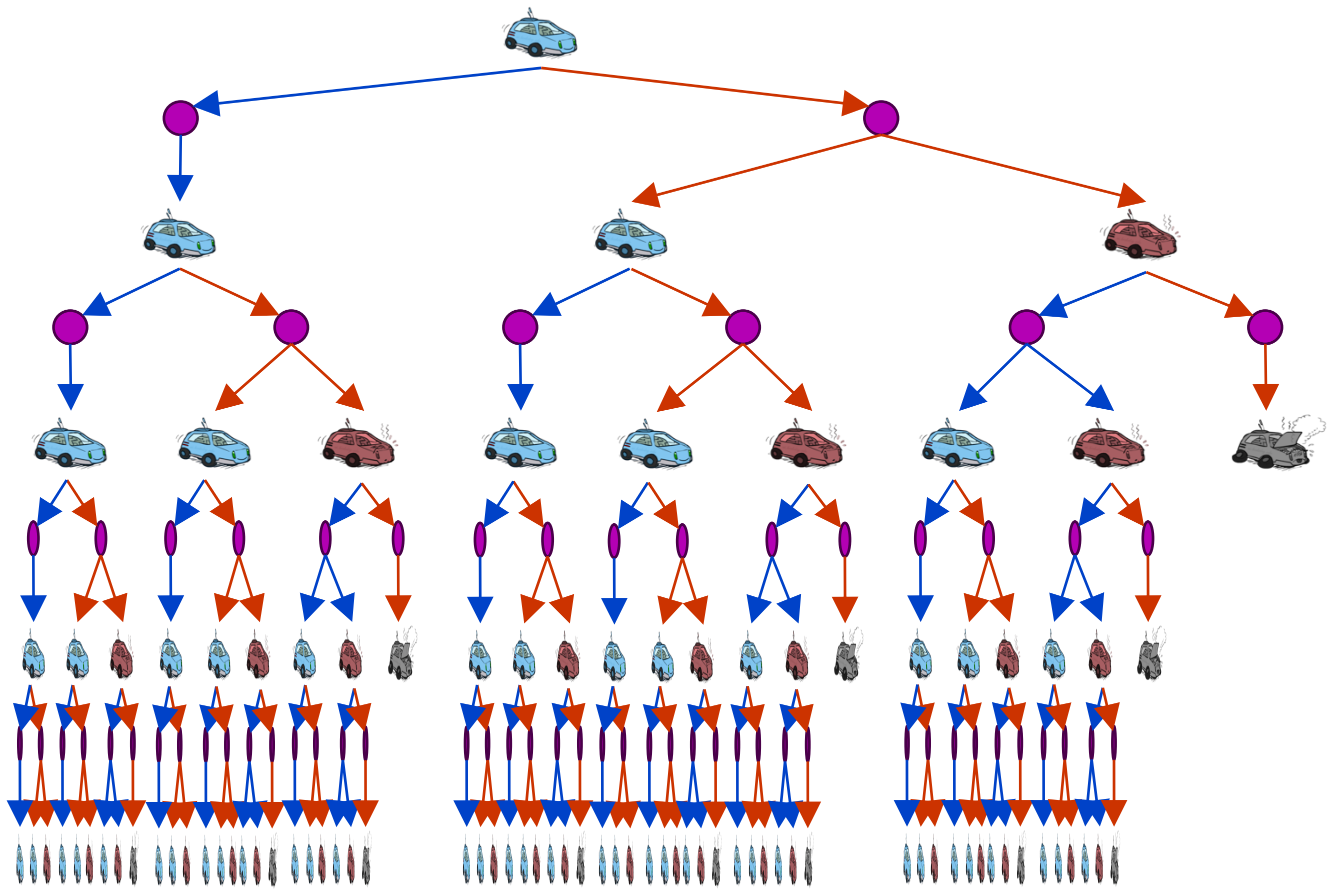
Co-financed by the Connecting Europe Facility of the European Union

Racing Search Tree



Co-financed by the Connecting Europe Facility of the European Union

Racing Search Tree



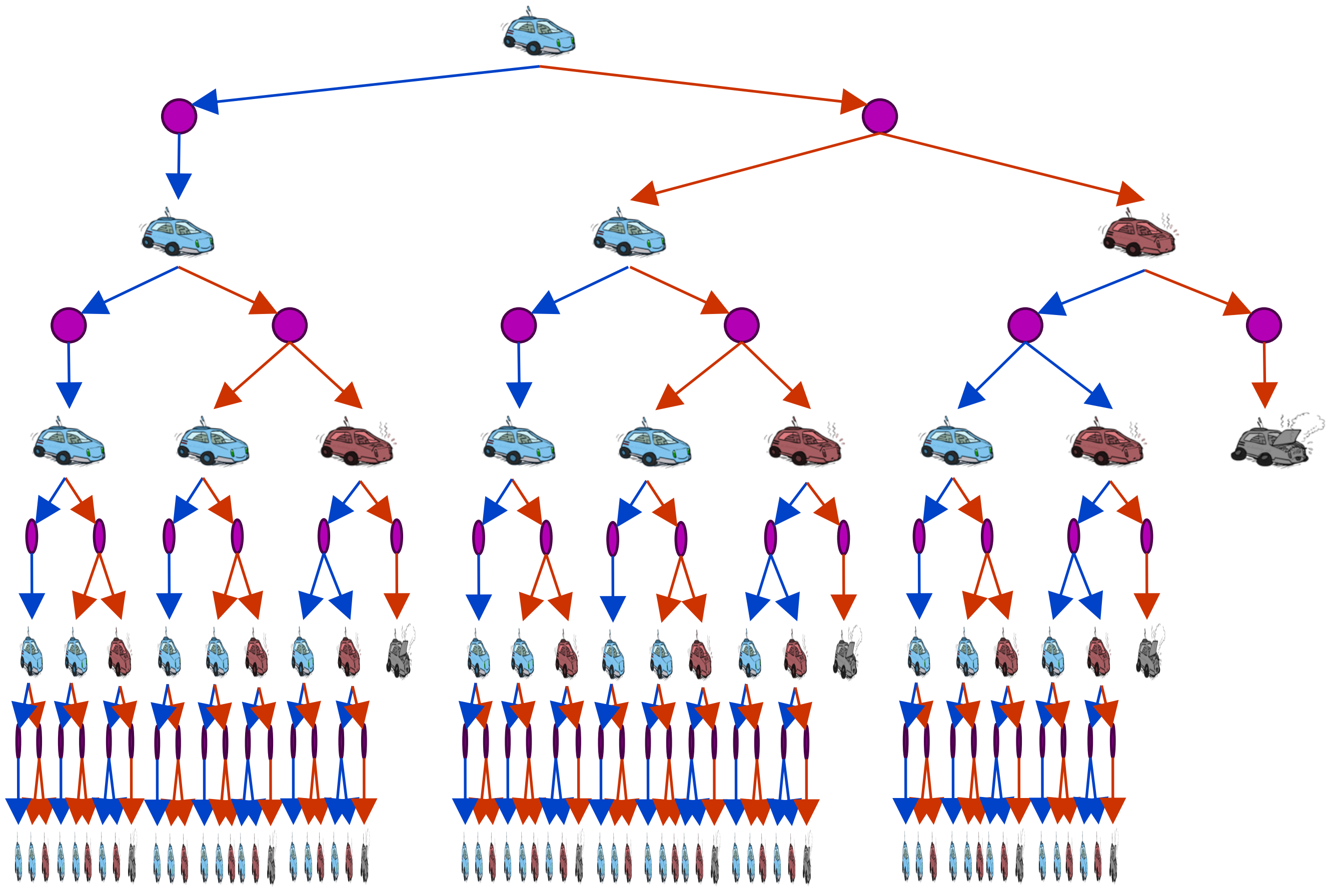
Co-financed by the Connecting Europe Facility of the European Union

Racing Search Tree

- We're doing way too much work with expectimax!

- Problem: States are repeated
  - Idea: Only compute needed quantities once

- Problem: Tree goes on forever
  - Idea: Do a depth-limited computation, but with increasing depths until change is small
  - Note: deep parts of the tree eventually don't matter if $\gamma < 1$



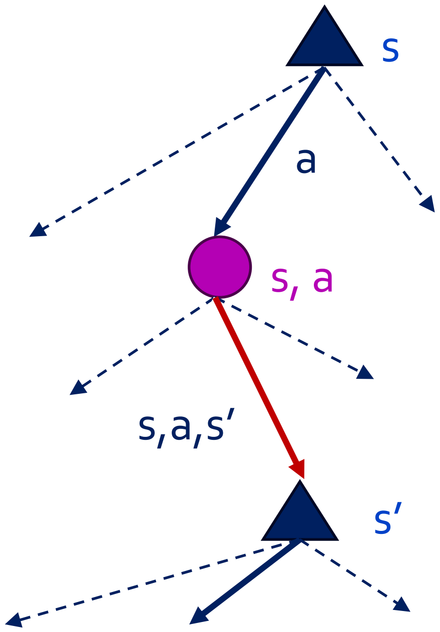
Co-financed by the Connecting Europe Facility of the European Union

Optimal Quantities

- The value (utility) of a state s:
  - $V^*(s)$ = expected utility starting in s and acting optimally

- The value (utility) of a q-state (s,a):
  - $Q^*(s,a)$ = expected utility starting out having taken action a from state s and (thereafter) acting optimally

- The optimal policy:
  - $\pi^*(s)$ = optimal action from state s



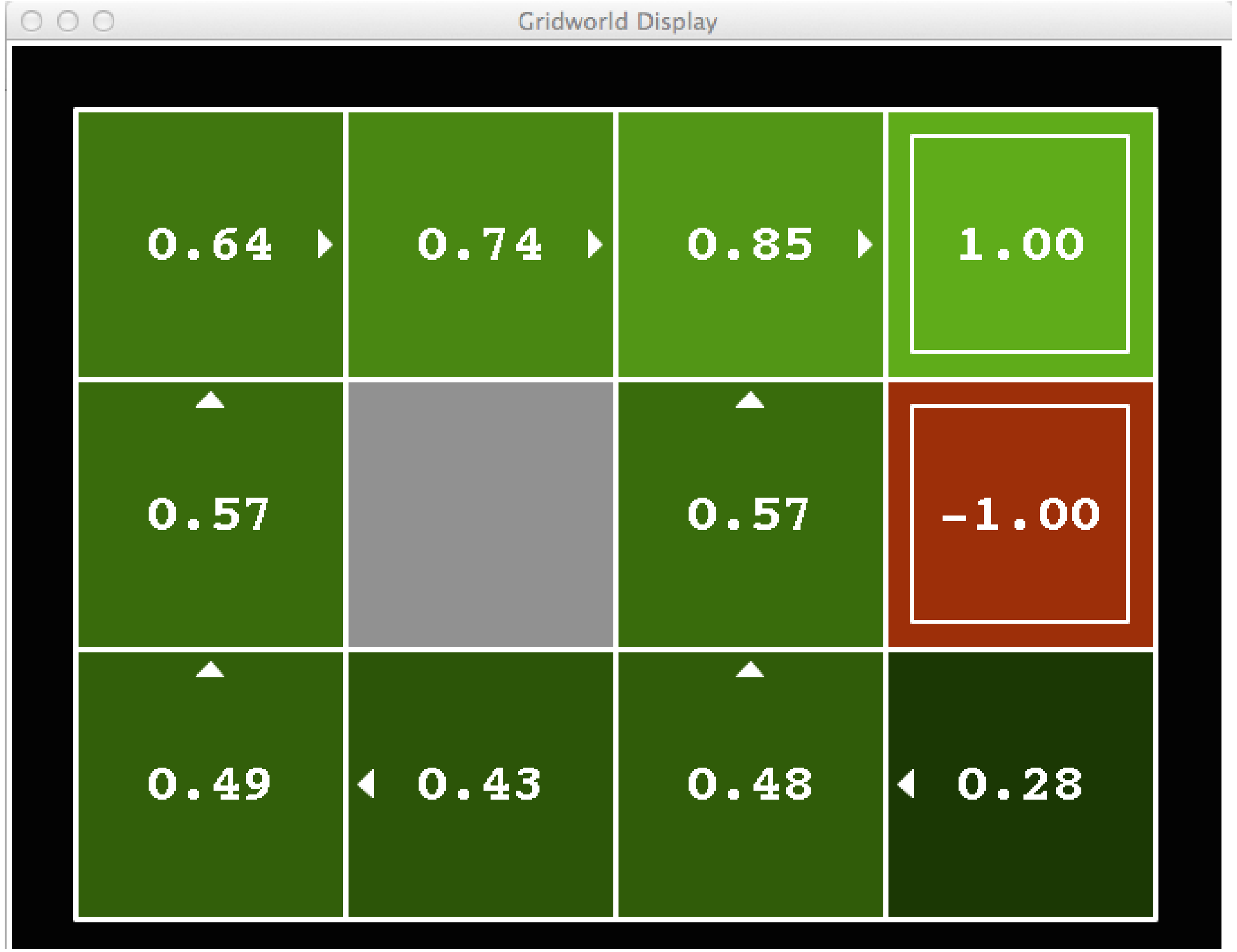
s is a *state*

(s, a) is a *q-state*

(s,a,s') is a *transition*

Co-financed by the Connecting Europe Facility of the European Union

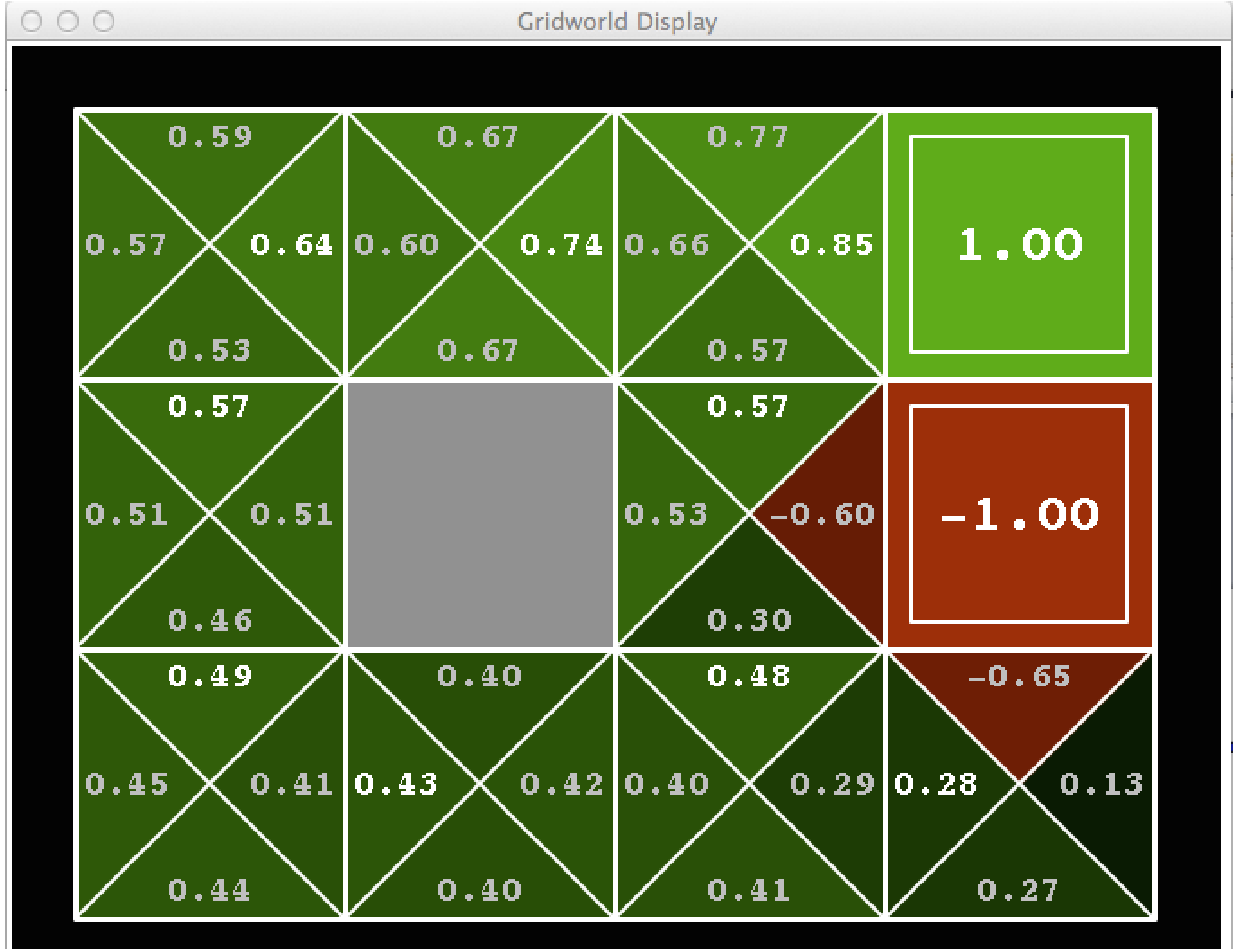
Gridworld V* Values



Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

Gridworld Q* Values

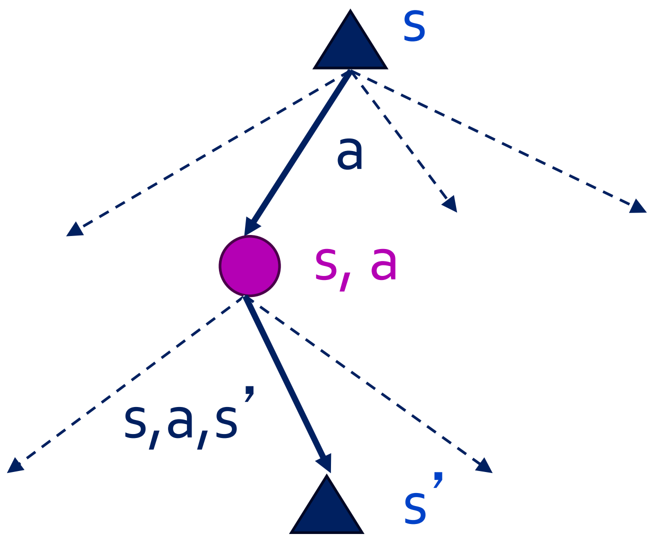


Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

Values of States

- Recursive definition of value:



$$V^*(s) = \max_a \quad Q^*(s,a)$$

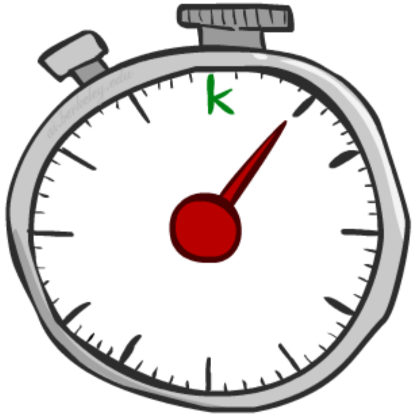
$$Q^*(s,a) = \sum_{s'} T(s,a,s')[\, R(s,a,s') + \gamma V^*(s') \,]$$

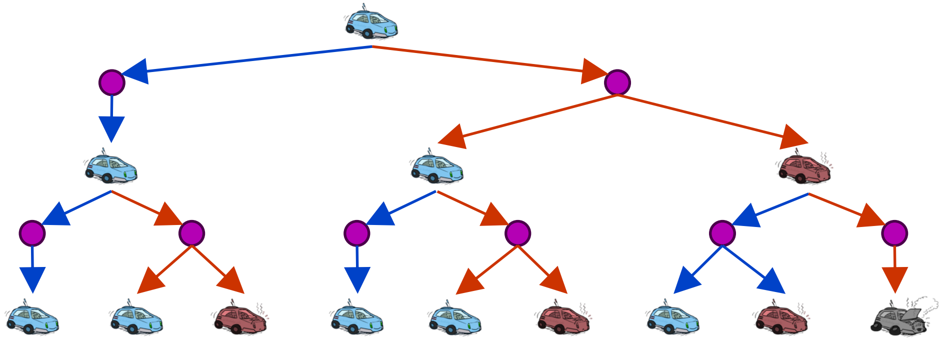
$$V^*(s) = \max_a \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma\, V^*(s')]$$

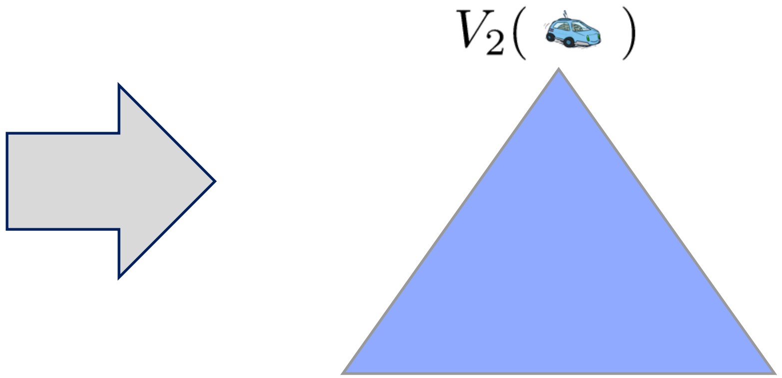
Co-financed by the Connecting Europe Facility of the European Union

Time-Limited Values

- Key idea: time-limited values
- Define $V_k(s)$ to be the optimal value of s if the game ends in k more time steps
  - Equivalently, it's what a depth-k expectimax would give from s

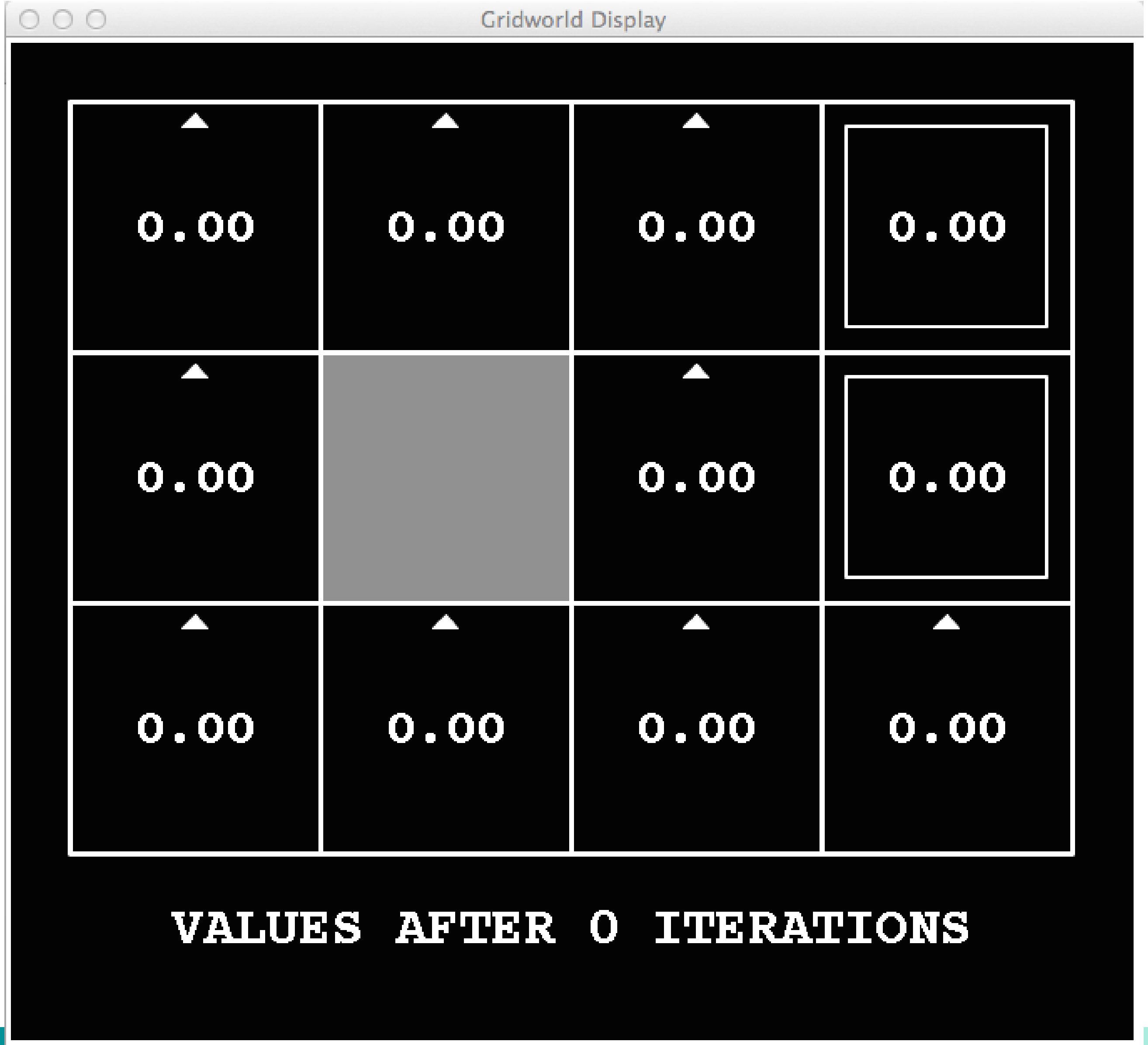






Co-financed by the Connecting Europe Facility of the European Union

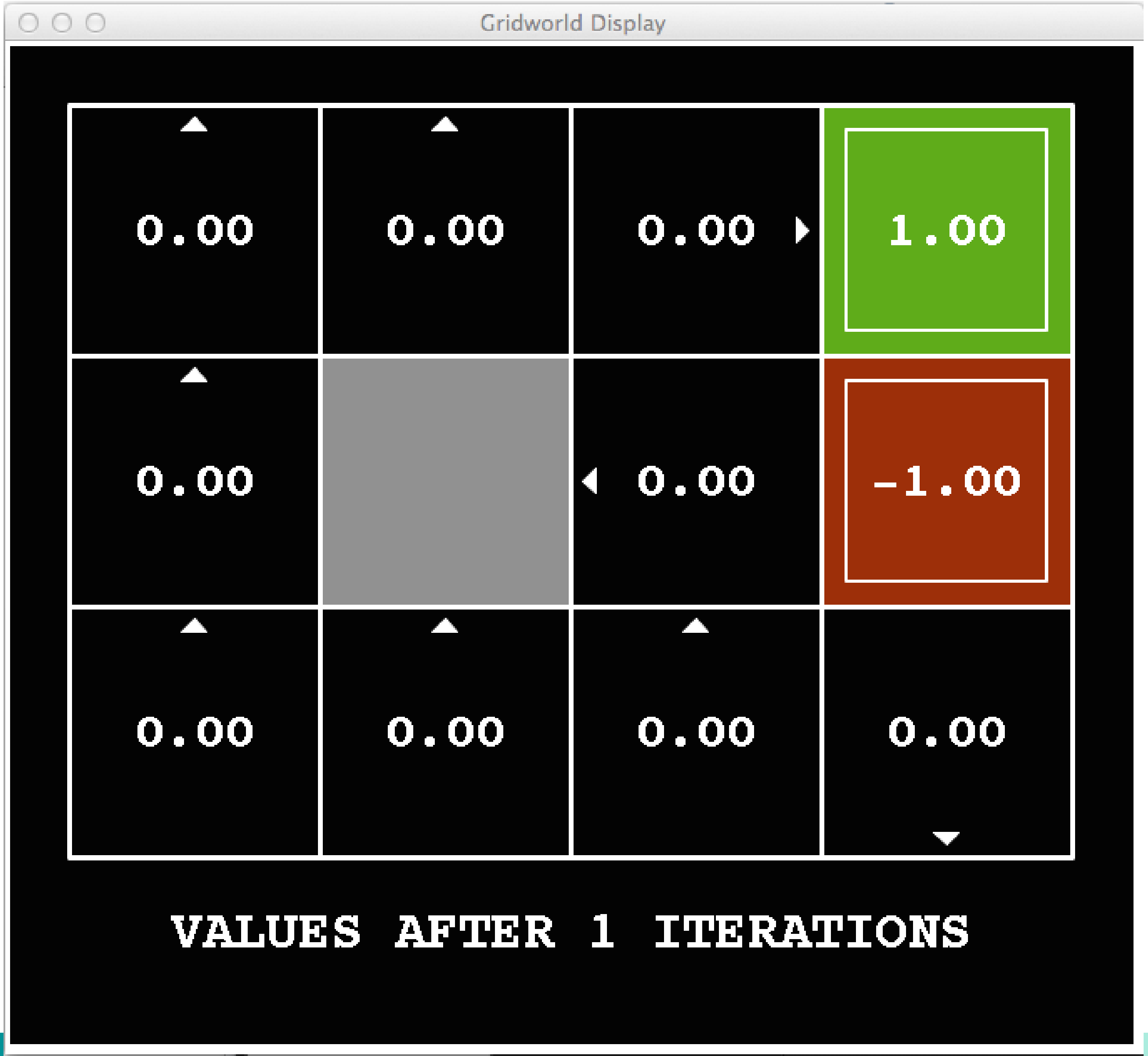
k=o



Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

k=1



Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

k=2

Noise = 0.2
Discount = 0.9
Living reward = 0

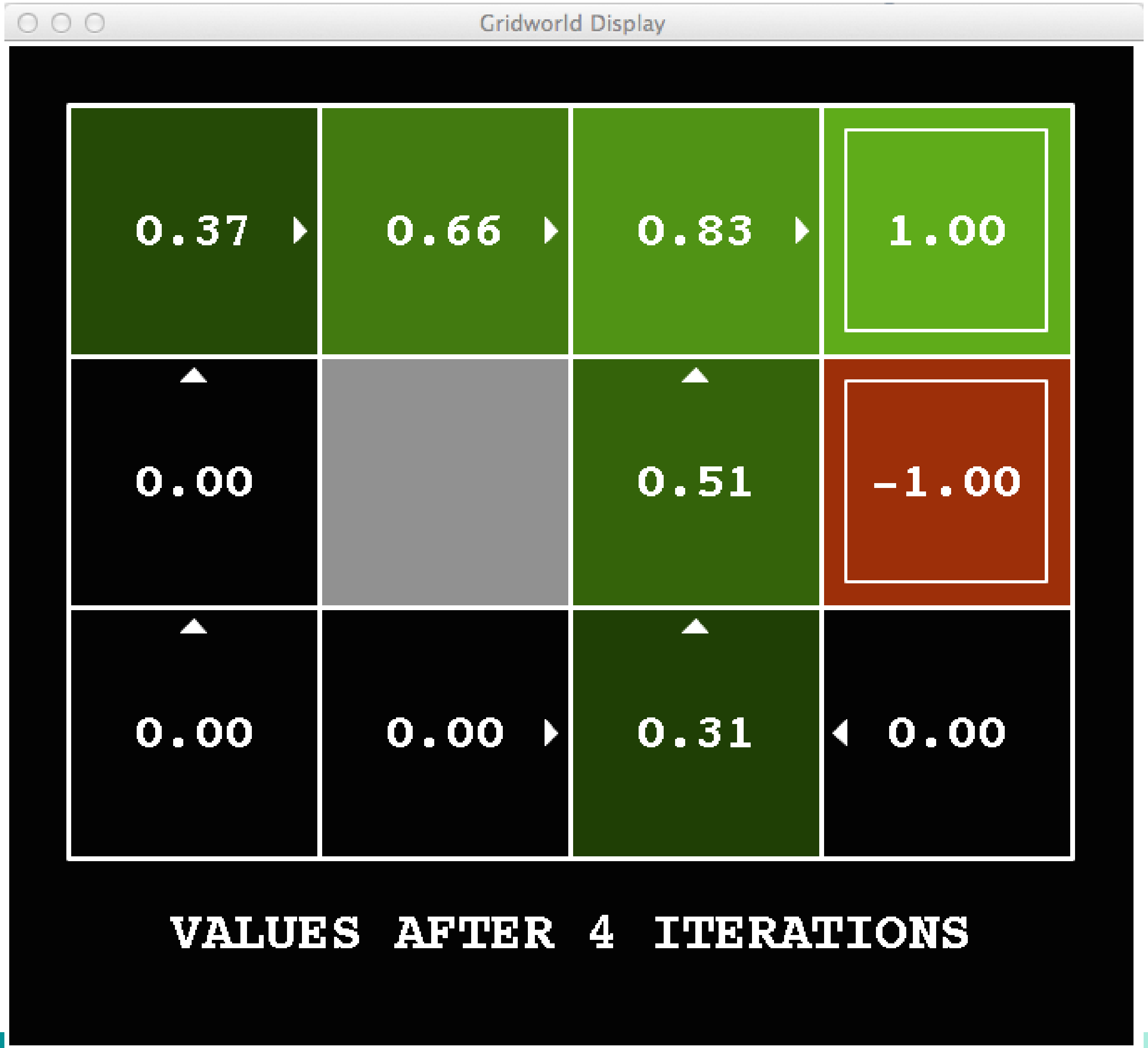
Co-financed by the Connecting Europe Facility of the European Union

k=3

Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

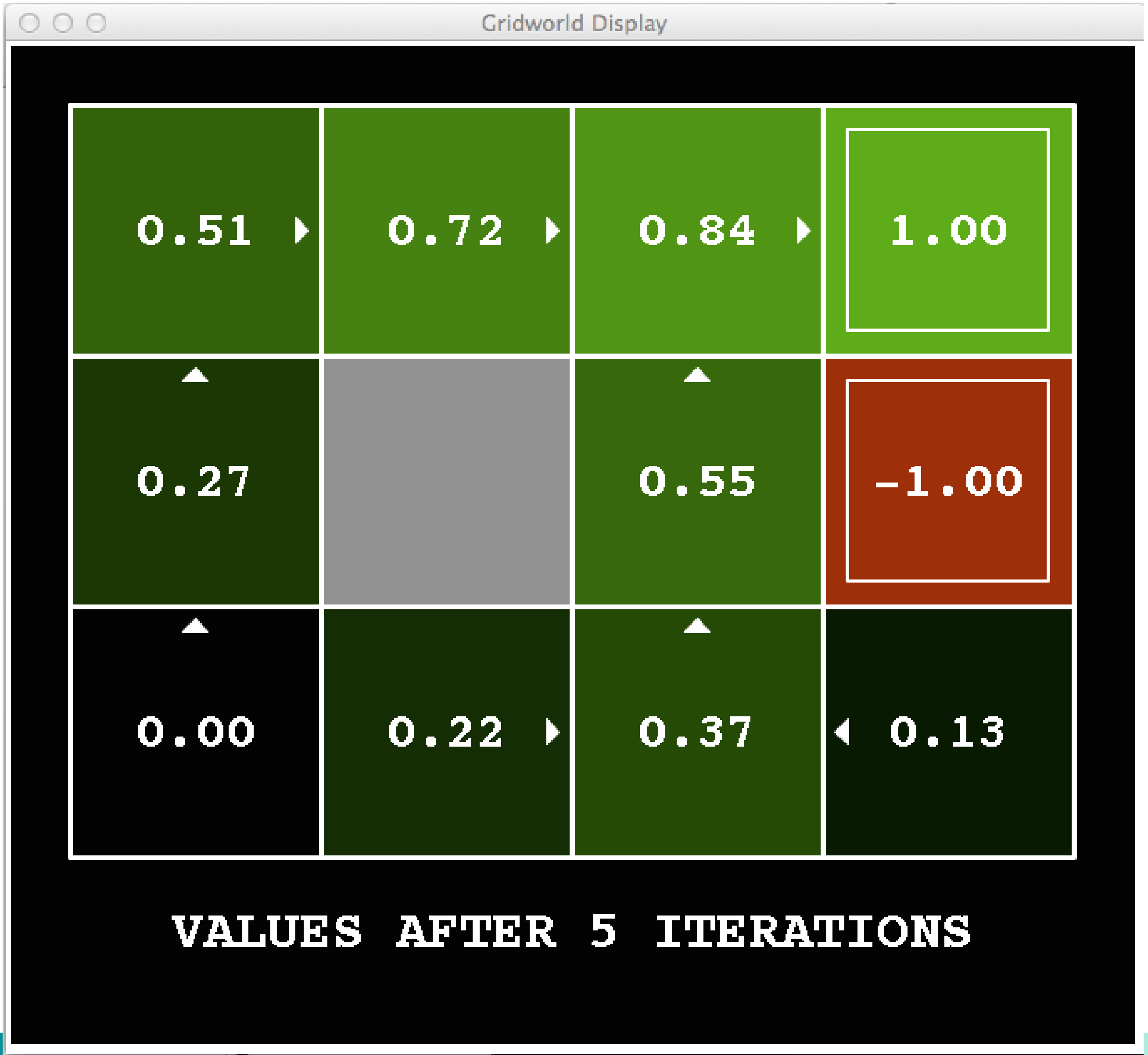
k=4



Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

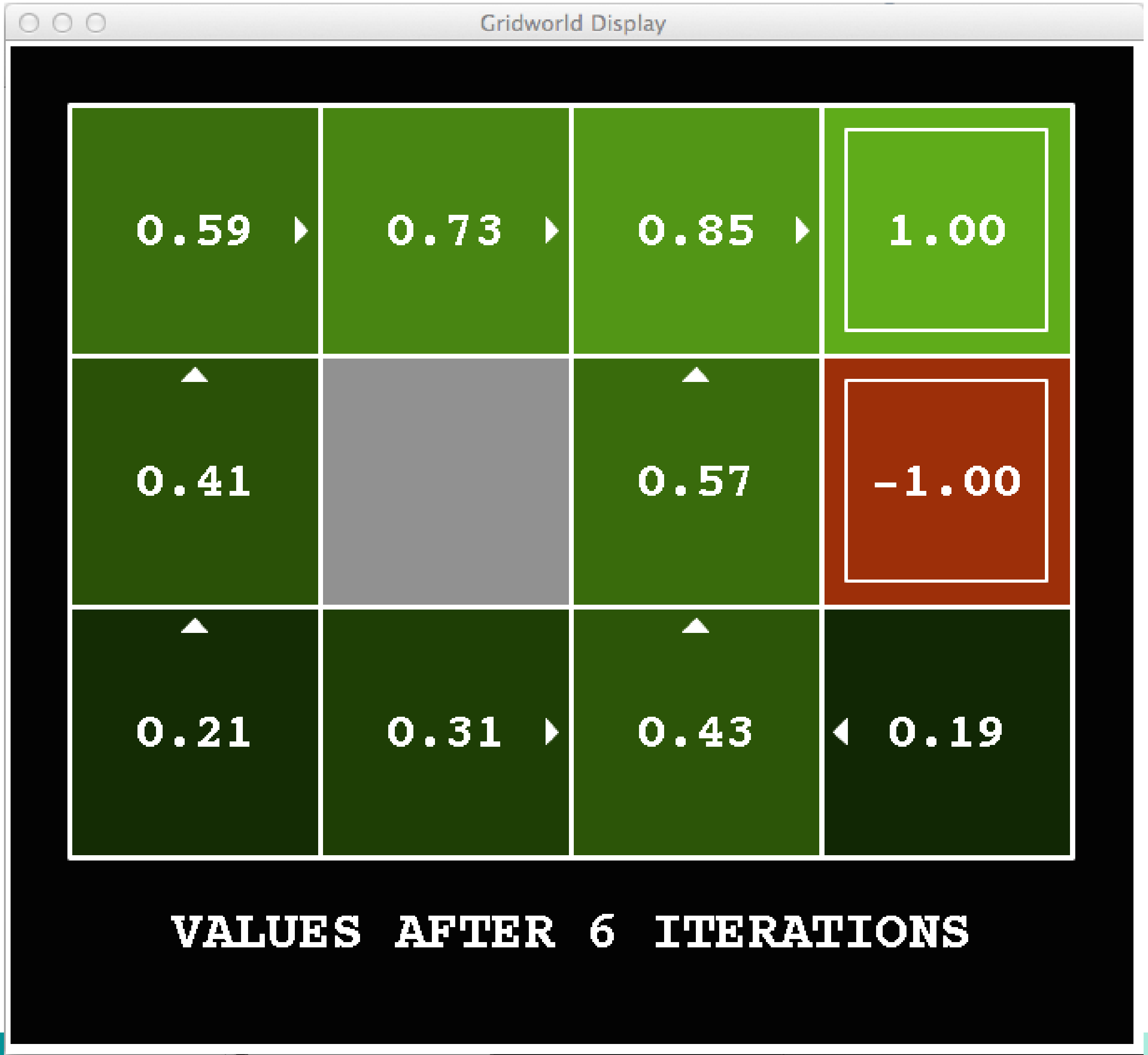
k=5



Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

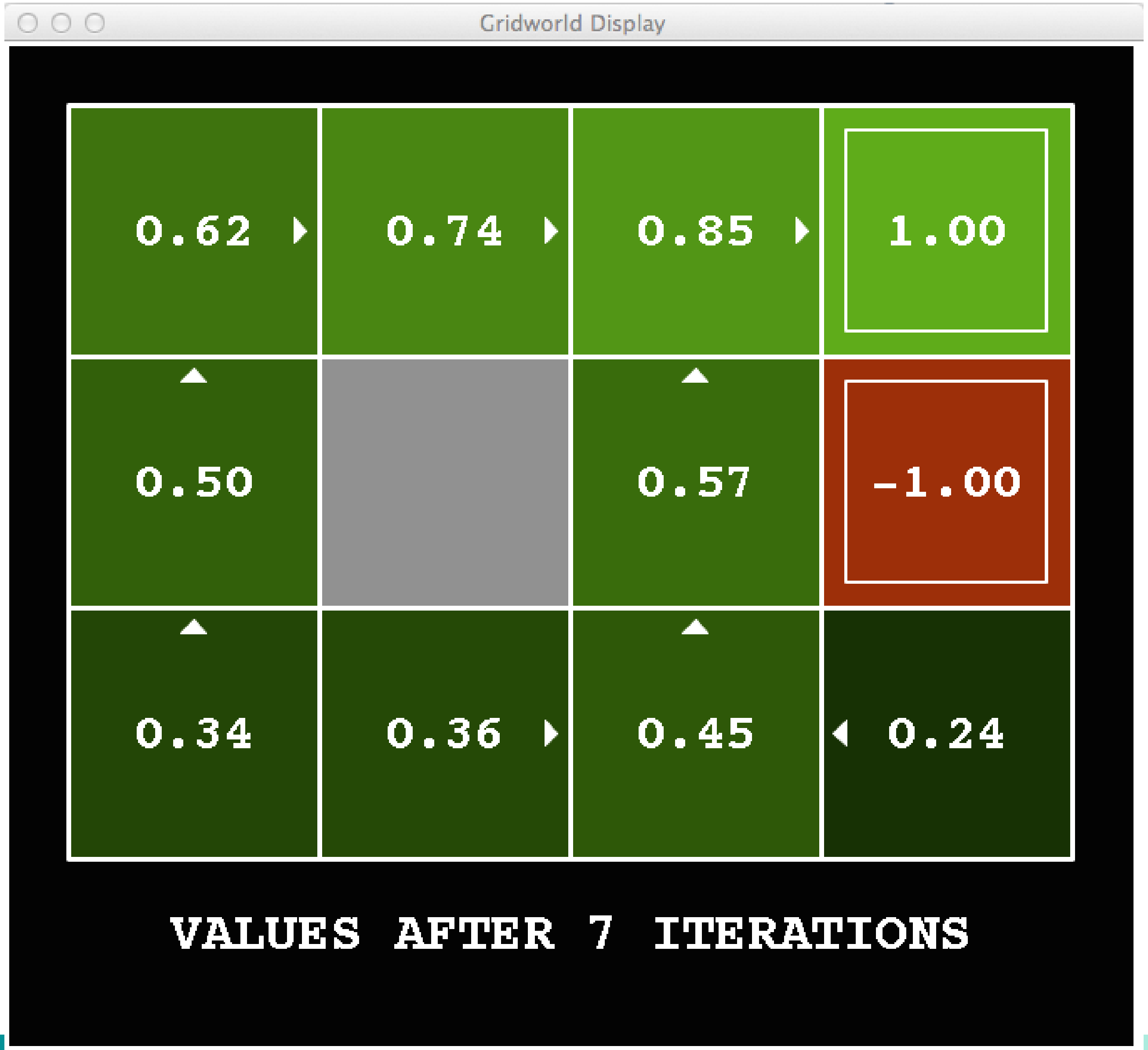
k=6



Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

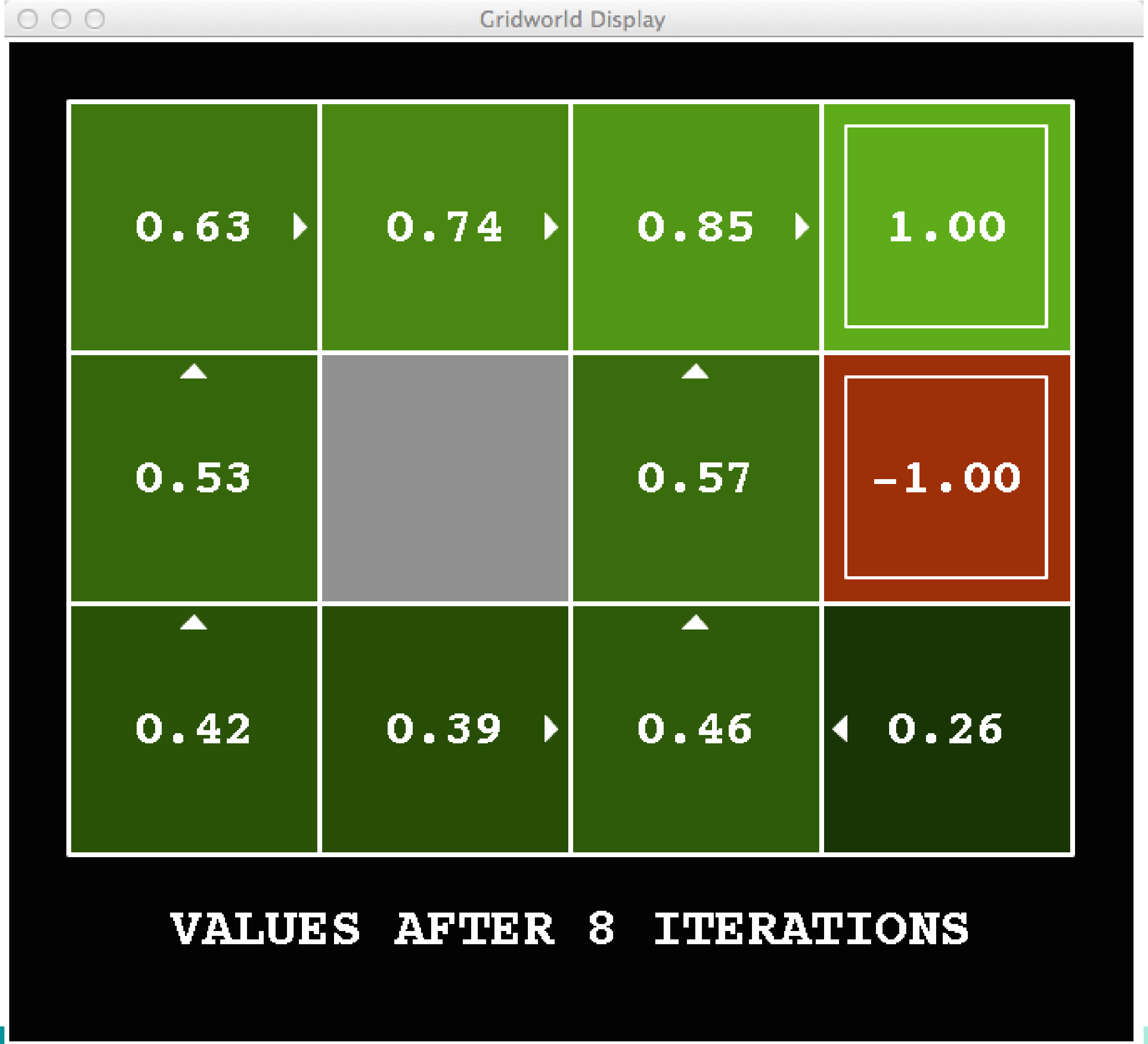
k=7



Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

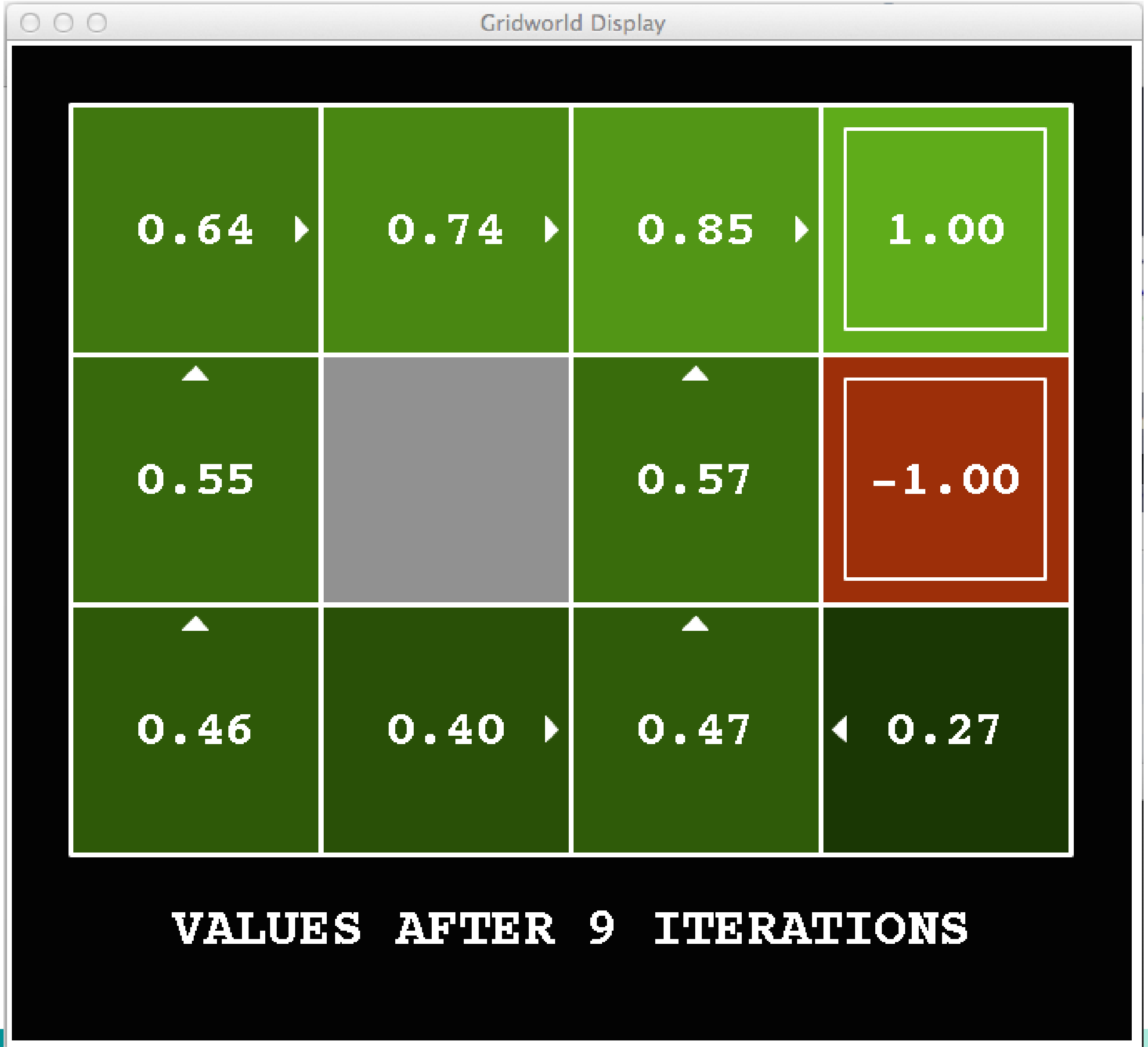
k=8



Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

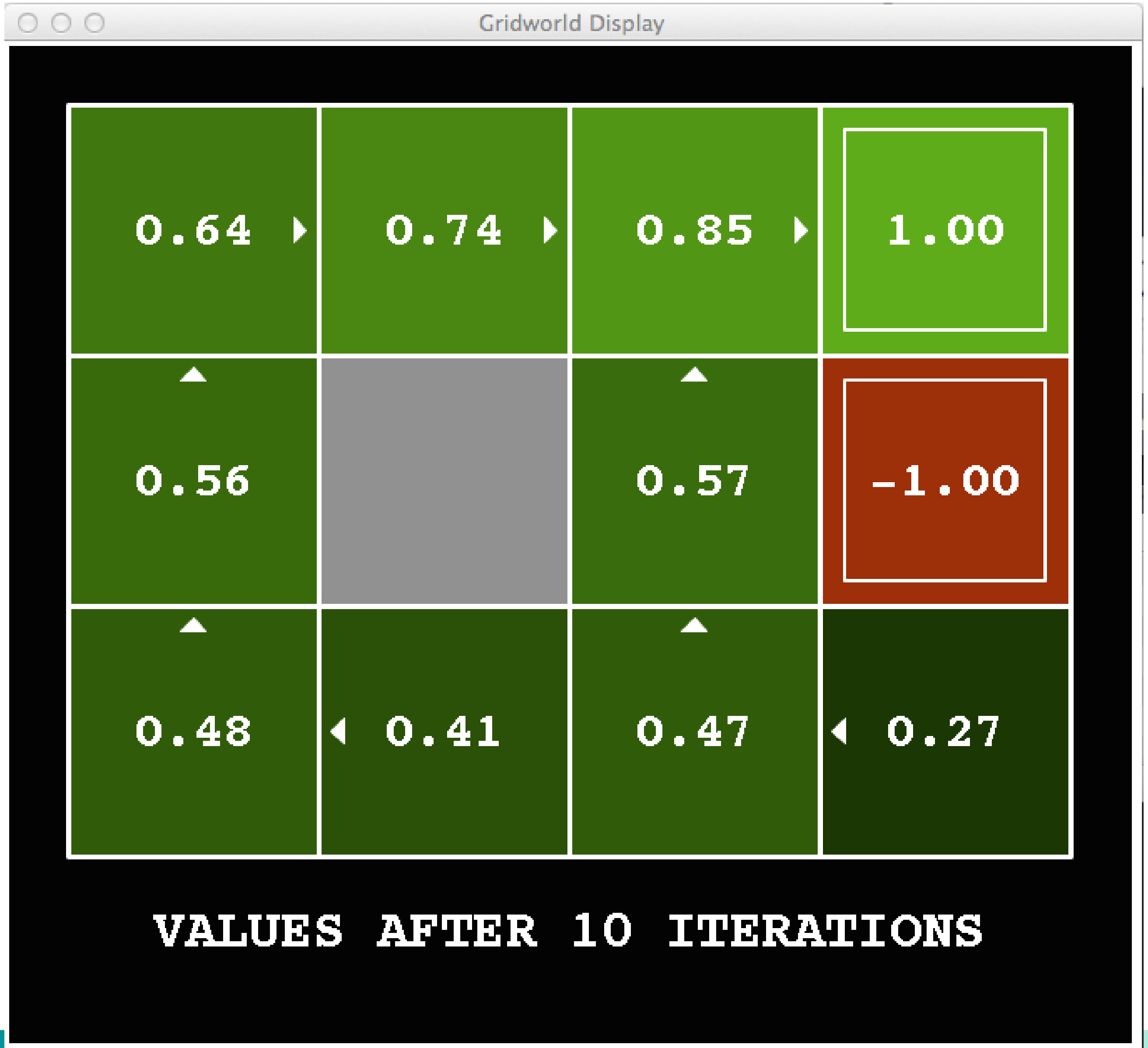
k=9



Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

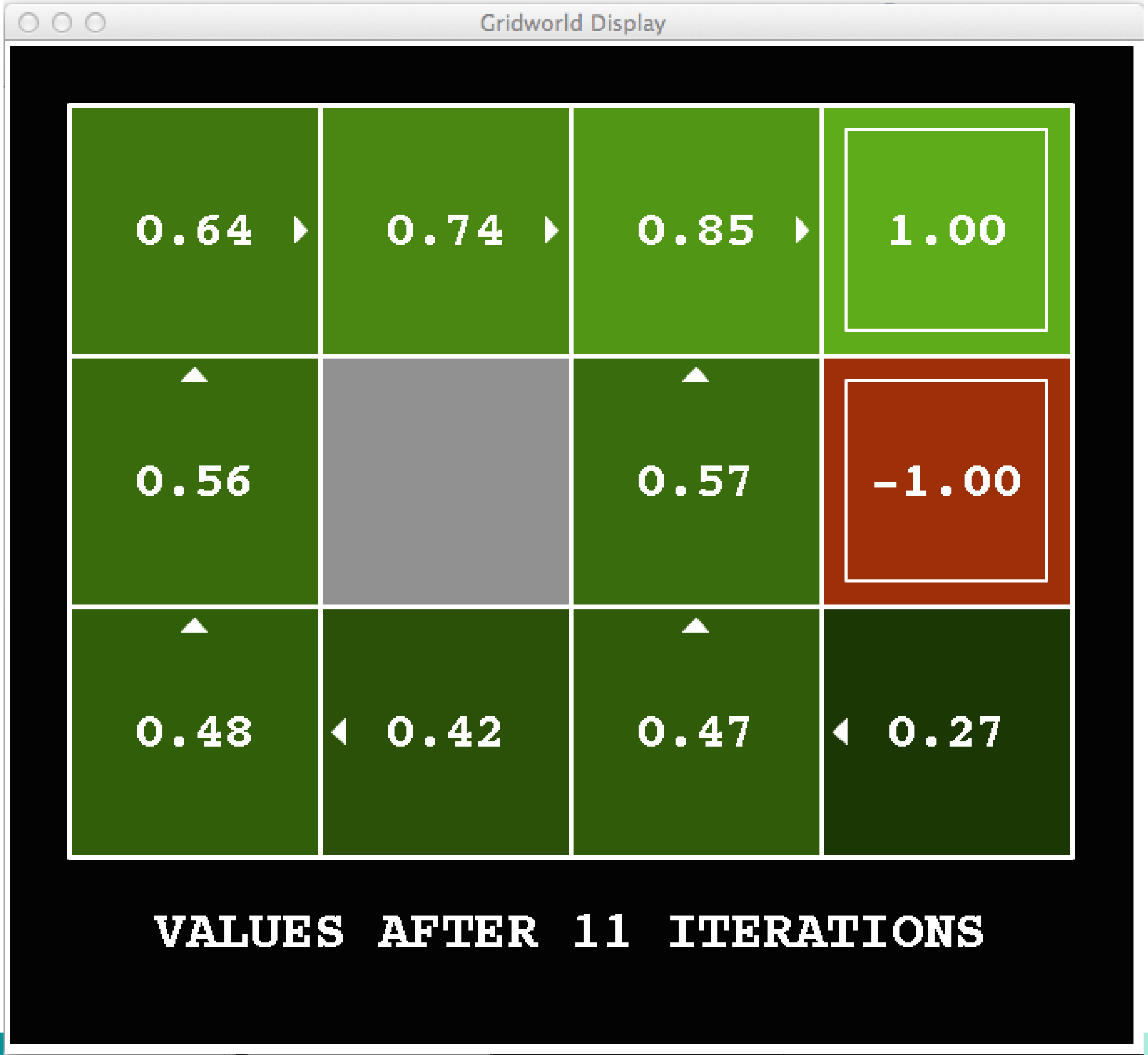
k=10



Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

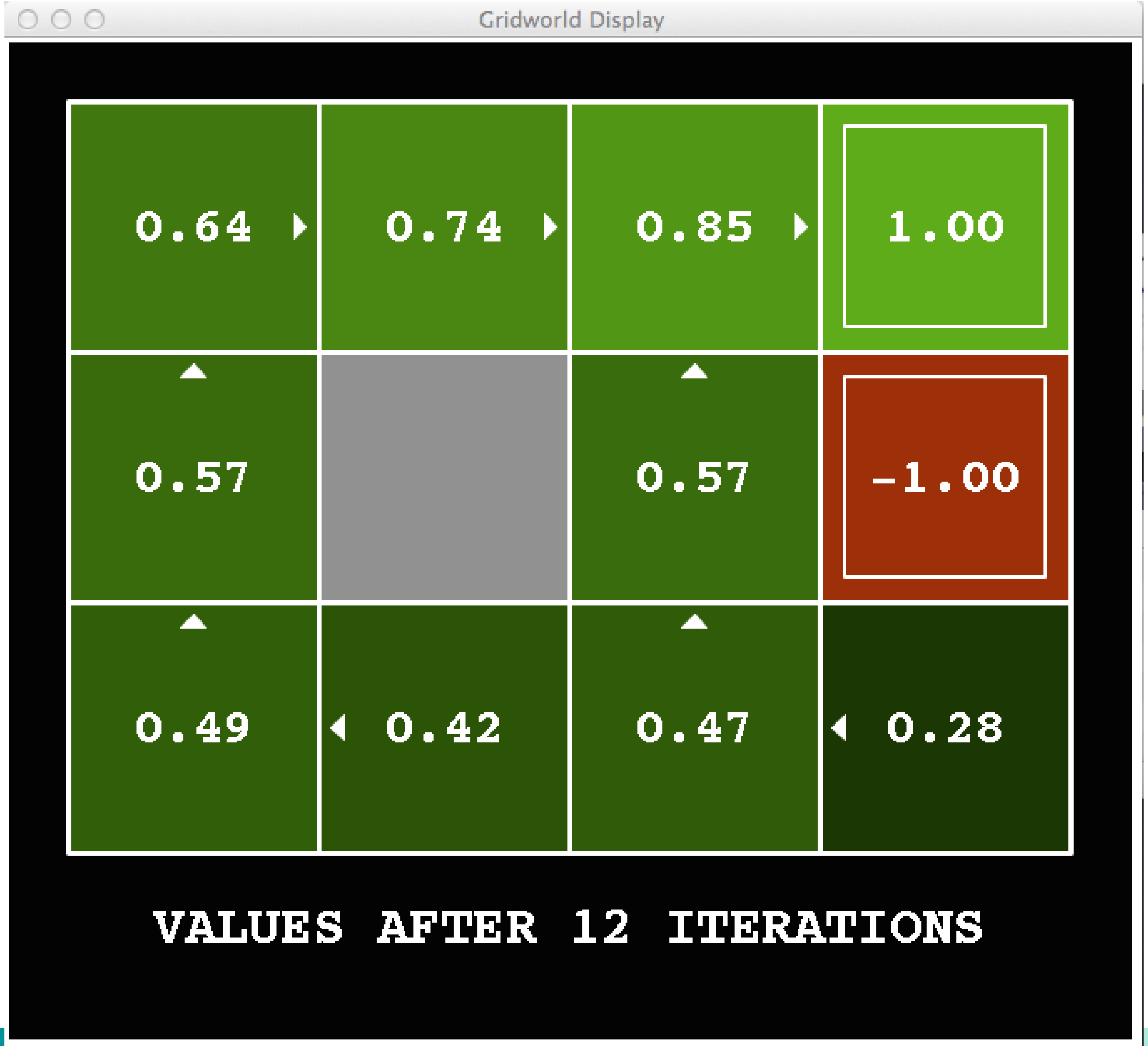
k=11



Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

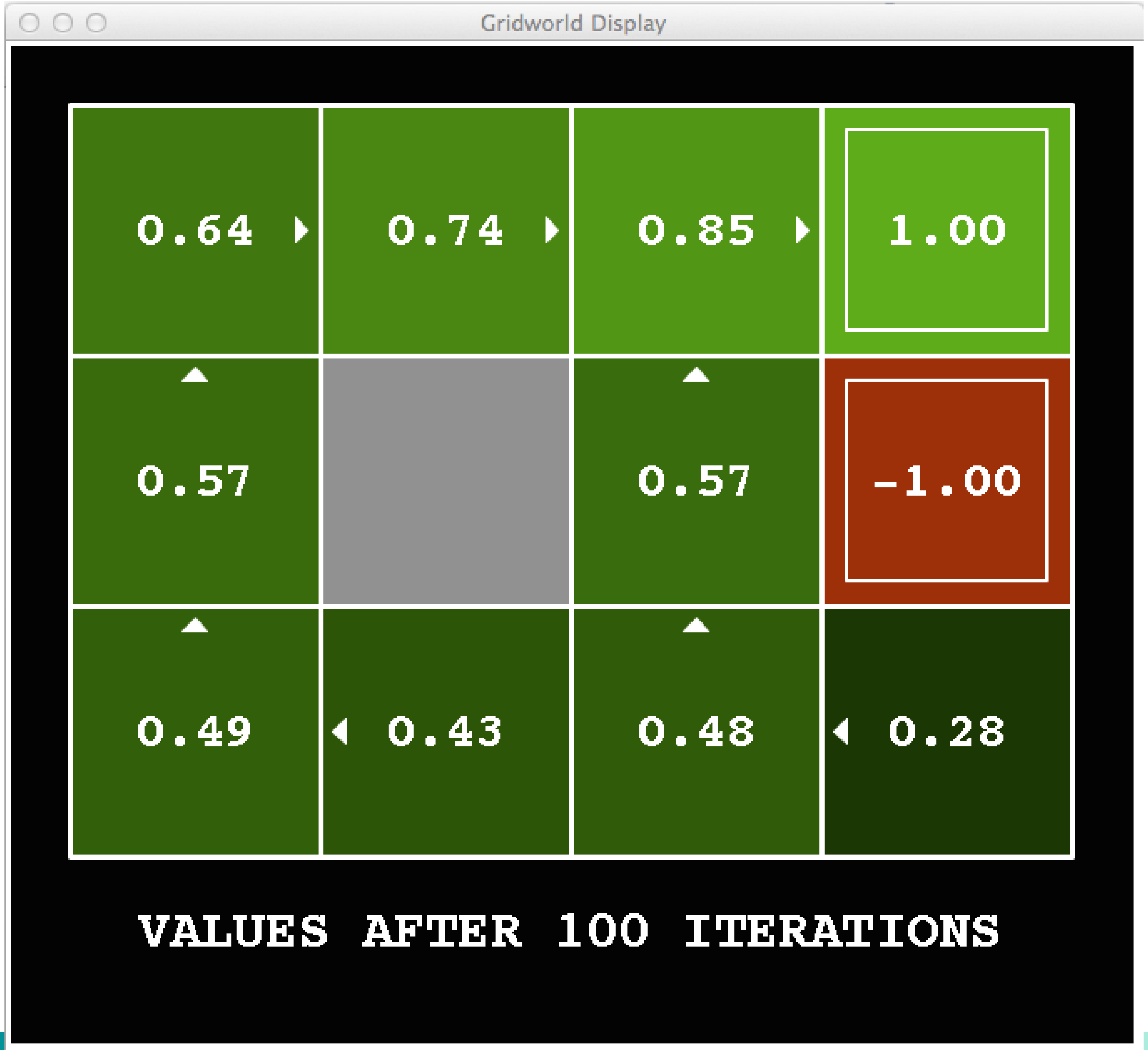
k=12



Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

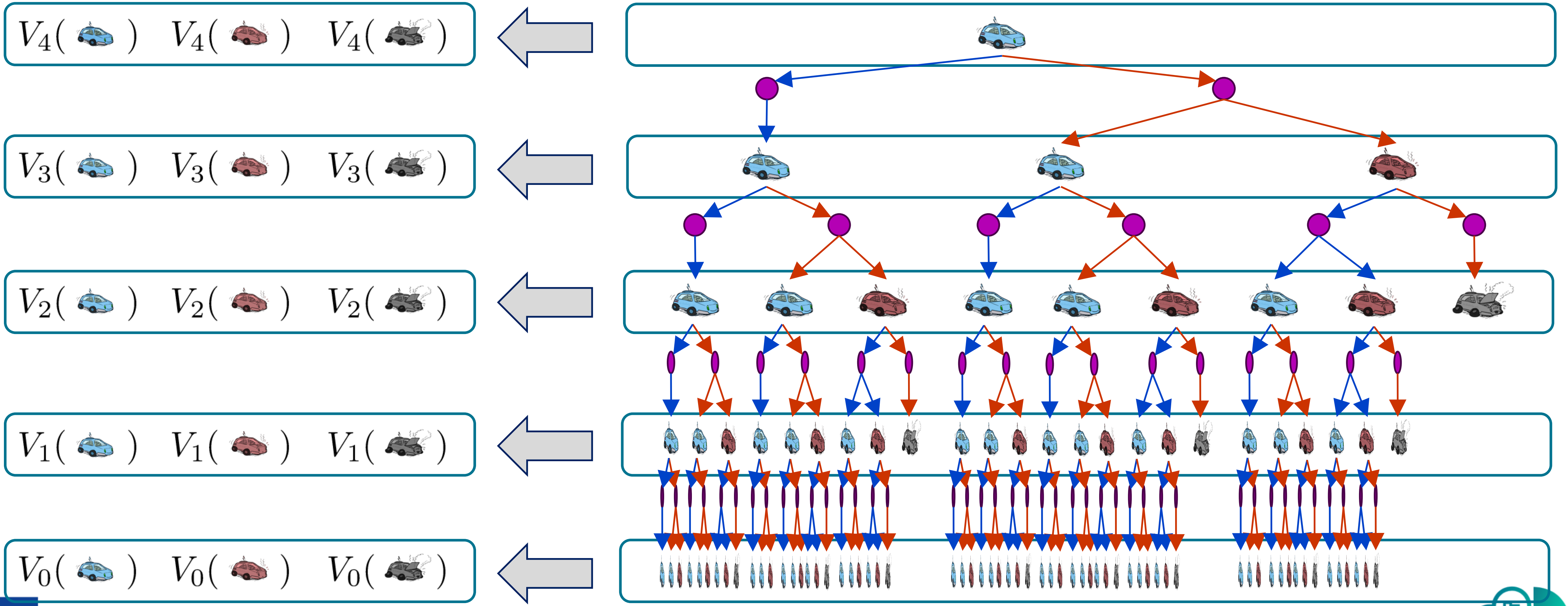
k=100



Noise = 0.2
Discount = 0.9
Living reward = 0

Co-financed by the Connecting Europe Facility of the European Union

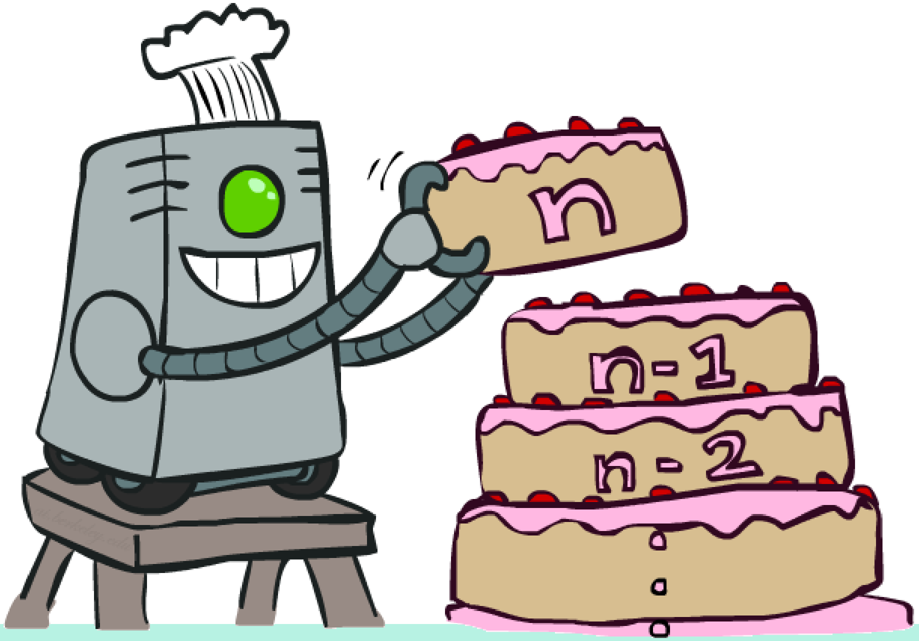
Computing Time-Limited Values



Co-financed by the Connecting Europe Facility of the European Union

Advanced Topics in AI

Next: Value Iteration



Instructor: Prof. Dr. techn. Wolfgang Nejdl

Leibniz University Hannover

Co-financed by the Connecting Europe Facility of the European Union