Correlated-Q Learning

Greenwald & Hall, 2003

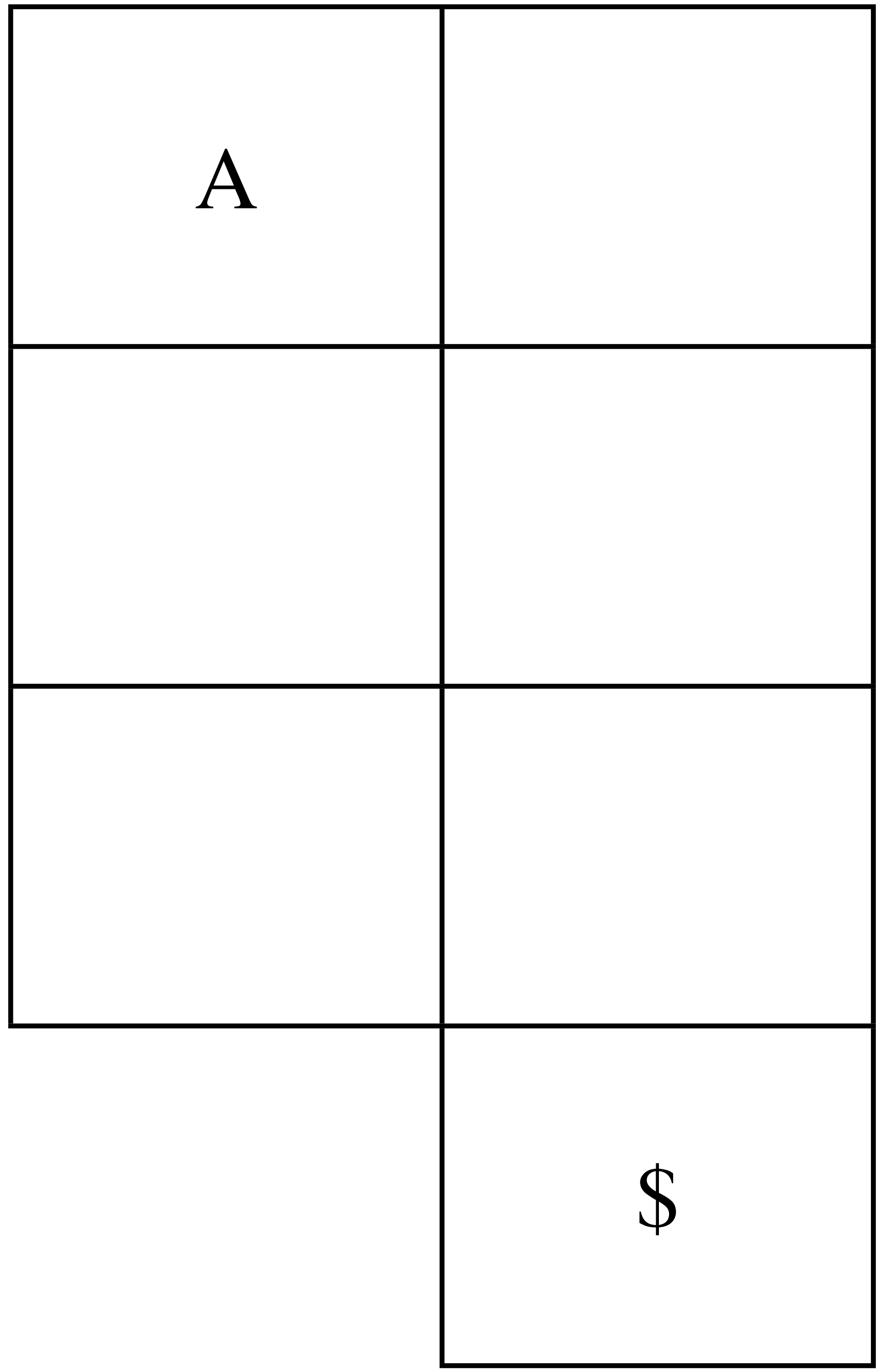
and

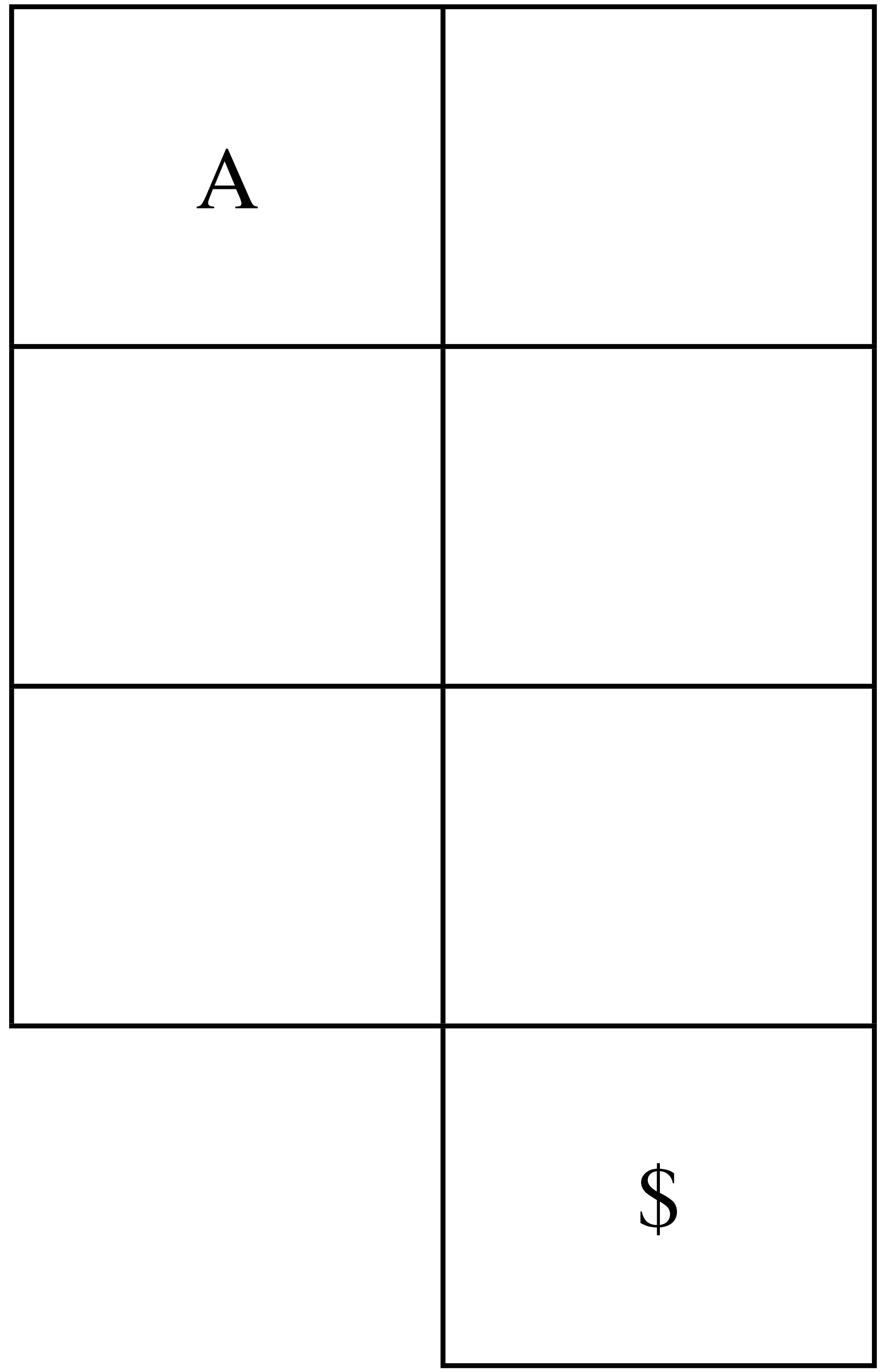
Cooperation In Strategic Games, Revisited

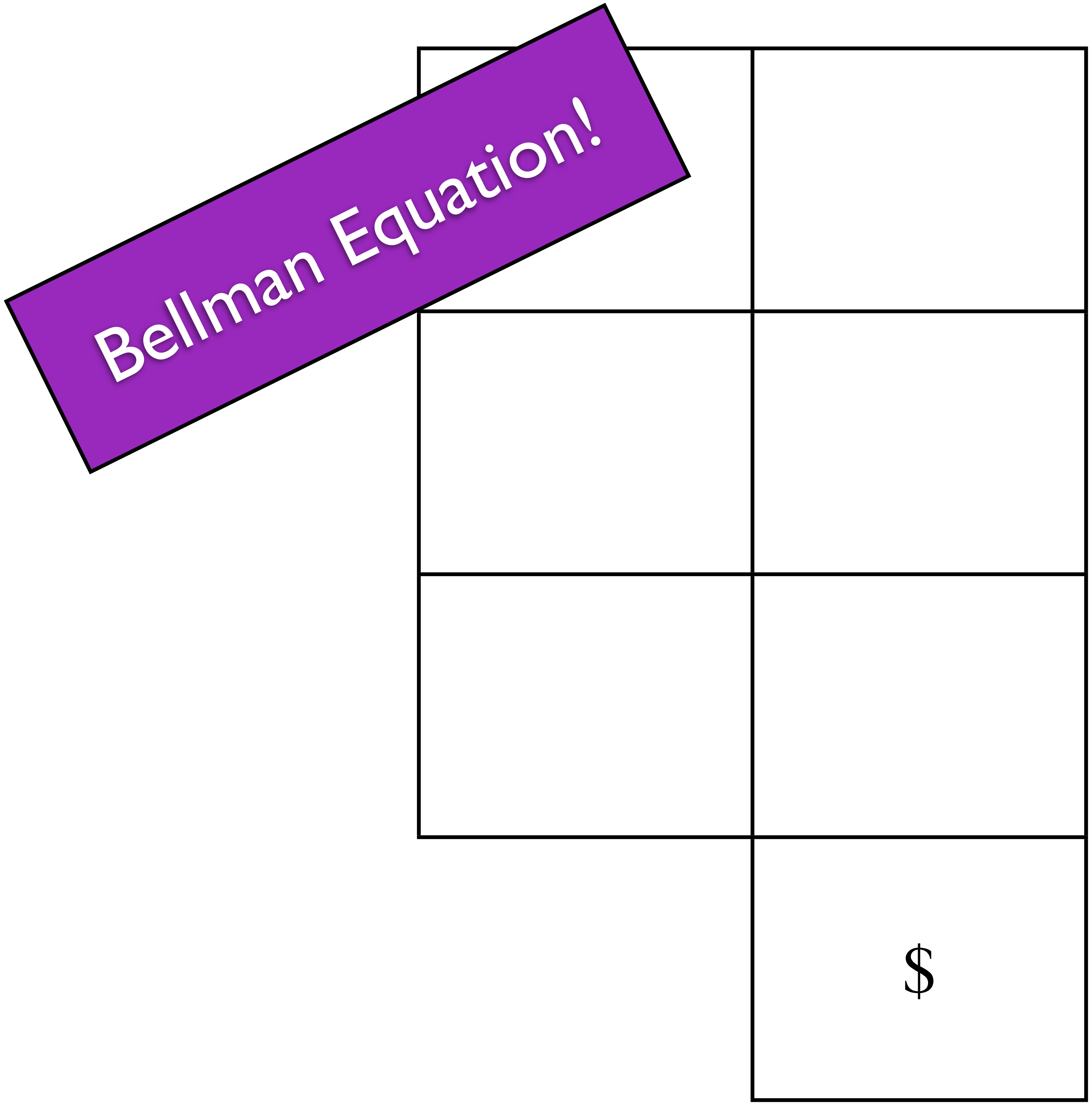
Kalai & Kalai

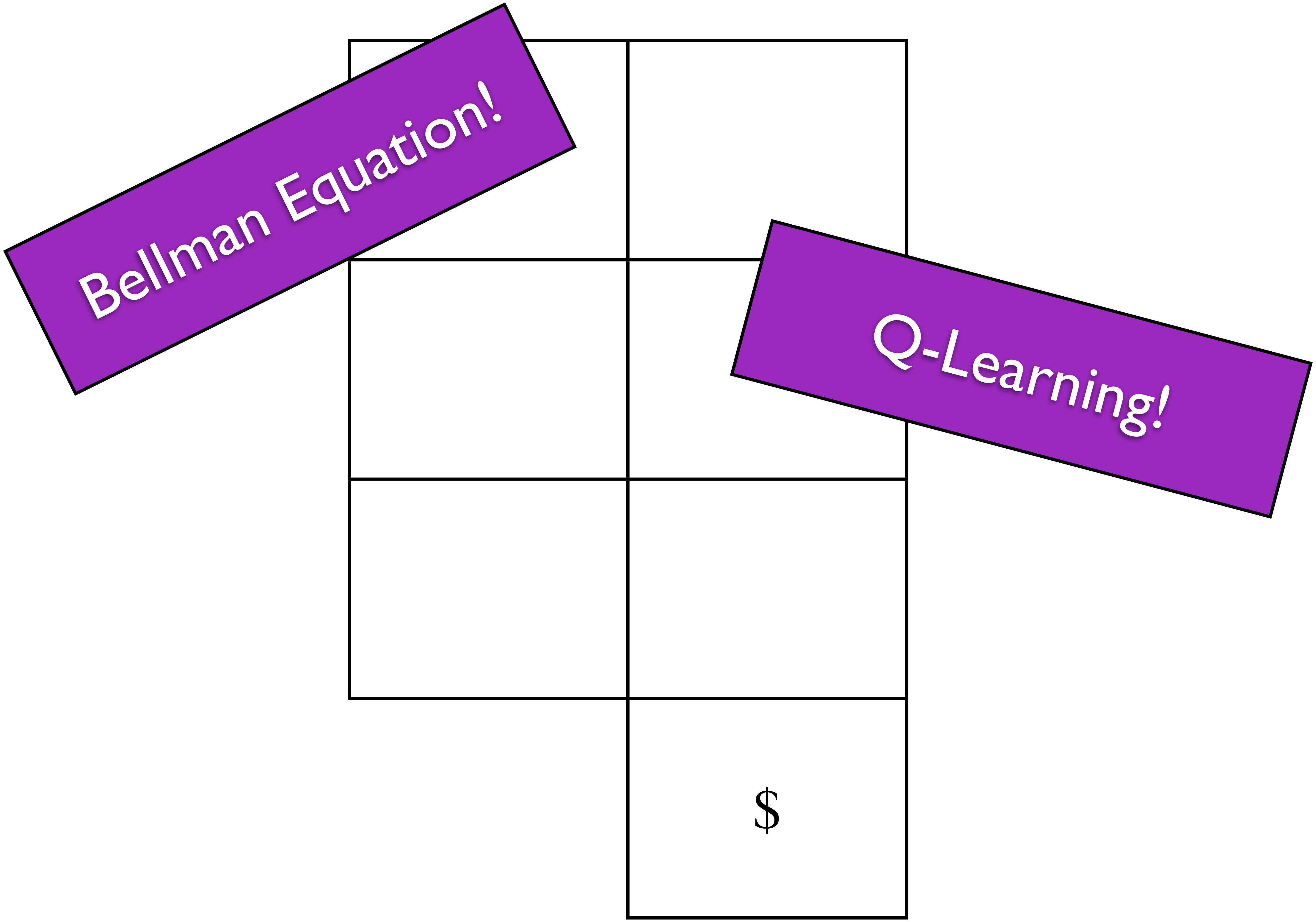
Adam Abeshouse, Betsy Hilliard and Eric Sodomka

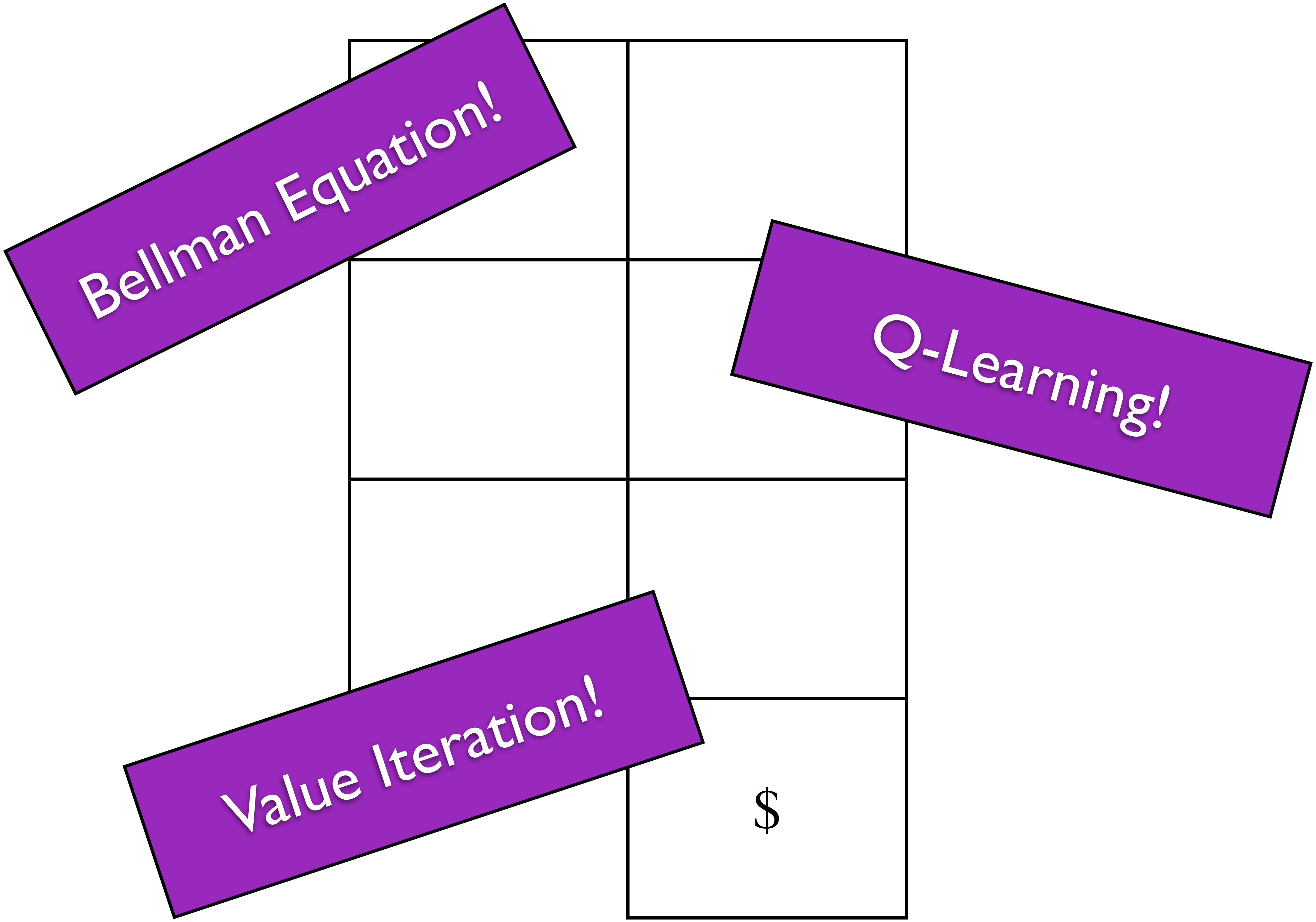
December 12, 2012

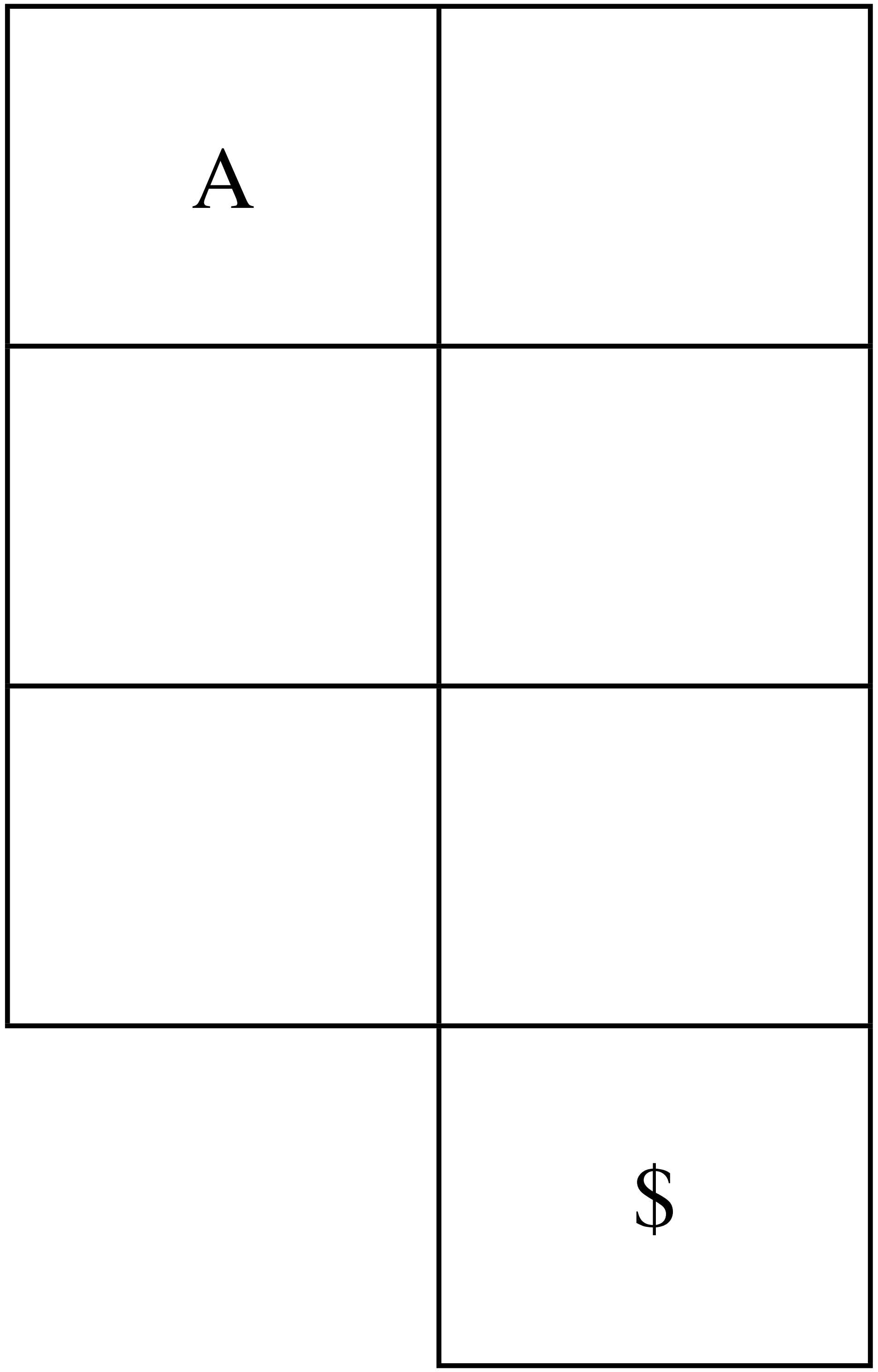

A
$

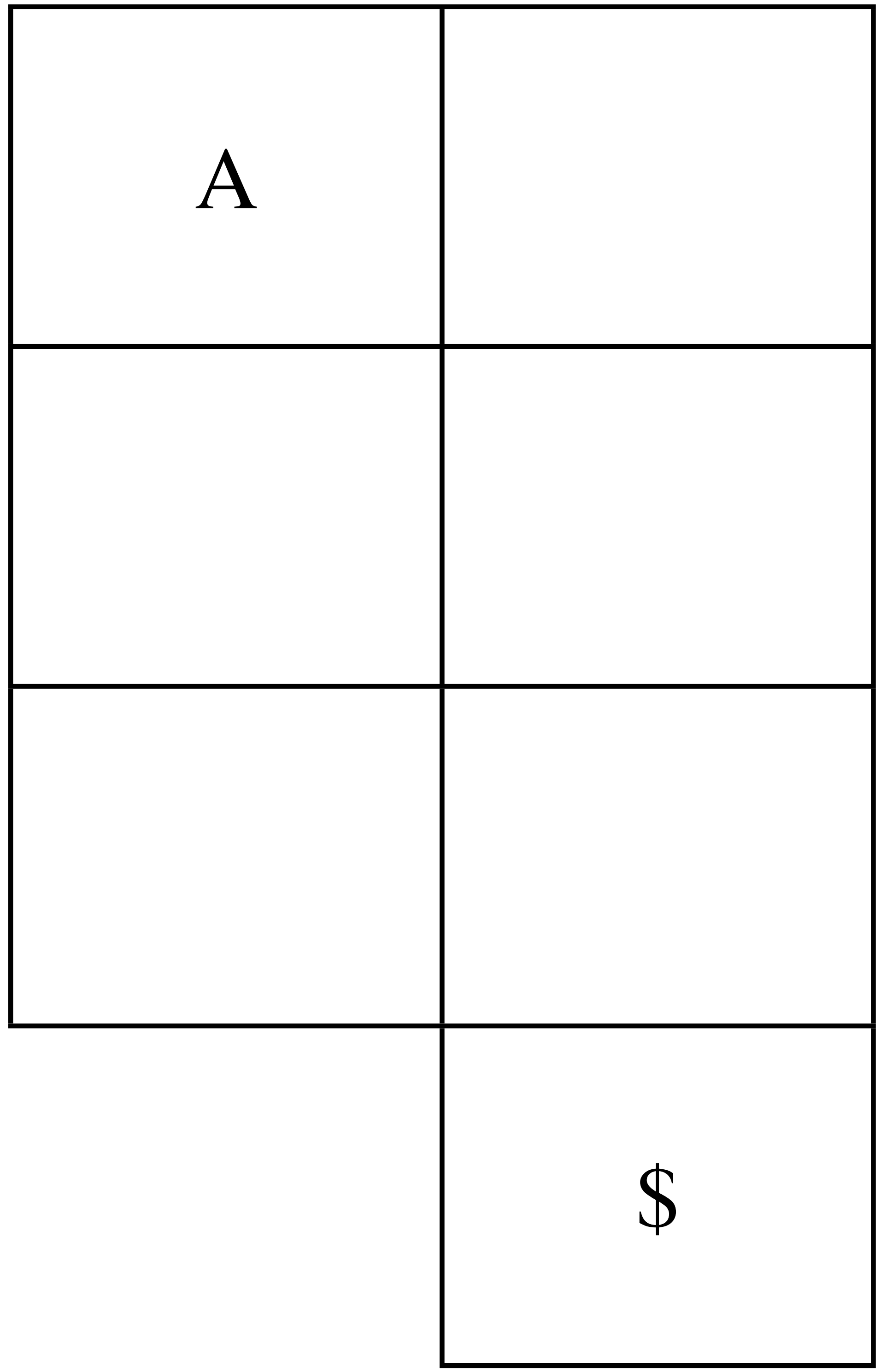

A
$

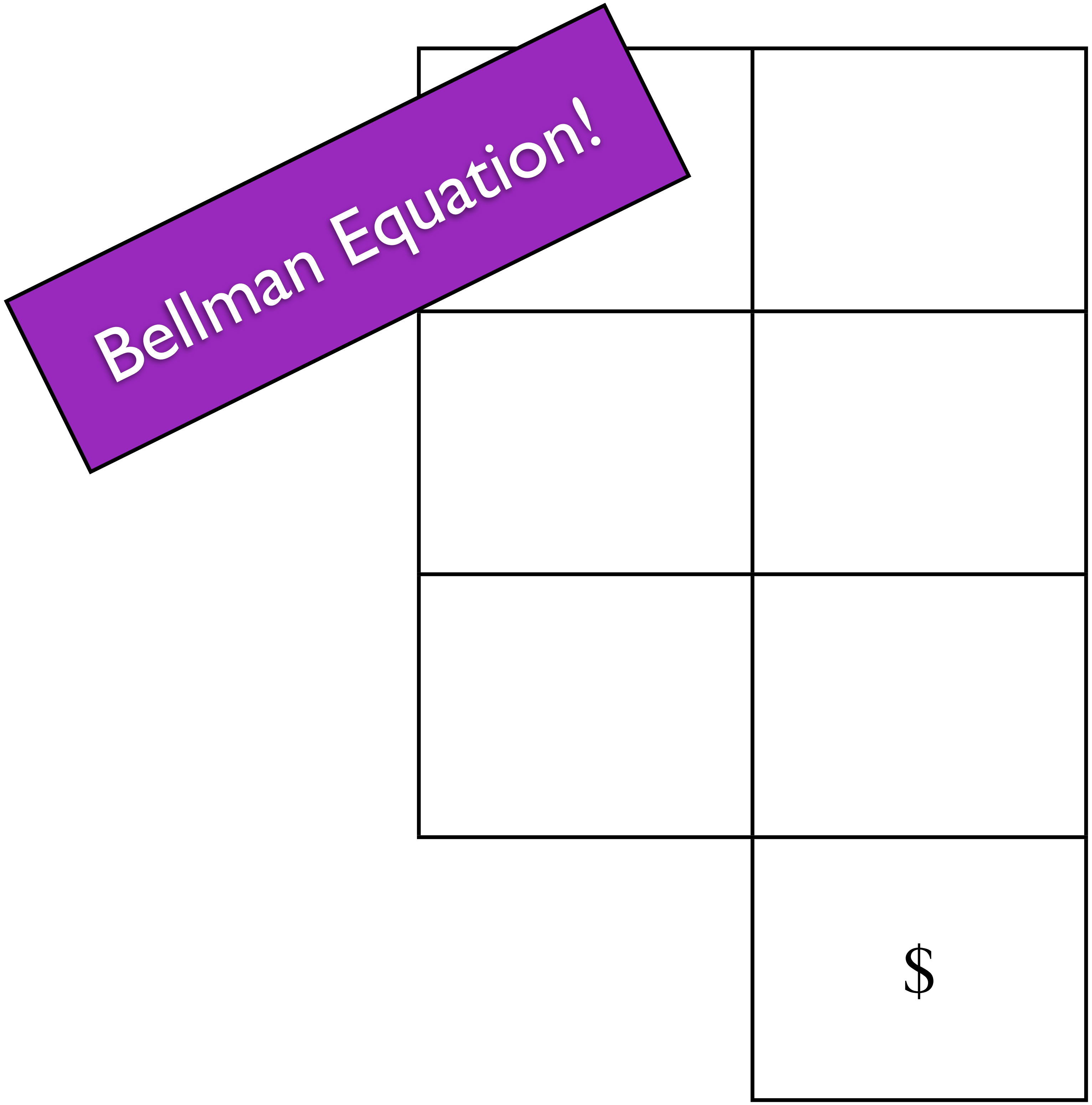

Bellman Equation!
$

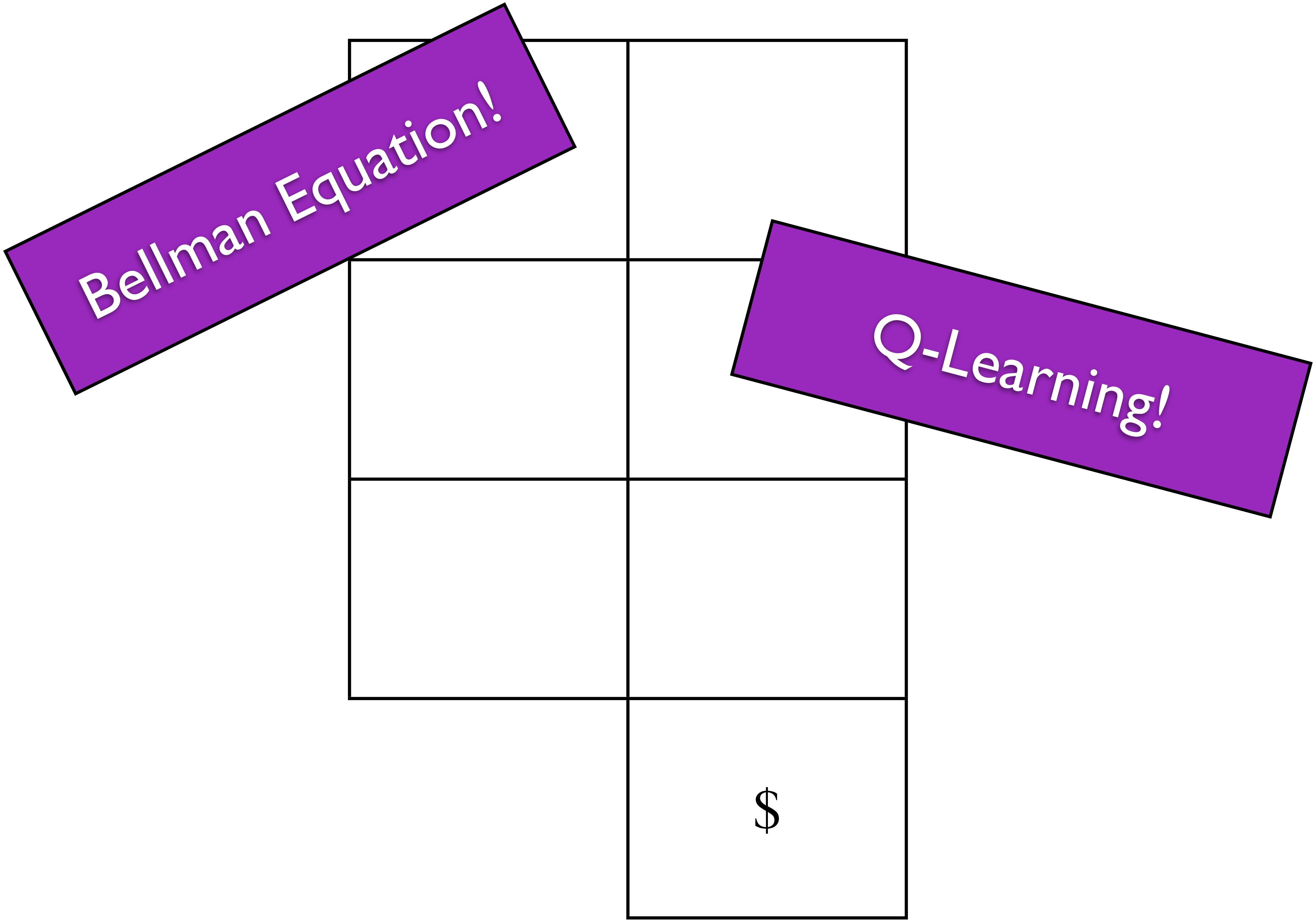

Bellman Equation!
Q-Learning!
$

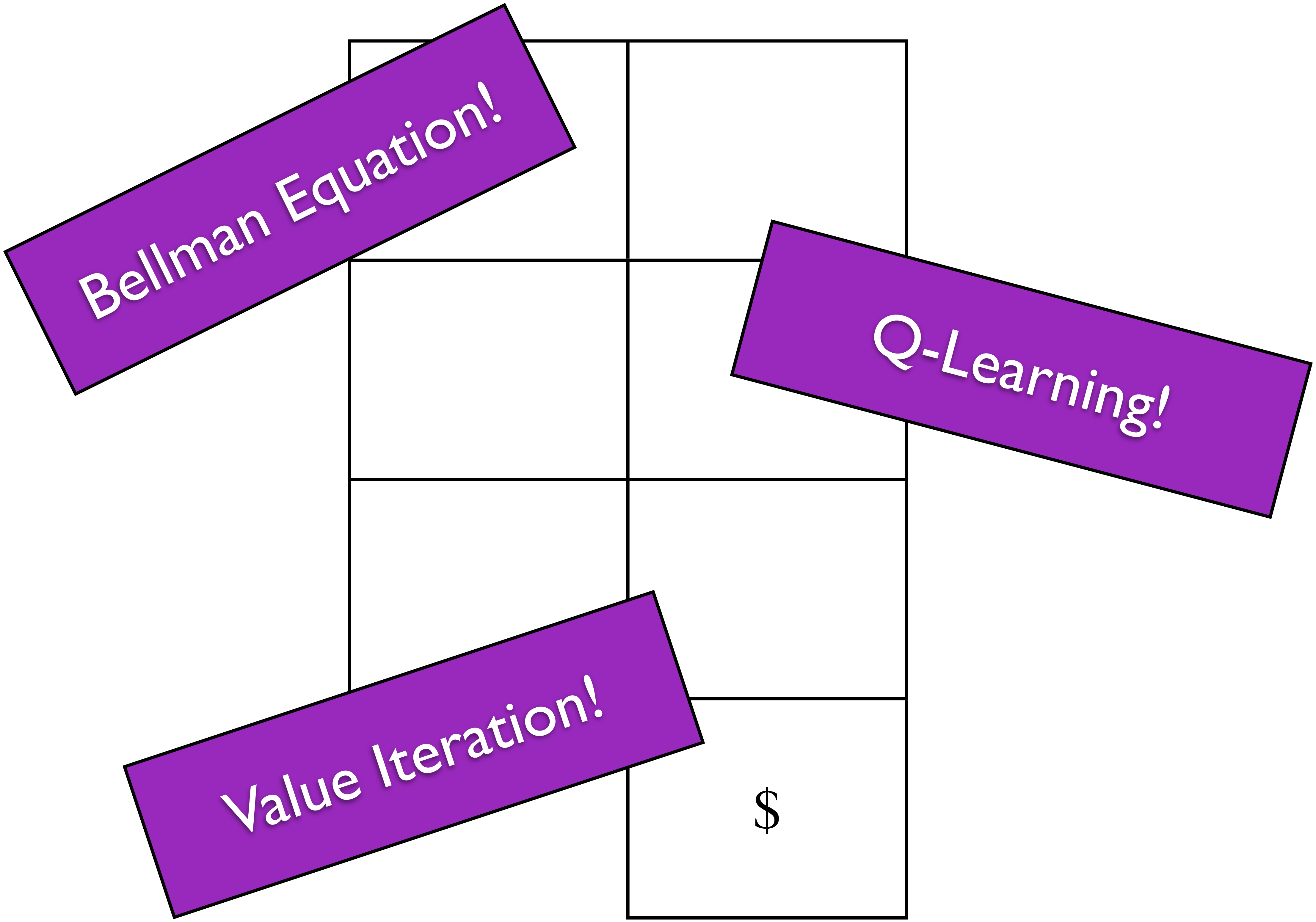

Bellman Equation!
Q-Learning!
Value Iteration!
$


A
$

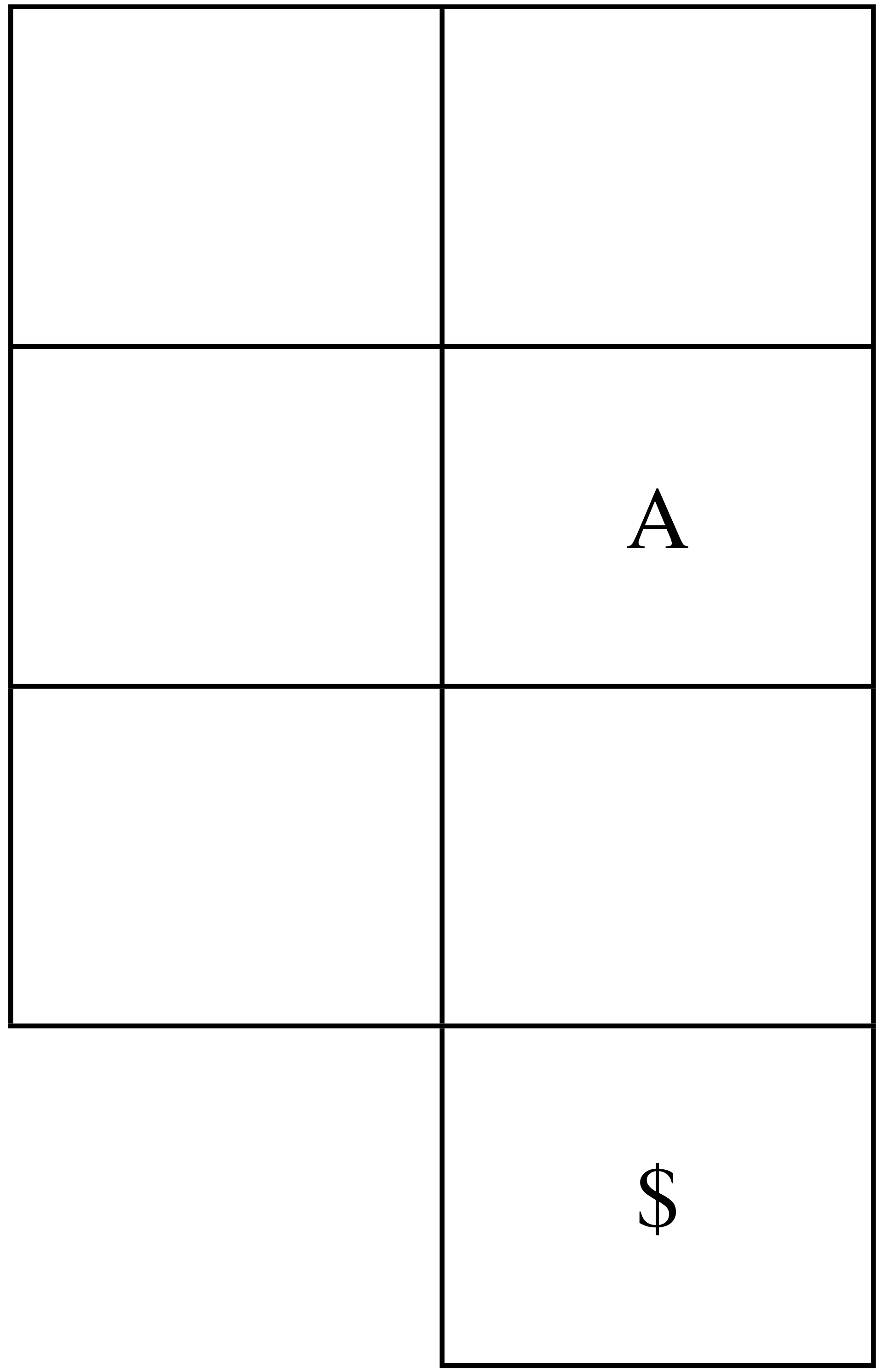

A
$


Bellman Equation!
$

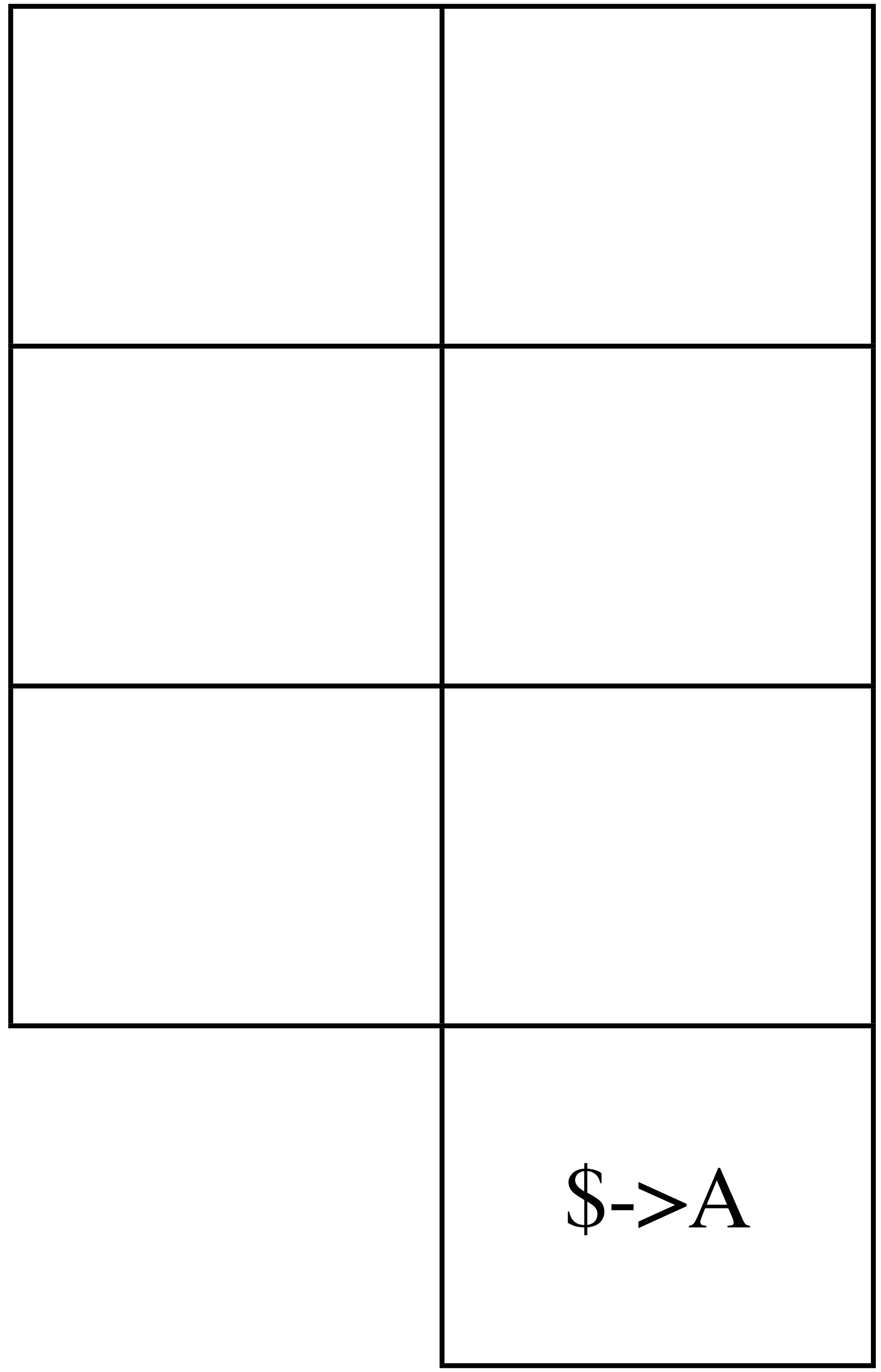

Bellman Equation!
Q-Learning!
$

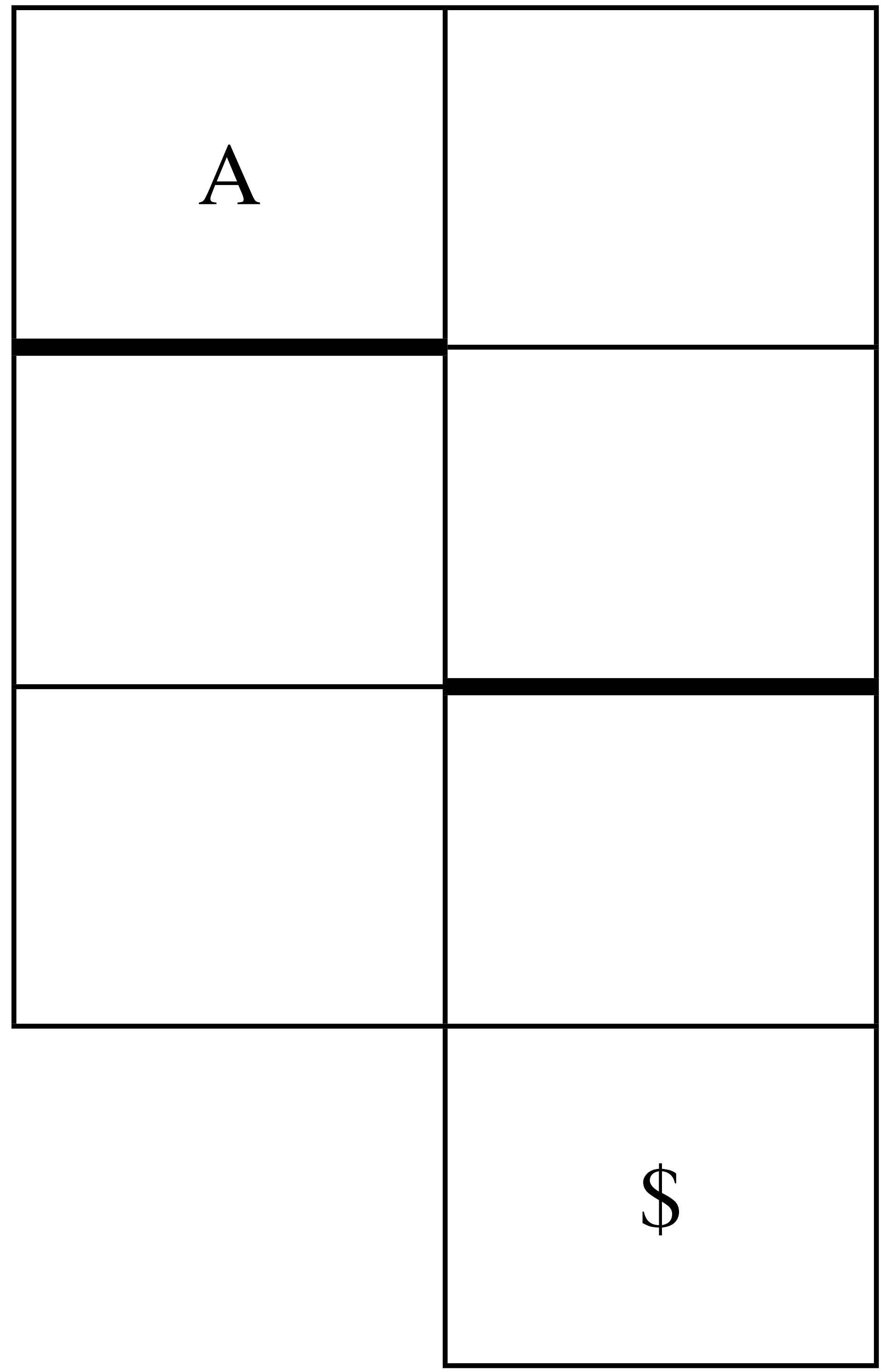

Bellman Equation!
Q-Learning!
Value Iteration!
$

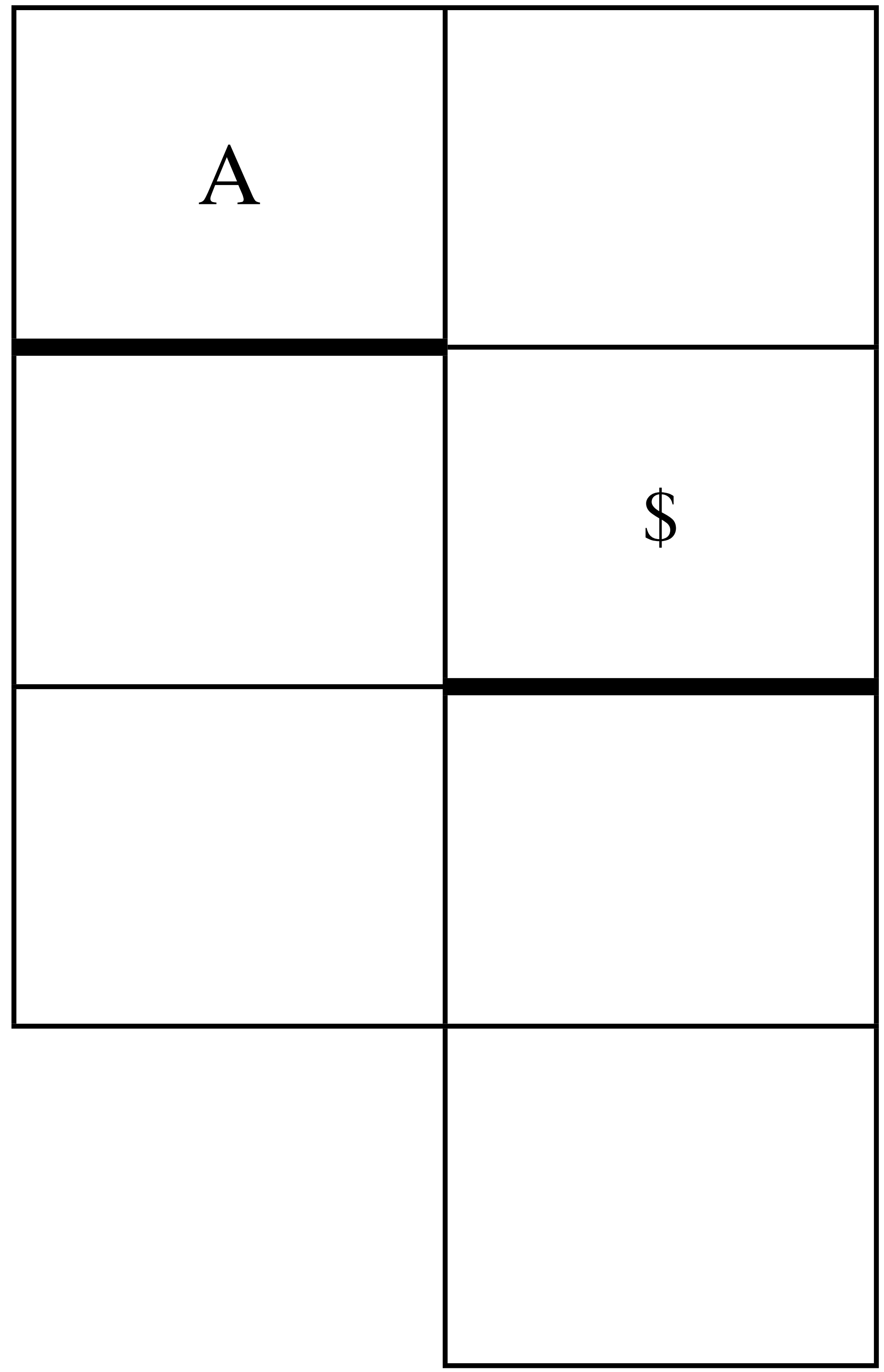
	A
	$


A
$

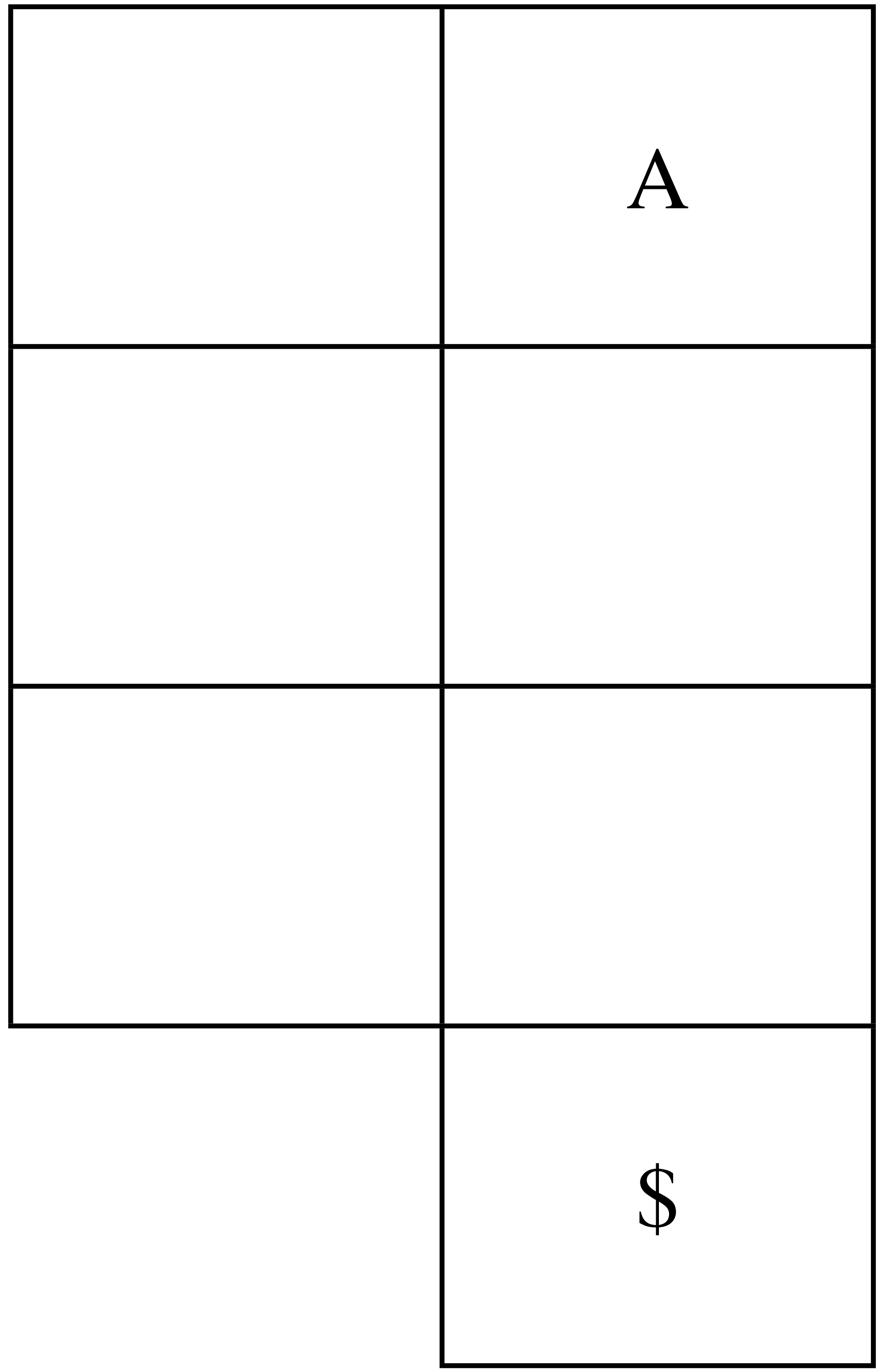
	A
	$

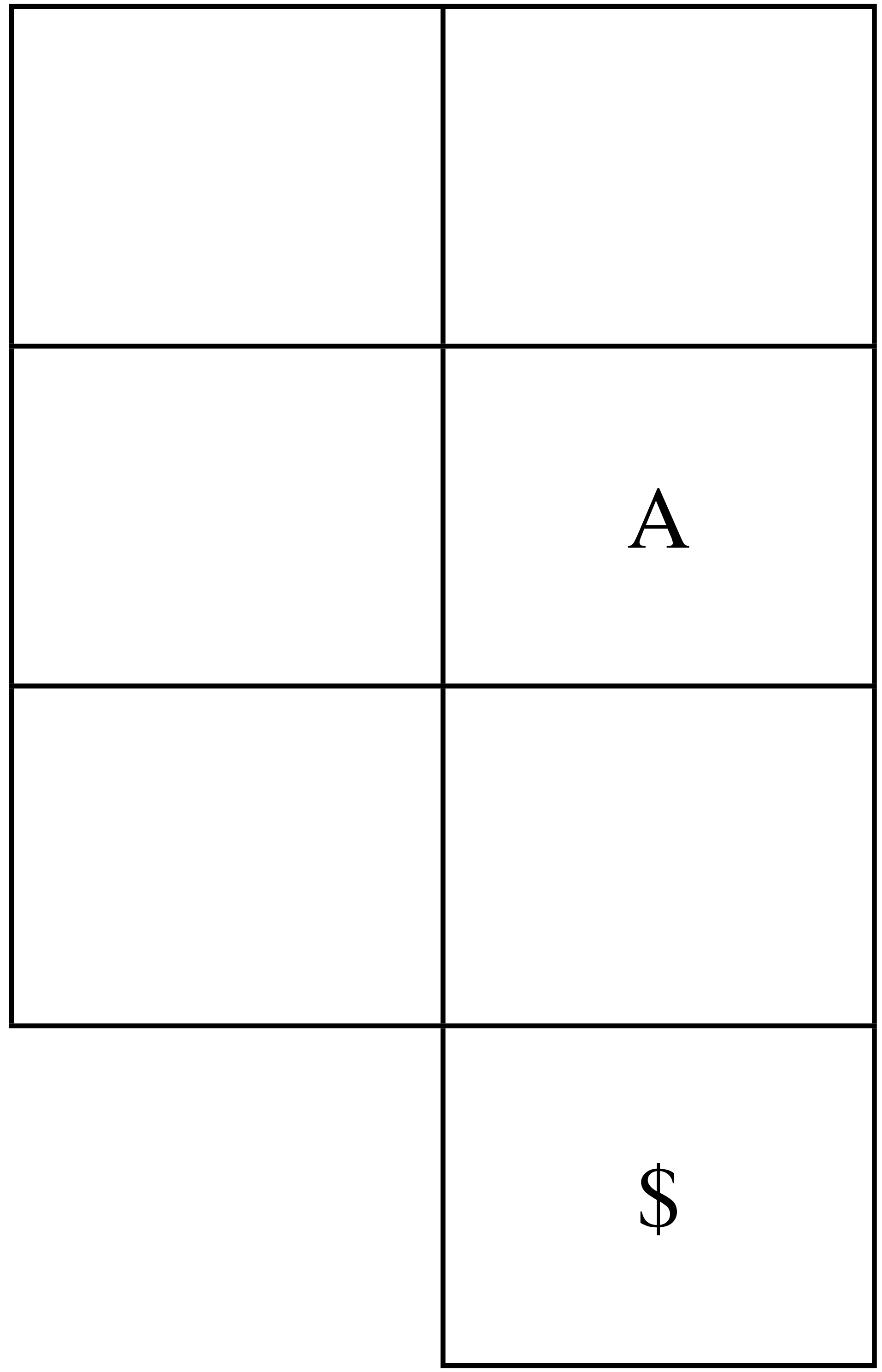

$->A

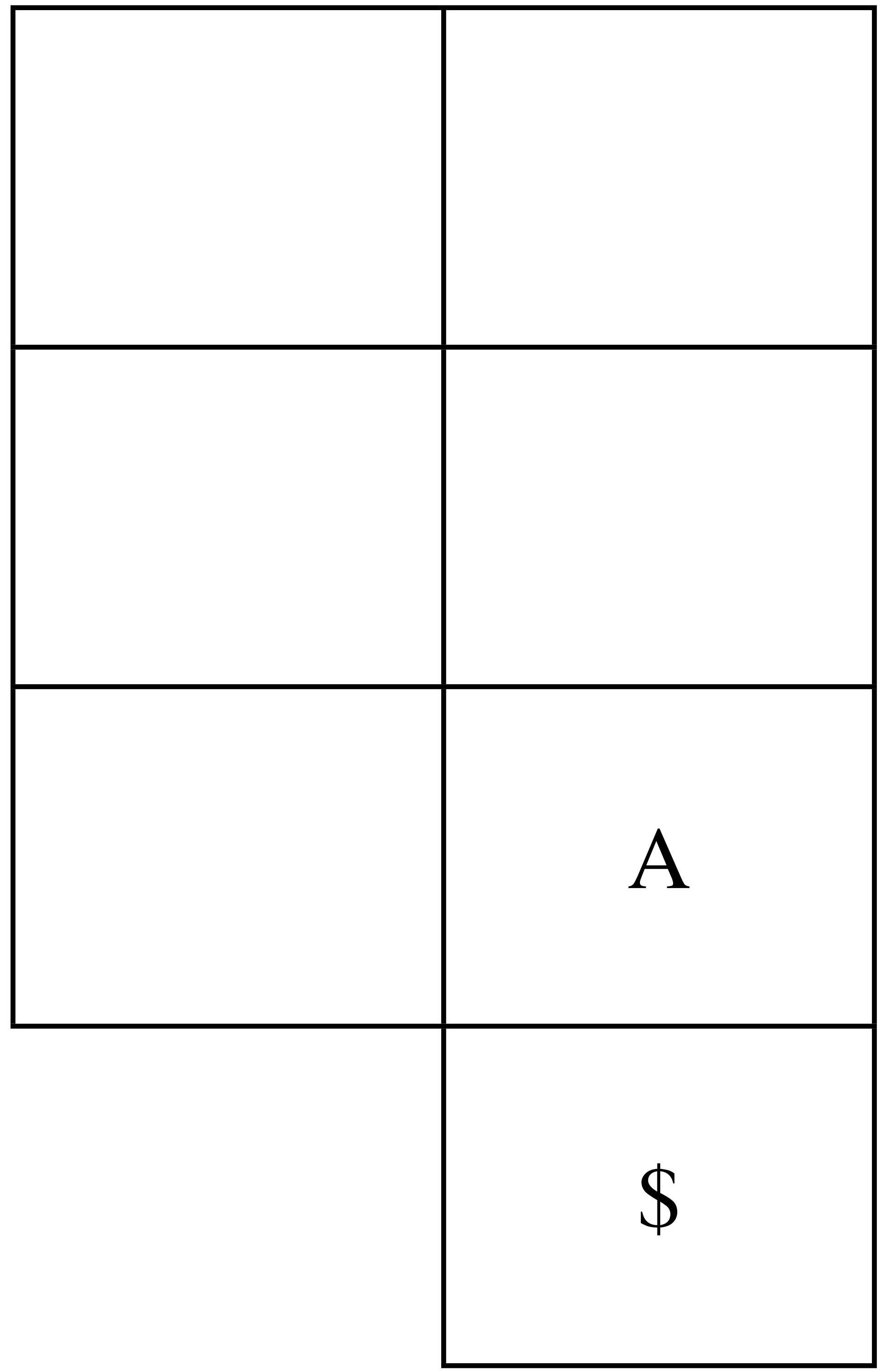

A
$

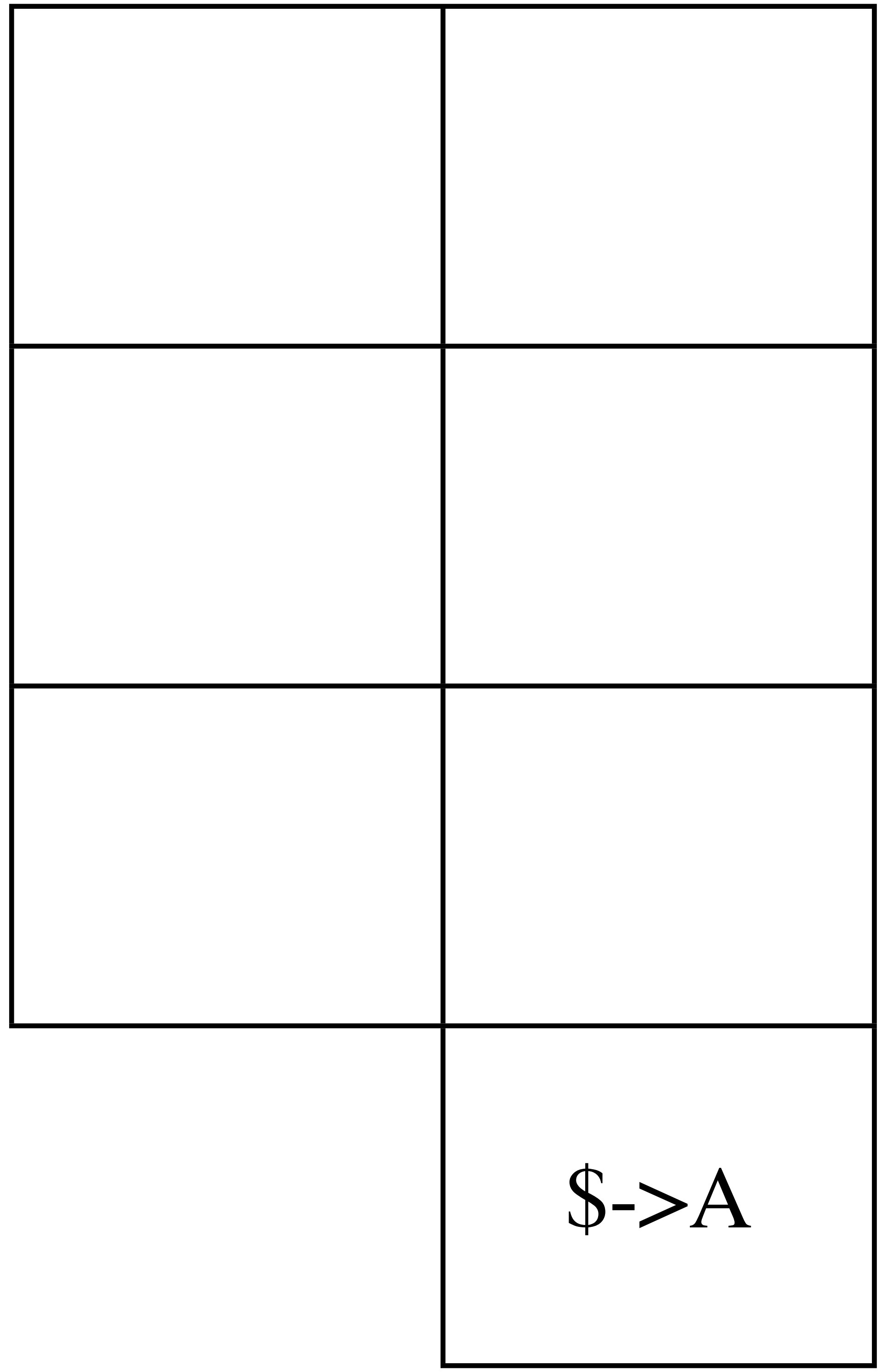

A
$

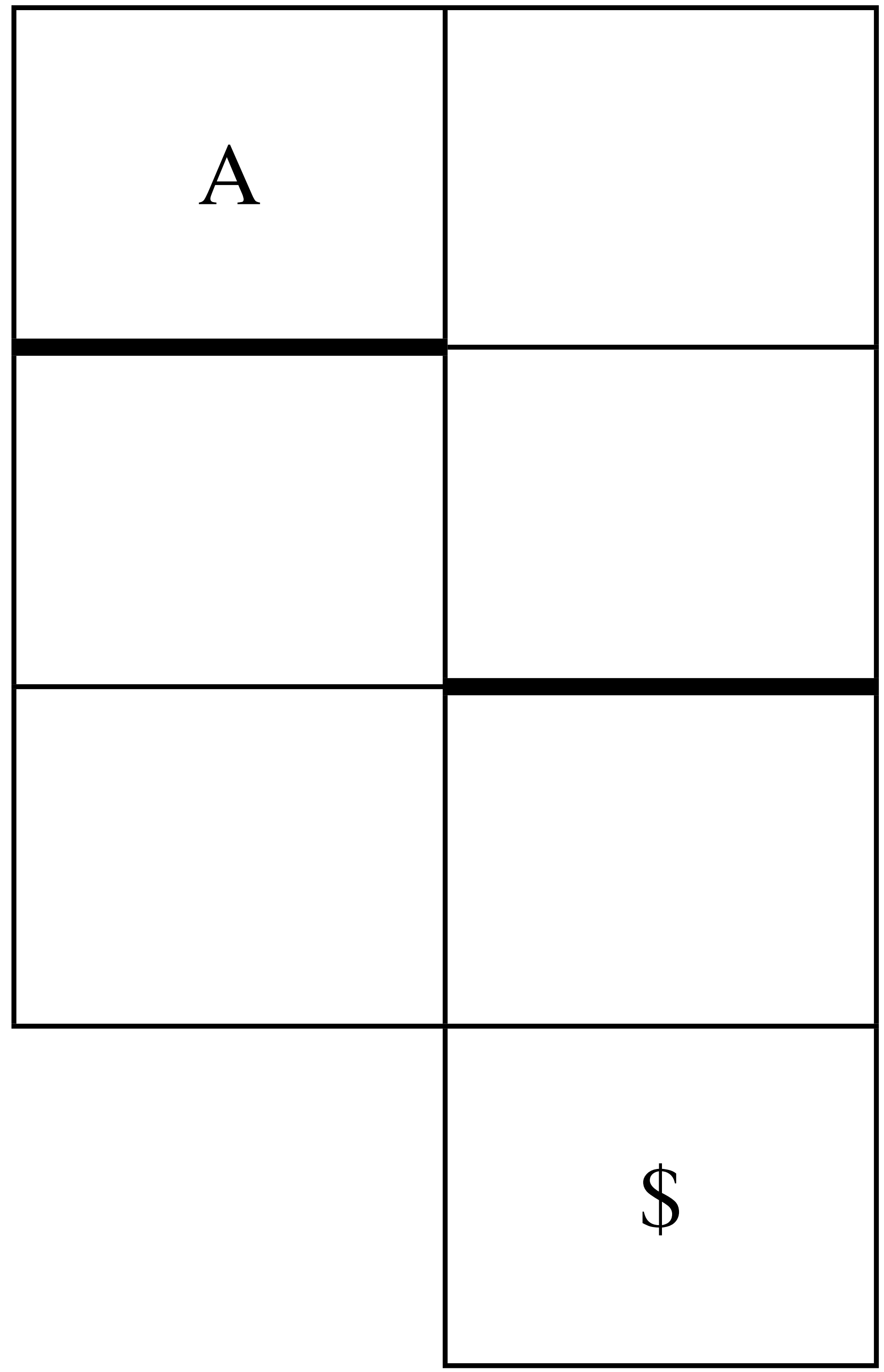
A	B $a
	$a
	$b


A
$


A
$


A
$


$->A


A
$

A	B $a
	$a
	$b

Correlated-Q Learning

Greenwald and Hall, 2003

- Describes an algorithm to extend Q-learning (and VI) to multiagent settings
- Uses "solution concepts" because "max" no longer makes sense
  - Friend/F

    proof of convergance

But agents in the real world have more options...

  - Correlated (Greenwald, Hall, 2003) no proof of convergance

Hot Dog Stand Game

B

A	beach	park
park	120, 40	240, 200
beach	600, 80	300, 100

Coco Value

B

A	beach	park
park	120, 40	240, 200
beach	600, 80	300, 100

$$\frac{A+B}{2}, \frac{A+B}{2}$$

Cooperative

	beach	park
park	130, 130	220, 220
beach	340, 340	200, 200

$$\frac{A-B}{2}, \frac{B-A}{2}$$

Competitive

	beach	park
park	45, -45	20, -20
beach	270, -270	100, -100

Coco Value Solution

Cooperative

	beach	park
park	130, 130	220, 220
beach	340, 340	200, 200

COCO value =(340+100, 340+-100)=(440,240)

But one could calculate many such values...

park	45, -45	20, -20
beach	270, -270	100, -100

A

		park
park	120, 40	240, 200
beach	600, 80	300, 100

A then pays B $240

Cooperation In Strategic Games, Revisited

Kalai and Kalai, 2011

Axioms

1. Pareto Optimality: max sum of values

2. Payoff dominance: if an agent always dominate, value greater

3. Shift invariance: additive shift by c changes value by c

4. Redundant mixed strategies: convex combinations of moves can be removed with no effect on the value

5. Monotonicity in strategies: adding a strategy can only increase your value

Also: unique and efficiently computable and the **only** value that satisfies 1-5

Original Table:

Figure 2. Convergence in the grid games: all algorithms are converging. The CE-Q algorithm shown is uCE-Q.

Grid Games	GG1	GG2	GG3
Algorithm			
Q			
Foe-Q			
Friend-Q			
uCE-Q			
eCE-Q			
rCE-Q			
lCE-Q			

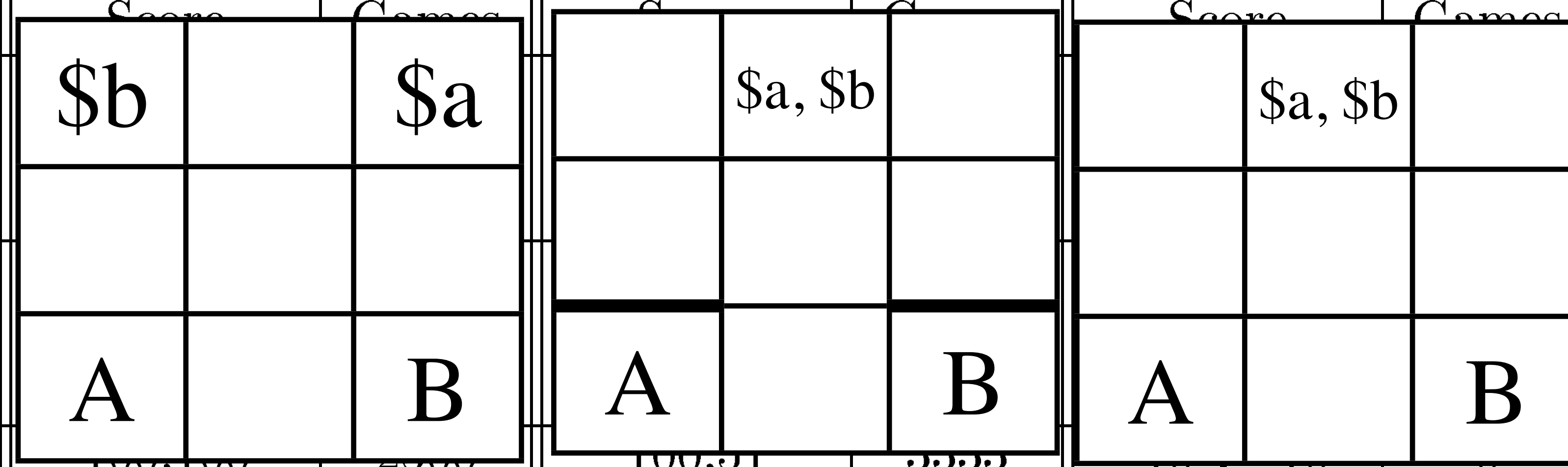


Table 2. Grid Games played repeatedly, allowing 10^4 moves. Average scores are shown. The number of games played varied with the agents' policies: some move directly to the goal, while others digress.

Original Table:

Figure 2. Convergence in the grid games: all algorithms are converging. The CE-Q algorithm shown is uCE-Q.

Grid Games	GG1		GG2		GG3	
Algorithm	Score	Games	Score	Games	Score	Games
Q	100,100	2500	49,100	3333	100,125	3333
Foe-Q	0,0	0	67,68	3003	120,120	3333
Friend-Q	$-10^4, -10^4$	0	$-10^4, -10^4$	0	$-10^4, -10^4$	0
uCE-Q	100,100	2500	50,100	3333	116,116	3333
eCE-Q	100,100	2500	51,100	3333	117,117	3333
rCE-Q	100,100	2500	100,49	3333	125,100	3333
lCE-Q	100,100	2500	100,51	3333	$-10^4, -10^4$	0

Table 2. Grid Games played repeatedly, allowing 10^4 moves. Average scores are shown. The number of games played varied with the agents' policies: some move directly to the goal, while others digress.

Original Table:

Figure 2. Convergence in the grid games: all algorithms are converging. The CE-Q algorithm shown is uCE-Q.

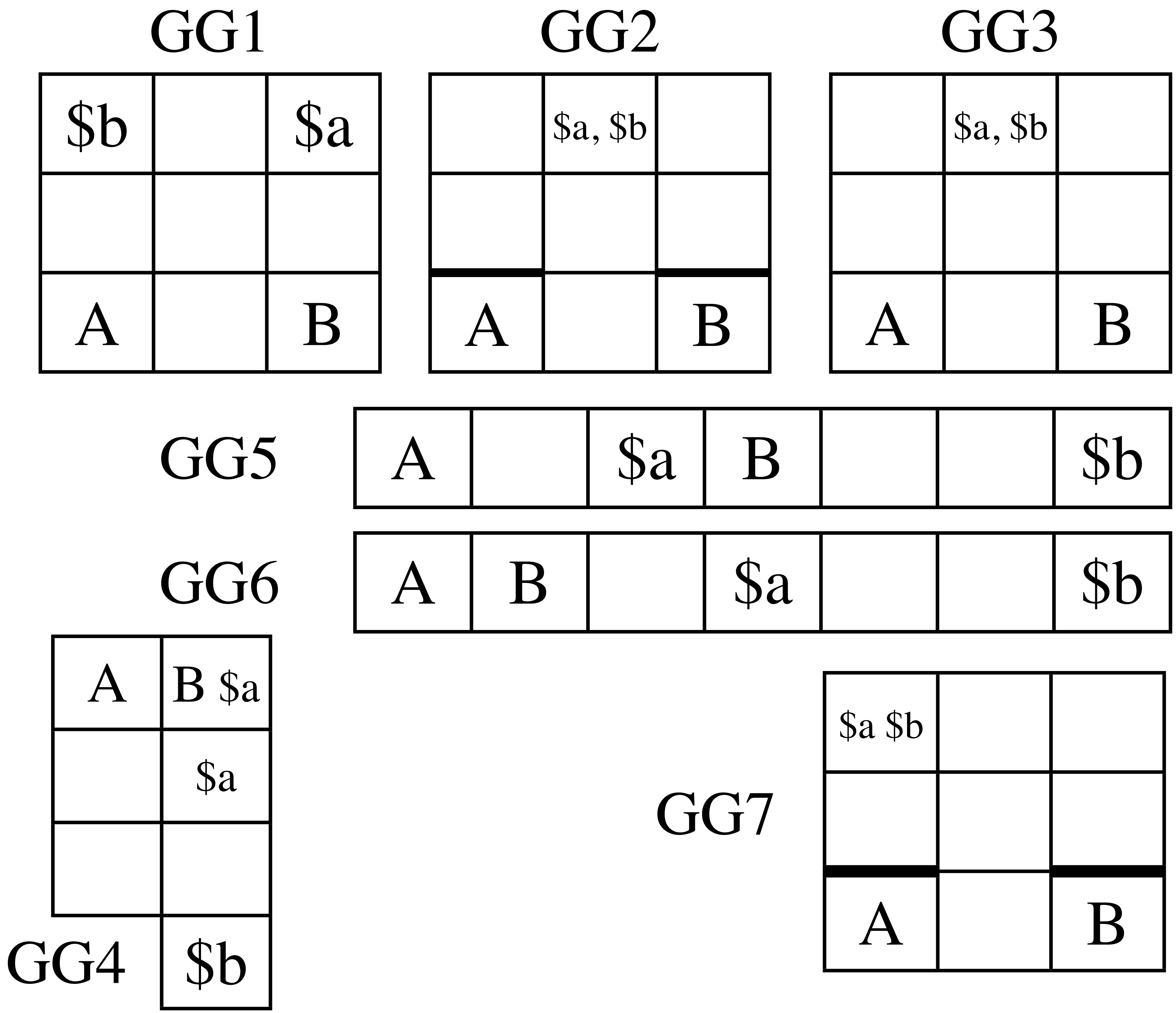
Grid Games	GG1		GG2		GG3	
Algorithm	Score	Games	Score	Games	Score	Games
Q	100,100	2500	49,100	3333	100,125	3333
Foe-Q	0,0	0	67,68	3003	120,120	3333
Friend-Q	$-10^4, -10^4$	0	$-10^4, -10^4$	0	$-10^4, -10^4$	0
uCE-Q	100,100	2500	50,100	3333	116,116	3333
eCE-Q	100,100	2500	51,100	3333	117,117	3333
rCE-Q	100,100	2500	100,49	3333	125,100	3333
lCE-Q	100,100	2500	100,51	3333	$-10^4, -10^4$	0

Table 2. Grid Games played repeatedly, allowing 10^4 moves. Average scores are shown. The number of games played varied with the agents' policies: some move directly to the goal, while others digress.

Our Table:

Grid Games	GG1		GG2		GG3	
Algorithm	Score	Games	Score	Games	Score	Games
Friend-Q	$-10^4, -10^4$	0	$-10^4, -10^4$	0	$-10^4, -10^4$	0
uCE-Q	100,100	2500	50,100	3333	117,117	3333
eCE-Q	100,100	2500	100,50	3333	117,117	3333
rCE-Q	100,100	2500	49,100	3333	100,125	3333
lCE-Q	100,100	2500	52, 100	3333	$-10^4, -10^4$	0

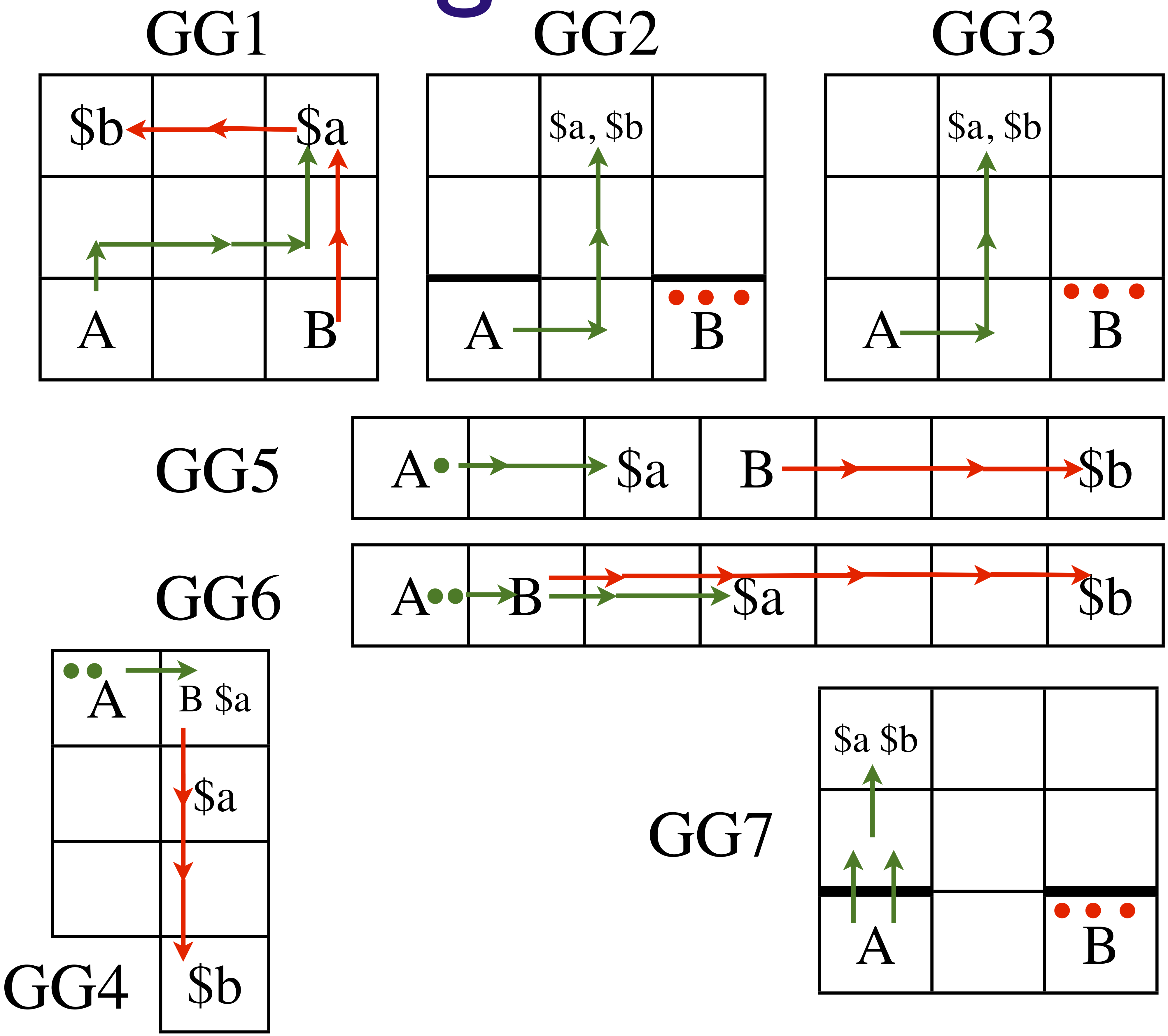
Grid Games


GG1
$b
$a
A
B
GG2
$a, $b
A
B
GG3
$a, $b
A
B
GG5
A
$a
B
$b
GG6
A
B
$a
$b
A
B $a
$a
GG4
$b
$a $b
GG7
A
B

Coco Agents for Grid Games

Coco Results

Coco Agents' Policies



Comparing Solution Concepts

Future work

- What about more than 2 agents?
- Coco in the real world:
  - asymmetric transfer payments?
  - coco vs. negotiation?
- What if humans are allowed (encouraged? forced?) to play games with coco values?

Citations and References

- http://www.istockphoto.com/stock-illustration-14738687-family-fun-beach-hot-dog-stand-couple-serving-to-children.php

- http://mayordave.org.uk/en/article/2010/430632/scores-on-the-doors-puts-knowledge-in-the-hands-of-local-restaurant-goers

[1] Enrique Munoz de Cote and Michael L. Littman. A polynomial-time nash equilibrium algorithm for repeated stochastic games. *CoRR*, abs/1206.3277, 2012.

[2] Amy Greenwald and Keith Hall. Correlated-q learning. In *In AAAI Spring Symposium*, pages 242–249. AAAI Press, 2003.

[3] Amy Greenwald, Keith Hall, and Martin Zinkevich. Correlated q-learning, 2005.

[4] Adam Kalai and Ehud Kalai. Cooperation in two person games, revisited. *SIGecom Exch.*, 10(1):13–16, March 2011.

[5] Adam Tauman Kalai and Ehud Kalai. Cooperation and competition in strategic games with private information. In *Proceedings of the 11th ACM conference on Electronic commerce*, EC '10, pages 345–346, New York, NY, USA, 2010. ACM.

[6] Adam Tauman Kalai and Ehud Kalai. A cooperative value for bayesian games, 2010.