Multi-agent learning

Teaching strategies

Gerard Vreeswijk, Intelligent Software Systems, Computer Science Department, Faculty of Sciences, Utrecht University, The Netherlands.

Monday 14th June, 2021

Plan for Today

Part I: Preliminaries

Plan for Today

Part I: Preliminaries

1. Teacher possesses memory of $k = 0$ rounds: Bully

Plan for Today

Part I: Preliminaries

1. Teacher possesses memory of $k = 0$ rounds: Bully
2. Teacher possesses memory of $k = 1$ round: Godfather

Plan for Today

Part I: Preliminaries

1. Teacher possesses memory of $k = 0$ rounds: Bully
2. Teacher possesses memory of $k = 1$ round: Godfather
3. Teacher possesses memory of $k > 1$ rounds: {lenient, strict} Godfather

Plan for Today

Part I: Preliminaries

1. Teacher possesses memory of $k = 0$ rounds: Bully
2. Teacher possesses memory of $k = 1$ round: Godfather
3. Teacher possesses memory of $k > 1$ rounds: {lenient, strict} Godfather
4. Teacher is represented by a finite machine: Godfather++

Plan for Today

Part I: Preliminaries

1. Teacher possesses memory of $k = 0$ rounds: Bully
2. Teacher possesses memory of $k = 1$ round: Godfather
3. Teacher possesses memory of $k > 1$ rounds: {lenient, strict} Godfather
4. Teacher is represented by a finite machine: Godfather++

Part II: Crandall & Goodrich (2005) SPaM: an algorithm that claims to integrate follower and teacher algorithms.

Plan for Today

Part I: Preliminaries

1. Teacher possesses memory of $k = 0$ rounds: Bully
2. Teacher possesses memory of $k = 1$ round: Godfather
3. Teacher possesses memory of $k > 1$ rounds: {lenient, strict} Godfather
4. Teacher is represented by a finite machine: Godfather++

Part II: Crandall & Goodrich (2005) SPaM: an algorithm that claims to integrate follower and teacher algorithms.

Plan for Today

Part I: Preliminaries

1. Teacher possesses memory of $k = 0$ rounds: Bully
2. Teacher possesses memory of $k = 1$ round: Godfather
3. Teacher possesses memory of $k > 1$ rounds: {lenient, strict} Godfather
4. Teacher is represented by a finite machine: Godfather++

Part II: Crandall & Goodrich (2005) SPaM: an algorithm that claims to integrate follower and teacher algorithms.

1. Three points of criticism to Godfather++.

Plan for Today

Part I: Preliminaries

1. Teacher possesses memory of $k = 0$ rounds: Bully
2. Teacher possesses memory of $k = 1$ round: Godfather
3. Teacher possesses memory of $k > 1$ rounds: {lenient, strict} Godfather
4. Teacher is represented by a finite machine: Godfather++

Part II: Crandall & Goodrich (2005) SPaM: an algorithm that claims to integrate follower and teacher algorithms.

1. Three points of criticism to Godfather++.
2. Core idea of SPaM: combine teacher and follower capabilities.

Plan for Today

Part I: Preliminaries

1. Teacher possesses memory of $k = 0$ rounds: Bully
2. Teacher possesses memory of $k = 1$ round: Godfather
3. Teacher possesses memory of $k > 1$ rounds: {lenient, strict} Godfather
4. Teacher is represented by a finite machine: Godfather++

Part II: Crandall & Goodrich (2005) SPaM: an algorithm that claims to integrate follower and teacher algorithms.

1. Three points of criticism to Godfather++.
2. Core idea of SPaM: combine teacher and follower capabilities.
3. Notion of guilt to trigger switches between teaching and following.

Literature

Literature

Michael L. Littman and Peter Stone (2001). "Leading best-response strategies in repeated games". Research note.

One of the first papers, if not the first paper, that mentions Bully and Godfather.

Literature

Michael L. Littman and Peter Stone (2001). "Leading best-response strategies in repeated games". Research note.

One of the first papers, if not the first paper, that mentions Bully and Godfather.

Michael L. Littman and Peter Stone (2005). "A polynomial-time Nash equilibrium algorithm for repeated games". In *Decision Support Systems* Vol. 39, pp. 55-66.

Paper that describes Godfather++.

Literature

Michael L. Littman and Peter Stone (2001). "Leading best-response strategies in repeated games". Research note.

One of the first papers, if not the first paper, that mentions Bully and Godfather.

Michael L. Littman and Peter Stone (2005). "A polynomial-time Nash equilibrium algorithm for repeated games". In *Decision Support Systems* Vol. 39, pp. 55-66.

Paper that describes Godfather++.

Jacob W. Crandall and Michael A. Goodrich (2005). "Learning to teach and follow in repeated games". In *AAAI Workshop on Multiagent Learning*, Pittsburgh, PA.

Paper that attempts to combine Fictitious Play and a modified Godfather++ to define an algorithm that "knows" when to teach and when to follow.

Literature

Michael L. Littman and Peter Stone (2001). "Leading best-response strategies in repeated games". Research note.

One of the first papers, if not the first paper, that mentions Bully and Godfather.

Michael L. Littman and Peter Stone (2005). "A polynomial-time Nash equilibrium algorithm for repeated games". In *Decision Support Systems* Vol. 39, pp. 55-66.

Paper that describes Godfather++.

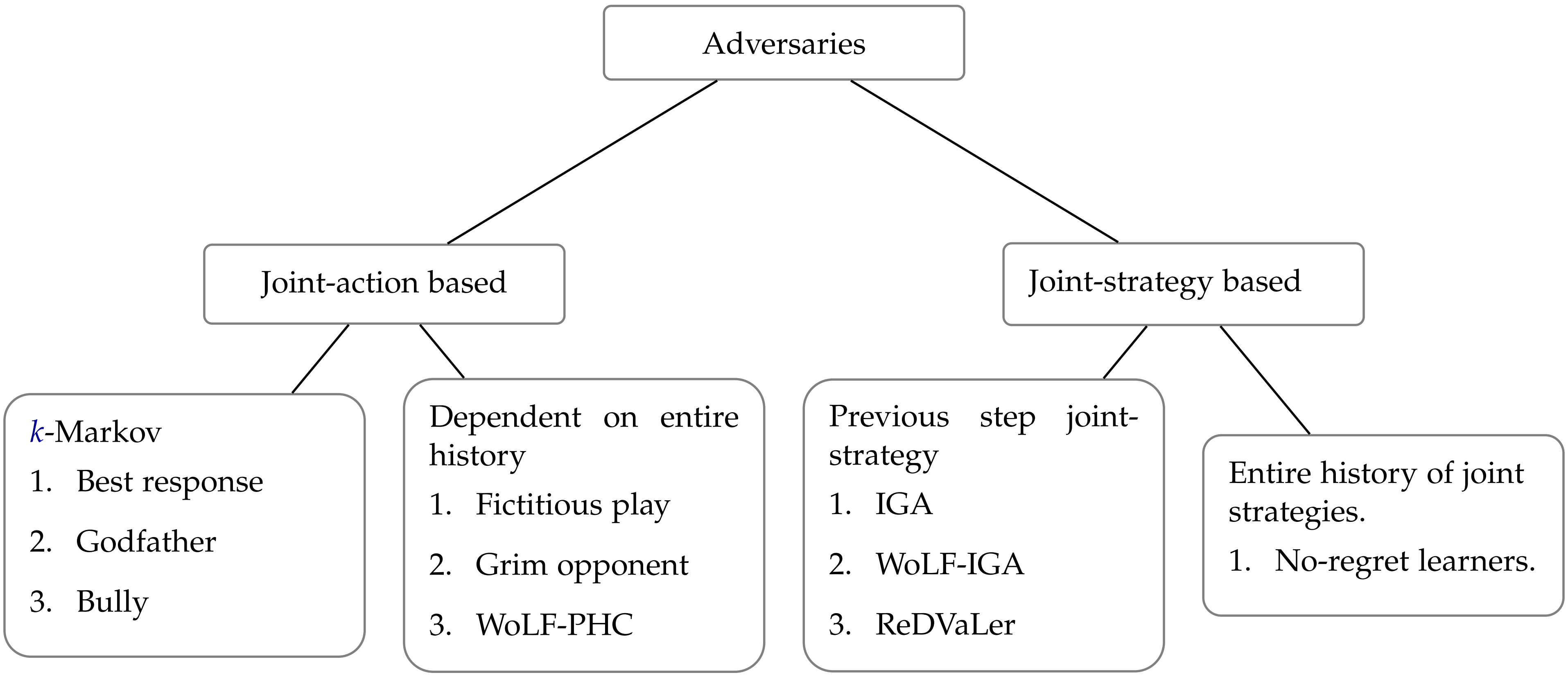
Jacob W. Crandall and Michael A. Goodrich (2005). "Learning to teach and follow in repeated games". In *AAAI Workshop on Multiagent Learning*, Pittsburgh, PA.

Paper that attempts to combine Fictitious Play and a modified Godfather++ to define an algorithm that "knows" when to teach and when to follow.

Doran Chakraborty and Peter Stone (2008). "Online Multiagent Learning against Memory Bounded Adversaries," *Machine Learning and Knowledge Discovery in Databases*, Lecture Notes in Artificial Intelligence Vol. 5212, pp. 211-26

Taxonomy of possible adversaries

(Taken from Chakraborty and Stone, 2008):



Bully

Bully

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Bully

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Example of finding a pure Bully strategy:

Bully

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Example of finding a pure Bully strategy:

$$\begin{array}{cc} & \begin{array}{ccc} L & M & R \end{array} \\ \begin{array}{c} T \\ C \\ B \end{array} & \begin{pmatrix} 3,6 & 8,6 & 7,3 \\ 8,1 & 6,3 & 7,3 \\ 3,5 & 9,2 & 7,5 \end{pmatrix} \end{array}$$

Bully

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Example of finding a pure Bully strategy:

$$\begin{array}{cc} & \begin{array}{ccc} L & M & R \end{array} \\ \begin{array}{c} T \\ C \\ B \end{array} & \begin{pmatrix} 3,6 & 8,6 & 7,3 \\ 8,1 & 6,3 & 7,3 \\ 3,5 & 9,2 & 7,5 \end{pmatrix} \end{array}$$

1. Find, for every action of yourself, the best response(s) of your opponent.

Bully

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Example of finding a pure Bully strategy:

$$\begin{array}{c c} & \begin{array}{ccc} L & M & R \end{array} \\ \begin{array}{c} T \\ C \\ B \end{array} & \begin{pmatrix} 3,6 & 8,6 & 7,3 \\ 8,1 & 6,3 & 7,3 \\ 3,5 & 9,2 & 7,5 \end{pmatrix} \end{array}$$

1. Find, for every action of yourself, the best response(s) of your opponent.

   This yields L and M for T, M and R for C and L and R for B.

Bully

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Example of finding a pure Bully strategy:

$$\begin{array}{c} \\ T \\ C \\ B \end{array} \begin{array}{c} \begin{array}{ccc} \;\;L\;\; & \;\;M\;\; & \;\;R\;\; \end{array} \\ \begin{pmatrix} 3,6 & 8,6 & 7,3 \\ 8,1 & 6,3 & 7,3 \\ 3,5 & 9,2 & 7,5 \end{pmatrix} \end{array}$$

1. Find, for every action of yourself, the best response(s) of your opponent.

   This yields L and M for T, M and R for C and L and R for B.

2. For these opponents actions, you'll receive

Bully

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Example of finding a pure Bully strategy:

$$\begin{array}{cc} & \begin{array}{ccc} L & M & R \end{array} \\ \begin{array}{c} T \\ C \\ B \end{array} & \left(\begin{array}{ccc} 3,6 & 8,6 & 7,3 \\ 8,1 & 6,3 & 7,3 \\ 3,5 & 9,2 & 7,5 \end{array} \right) \end{array}$$

1. Find, for every action of yourself, the best response(s) of your opponent.

   This yields L and M for T, M and R for C and L and R for B.

2. For these opponents actions, you'll receive

   3 and 8 for T, 6 and 7 for C and 3 and 9 for B.

Bully

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Example of finding a pure Bully strategy:

$$\begin{array}{c} \\ T \\ C \\ B \end{array} \begin{array}{c} \begin{array}{ccc} \text{L} & \text{M} & \text{R} \end{array} \\ \begin{pmatrix} 3,6 & 8,6 & 7,3 \\ 8,1 & 6,3 & 7,3 \\ 3,5 & 9,2 & 7,5 \end{pmatrix} \end{array}$$

1. Find, for every action of yourself, the best response(s) of your opponent.

   This yields L and M for T, M and R for C and L and R for B.

2. For these opponents actions, you'll receive

   3 and 8 for T, 6 and 7 for C and 3 and 9 for B.

   Now choose one, and only one, of the actions with a highest security value.

Bully

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Example of finding a pure Bully strategy:

$$\begin{array}{cc} & \begin{array}{ccc} \text{L} & \text{M} & \text{R} \end{array} \\ \begin{array}{c} \text{T} \\ \text{C} \\ \text{B} \end{array} & \begin{pmatrix} 3,6 & 8,6 & 7,3 \\ 8,1 & 6,3 & 7,3 \\ 3,5 & 9,2 & 7,5 \end{pmatrix} \end{array}$$

1. Find, for every action of yourself, the best response(s) of your opponent.

   This yields L and M for T, M and R for C and L and R for B.

2. For these opponents actions, you'll receive

   3 and 8 for T, 6 and 7 for C and 3 and 9 for B.

   Now choose one, and only one, of the actions with a highest security value. Here that would be C with security value 6.

Bully sometimes must be mixed

Bully sometimes must be mixed

Consider matching pennies:

	H	T
H	$(1, -1)$	$(-1, 1)$
T	$(-1, 1)$	$(1, -1)$

Bully sometimes must be mixed

Consider matching pennies:

	H	T
H	$(1, -1)$	$(-1, 1)$
T	$(-1, 1)$	$(1, -1)$

The best way to profit from a pure follower is to randomise 50%.

Bully sometimes must be mixed

Consider matching pennies:

	H	T
H	$(1,-1)$	$(-1,1)$
T	$(-1,1)$	$(1,-1)$

The best way to profit from a pure follower is to randomise 50%. Your expected payoff then is 0.5.

Bully sometimes must be mixed

Consider matching pennies:

	H	T
H	$(1, -1)$	$(-1, 1)$
T	$(-1, 1)$	$(1, -1)$

The best way to profit from a pure follower is to randomise 50%. Your expected payoff then is 0.5.

A more complicated case

Bully sometimes must be mixed

Consider matching pennies:

	H	T
H	$(1,-1)$	$(-1,1)$
T	$(-1,1)$	$(1,-1)$

The best way to profit from a pure follower is to randomise 50%. Your expected payoff then is 0.5.

A more complicated case. Suppose Row's strategy α is $(\alpha_1, \alpha_2, \alpha_3)$:

		L	C	R
α_1	T	$(0,1)$	$(0,0)$	$(4,0)$
α_2	M	$(8,0)$	$(1,1)$	$(8,0)$
α_3	B	$(4,0)$	$(0,0)$	$(0,1)$
Col's $E[\$]$		α_1	α_2	α_3

Bully sometimes must be mixed

Consider matching pennies:

	H	T
H	$(1, -1)$	$(-1, 1)$
T	$(-1, 1)$	$(1, -1)$

The best way to profit from a pure follower is to randomise 50%. Your expected payoff then is 0.5.

A more complicated case. Suppose Row's strategy α is $(\alpha_1, \alpha_2, \alpha_3)$:

		L	C	R
α_1	T	$(0, 1)$	$(0, 0)$	$(4, 0)$
α_2	M	$(8, 0)$	$(1, 1)$	$(8, 0)$
α_3	B	$(4, 0)$	$(0, 0)$	$(0, 1)$
Col's $E[\$]$		α_1	α_2	α_3

- Max. payoff for pure Bully is 1

Bully sometimes must be mixed

Consider matching pennies:

	H	T
H	$(1,-1)$	$(-1,1)$
T	$(-1,1)$	$(1,-1)$

The best way to profit from a pure follower is to randomise 50%. Your expected payoff then is 0.5.

A more complicated case. Suppose Row's strategy α is $(\alpha_1, \alpha_2, \alpha_3)$:

		L	C	R
α_1	T	$(0,1)$	$(0,0)$	$(4,0)$
α_2	M	$(8,0)$	$(1,1)$	$(8,0)$
α_3	B	$(4,0)$	$(0,0)$	$(0,1)$
Col's $E[\$]$		α_1	α_2	α_3

- Max. payoff for pure Bully is 1, when $\alpha = (0,1,0)$, i.e., when row plays M throughout.

Bully sometimes must be mixed

Consider matching pennies:

	H	T
H	$(1,-1)$	$(-1,1)$
T	$(-1,1)$	$(1,-1)$

The best way to profit from a pure follower is to randomise 50%. Your expected payoff then is 0.5.

A more complicated case. Suppose Row's strategy α is $(\alpha_1,\alpha_2,\alpha_3)$:

		L	C	R
α_1	T	$(0,1)$	$(0,0)$	$(4,0)$
α_2	M	$(8,0)$	$(1,1)$	$(8,0)$
α_3	B	$(4,0)$	$(0,0)$	$(0,1)$
Col's $E[\$]$		α_1	α_2	α_3

- Max. payoff for pure Bully is 1, when $\alpha = (0,1,0)$, i.e., when row plays M throughout.
- Row's $E[\$]$ for mixed $\alpha = (1/2, 0, 1/2)$ is 2, whatever Col's reply $(\beta_1, 0, 1-\beta_1)$.

Bully sometimes must be mixed

Consider matching pennies:

	H	T
H	$(1,-1)$	$(-1,1)$
T	$(-1,1)$	$(1,-1)$

The best way to profit from a pure follower is to randomise 50%. Your expected payoff then is 0.5.

A more complicated case. Suppose Row's strategy α is $(\alpha_1, \alpha_2, \alpha_3)$:

		L	C	R
α_1	T	$(0,1)$	$(0,0)$	$(4,0)$
α_2	M	$(8,0)$	$(1,1)$	$(8,0)$
α_3	B	$(4,0)$	$(0,0)$	$(0,1)$
Col's $E[\$]$		α_1	α_2	α_3

- Max. payoff for pure Bully is 1, when $\alpha = (0,1,0)$, i.e., when row plays M throughout.
- Row's $E[\$]$ for mixed $\alpha = (1/2, 0, 1/2)$ is 2, whatever Col's reply $(\beta_1, 0, 1-\beta_1)$.
- Row's $E[\$]$ for mixed $\alpha = (1/3, 1/3, 1/3)$ may be as low as $1/3$ and as high as 4, since for this α Col is indifferent among L, C and R.

Bully sometimes must be mixed

Consider matching pennies:

	H	T
H	$(1,-1)$	$(-1,1)$
T	$(-1,1)$	$(1,-1)$

The best way to profit from a pure follower is to randomise 50%. Your expected payoff then is 0.5.

A more complicated case. Suppose Row's strategy α is $(\alpha_1, \alpha_2, \alpha_3)$:

		L	C	R
α_1	T	$(0,1)$	$(0,0)$	$(4,0)$
α_2	M	$(8,0)$	$(1,1)$	$(8,0)$
α_3	B	$(4,0)$	$(0,0)$	$(0,1)$
Col's $E[\$]$		α_1	α_2	α_3

- Max. payoff for pure Bully is 1, when $\alpha = (0,1,0)$, i.e., when row plays M throughout.
- Row's $E[\$]$ for mixed $\alpha = (1/2, 0, 1/2)$ is 2, whatever Col's reply $(\beta_1, 0, 1-\beta_1)$.
- Row's $E[\$]$ for mixed $\alpha = (1/3, 1/3, 1/3)$ may be as low as $1/3$ and as high as 4, since for this α Col is indifferent among L, C and R.
- Row's $E[\$]$ for mixed $\alpha = (0.3334, 0.3332, 0.3334)$ is $2 \cdot 0.3334 \cdot 2 + 0.3332 \cdot 8$

Bully sometimes must be mixed

Consider matching pennies:

	H	T
H	$(1, -1)$	$(-1, 1)$
T	$(-1, 1)$	$(1, -1)$

The best way to profit from a pure follower is to randomise 50%. Your expected payoff then is 0.5.

A more complicated case. Suppose Row's strategy α is $(\alpha_1, \alpha_2, \alpha_3)$:

		L	C	R
α_1	T	$(0, 1)$	$(0, 0)$	$(4, 0)$
α_2	M	$(8, 0)$	$(1, 1)$	$(8, 0)$
α_3	B	$(4, 0)$	$(0, 0)$	$(0, 1)$
Col's $E[\$]$		α_1	α_2	α_3

- Max. payoff for pure Bully is 1, when $\alpha = (0, 1, 0)$, i.e., when row plays M throughout.
- Row's $E[\$]$ for mixed $\alpha = (1/2, 0, 1/2)$ is 2, whatever Col's reply $(\beta_1, 0, 1 - \beta_1)$.
- Row's $E[\$]$ for mixed $\alpha = (1/3, 1/3, 1/3)$ may be as low as $1/3$ and as high as 4, since for this α Col is indifferent among L, C and R.
- Row's $E[\$]$ for mixed $\alpha = (0.3334, 0.3332, 0.3334)$ is $2 \cdot 0.3334 \cdot 2 + 0.3332 \cdot 8$, which is almost 4.

Bully sometimes must be mixed

Consider matching pennies:

	H	T
H	$(1,-1)$	$(-1,1)$
T	$(-1,1)$	$(1,-1)$

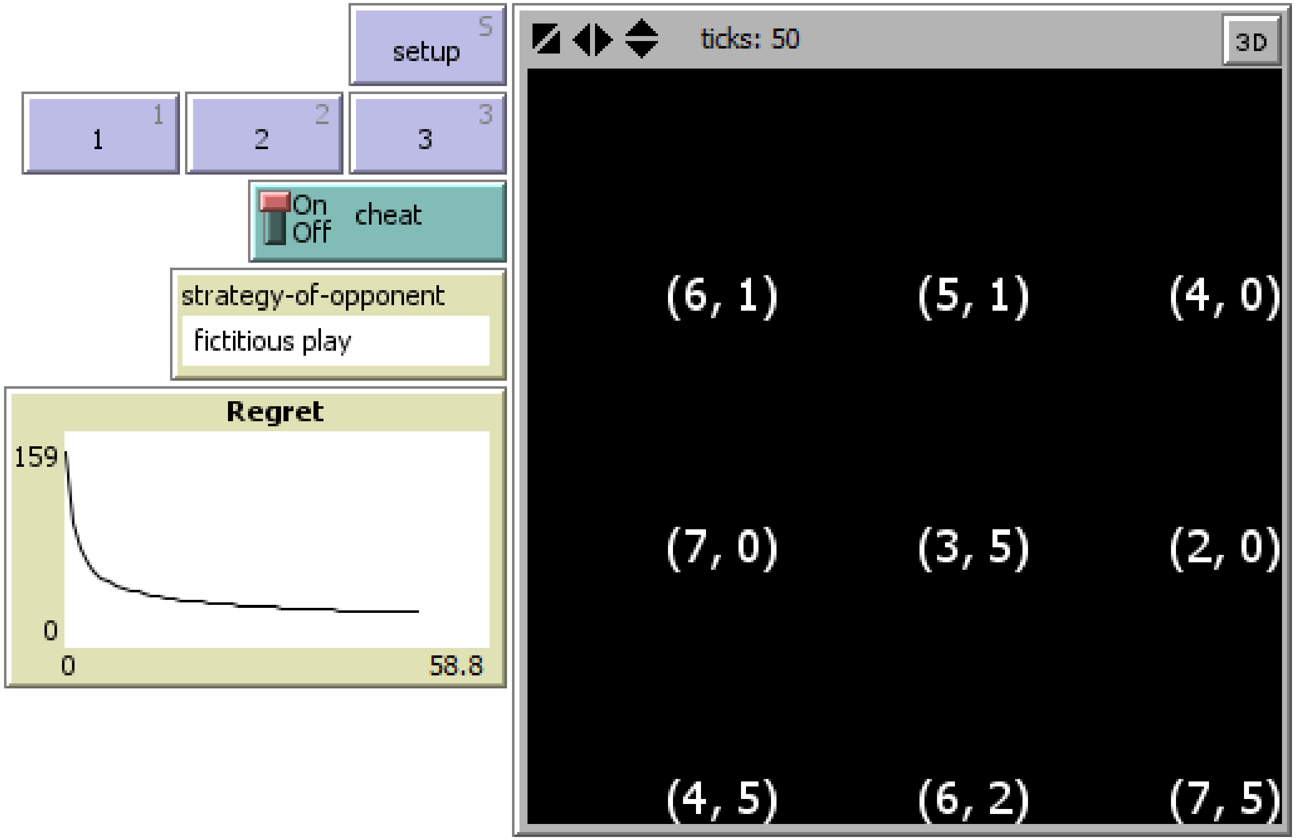
The best way to profit from a pure follower is to randomise 50%. Your expected payoff then is 0.5.

A more complicated case. Suppose Row's strategy α is $(\alpha_1, \alpha_2, \alpha_3)$:

		L	C	R
α_1	T	$(0,1)$	$(0,0)$	$(4,0)$
α_2	M	$(8,0)$	$(1,1)$	$(8,0)$
α_3	B	$(4,0)$	$(0,0)$	$(0,1)$
Col's $E[\$]$		α_1	α_2	α_3

- Max. payoff for pure Bully is 1, when $\alpha = (0,1,0)$, i.e., when row plays M throughout.
- Row's $E[\$]$ for mixed $\alpha = (1/2, 0, 1/2)$ is 2, whatever Col's reply $(\beta_1, 0, 1-\beta_1)$.
- Row's $E[\$]$ for mixed $\alpha = (1/3, 1/3, 1/3)$ may be as low as $1/3$ and as high as 4, since for this α Col is indifferent among L, C and R.
- Row's $E[\$]$ for mixed $\alpha = (0.3334, 0.3332, 0.3334)$ is $2 \cdot 0.3334 \cdot 2 + 0.3332 \cdot 8$, which is almost 4. Not bad!

Idea for an app to learn to play {against} Bully



Play against the computer. At the outset, the computer initializes to either Bully (with a probability of 50%) or pure fictitious play, the choice of which you can't see. After that, the computer won't change strategy. Try to press regret down as within few rounds as possible.

Bully: precise definition

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Bully: precise definition

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Surprisingly difficult to capture in an exact definition.

Bully: precise definition

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Surprisingly difficult to capture in an exact definition. The notion of best response helps us out:

Bully: precise definition

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Surprisingly difficult to capture in an exact definition. The notion of best response helps us out:

Bully_i

Bully: precise definition

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Surprisingly difficult to capture in an exact definition. The notion of best response helps us out:

$$\text{Bully}_i =_{Def}$$

Bully: precise definition

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Surprisingly difficult to capture in an exact definition. The notion of best response helps us out:

$$\text{Bully}_i =_{Def} \text{argmax}_{s_i \in S_i}$$

Bully: precise definition

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Surprisingly difficult to capture in an exact definition. The notion of best response helps us out:

$$\text{Bully}_i =_{Def} \text{argmax}_{s_i \in S_i} \min\{u_i(s_i, s_{-i})$$

Bully: precise definition

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Surprisingly difficult to capture in an exact definition. The notion of best response helps us out:

$$\mathrm{Bully}_i =_{Def} \mathrm{argmax}_{s_i \in S_i} \min\{u_i(s_i, s_{-i}) \mid s_{-i} \in \mathrm{BR}(s_i)\}.$$

Bully: precise definition

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Surprisingly difficult to capture in an exact definition. The notion of best response helps us out:

$$\text{Bully}_i =_{Def} \text{argmax}_{s_i \in S_i} \min\{u_i(s_i, s_{-i}) \mid s_{-i} \in \text{BR}(s_i)\}.$$

- Right most inner part (green): best response of opponent to s_i.

Bully: precise definition

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Surprisingly difficult to capture in an exact definition. The notion of best response helps us out:

$$\text{Bully}_i =_{Def} \text{argmax}_{s_i \in S_i} \min\{u_i(s_i, s_{-i}) \mid s_{-i} \in \text{BR}(s_i)\}.$$

- Right most inner part (green): best response of opponent to s_i.
- Middle inner part (as from min): guaranteed payoff for bullying opponent with s_i.

Bully: precise definition

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Surprisingly difficult to capture in an exact definition. The notion of best response helps us out:

$$\text{Bully}_i =_{Def} \text{argmax}_{s_i \in S_i} \min\{u_i(s_i, s_{-i}) \mid s_{-i} \in \text{BR}(s_i)\}.$$

- Right most inner part (green): best response of opponent to s_i.
- Middle inner part (as from min): guaranteed payoff for bullying opponent with s_i.
- Entire formula: choose s_i that maximises this guaranteed payoff.

Bully: precise definition

Play any strategy that gives you the highest payoff, assuming that your opponent is a mindless follower.

Surprisingly difficult to capture in an exact definition. The notion of best response helps us out:

$$\text{Bully}_i =_{Def} \text{argmax}_{s_i \in S_i} \min\{u_i(s_i, s_{-i}) \mid s_{-i} \in \text{BR}(s_i)\}.$$

- Right most inner part (green): best response of opponent to s_i.
- Middle inner part (as from min): guaranteed payoff for bullying opponent with s_i.
- Entire formula: choose s_i that maximises this guaranteed payoff.

  Recognise the maxmin = the security value in this formula!

Bully: precise definition (in parts)

Bully: precise definition (in parts)

- Let $\mathrm{BR}(s_i)$ be the set of all best responses to i's strategy s_i:

$$\mathrm{BR}(s_i) =_{Def} \operatorname{argmax}_{s_{-i}} \{u_{-i}(s_i, s_{-i}) \mid s_{-i} \in S_{-i}\}.$$

Bully: precise definition (in parts)

- Let $\mathrm{BR}(s_i)$ be the set of all best responses to i's strategy s_i:

$$\mathrm{BR}(s_i) =_{Def} \operatorname{argmax}_{s_{-i}} \{u_{-i}(s_i, s_{-i}) \mid s_{-i} \in S_{-i}\}.$$

- Let $\mathrm{Bully}_i(s_i)$ be the payoff guaranteed for playing s_i against mindless followers (i.e, best responders):

$$\mathrm{Bully}_i(s_i) =_{Def} \min\{u_i(s_i, s_{-i}) \mid s_{-i} \in \mathrm{BR}(s_i)\}$$

Bully: precise definition (in parts)

- Let $\mathrm{BR}(s_i)$ be the set of all best responses to i's strategy s_i:

$$\mathrm{BR}(s_i) =_{Def} \mathrm{argmax}_{s_{-i}}\{u_{-i}(s_i, s_{-i}) \mid s_{-i} \in S_{-i}\}.$$

- Let $\mathrm{Bully}_i(s_i)$ be the payoff guaranteed for playing s_i against mindless followers (i.e, best responders):

$$\mathrm{Bully}_i(s_i) =_{Def} \min\{u_i(s_i, s_{-i}) \mid s_{-i} \in \mathrm{BR}(s_i)\}$$

- The set of bully strategies is formed by:

$$\mathrm{Bully}_i =_{Def} \mathrm{argmax}_{s_i \in S_i} \mathrm{Bully}_i(s_i)$$

Bully: precise definition (in parts)

- Let $\mathrm{BR}(s_i)$ be the set of all best responses to i's strategy s_i:

$$\mathrm{BR}(s_i) =_{Def} \mathrm{argmax}_{s_{-i}} \{u_{-i}(s_i, s_{-i}) \mid s_{-i} \in S_{-i}\}.$$

- Let $\mathrm{Bully}_i(s_i)$ be the payoff guaranteed for playing s_i against mindless followers (i.e, best responders):

$$\mathrm{Bully}_i(s_i) =_{Def} \min\{u_i(s_i, s_{-i}) \mid s_{-i} \in \mathrm{BR}(s_i)\}$$

- The set of bully strategies is formed by:

$$\mathrm{Bully}_i =_{Def} \mathrm{argmax}_{s_i \in S_i} \mathrm{Bully}_i(s_i)$$

- Bully is stateless (a.k.a. memoryless, i.e, memory of $k = 0$ rounds), hence keeps playing the same action throughout.

Godfather

Godfather (Littman and Stone, 2001)

Godfather (Littman and Stone, 2001)

- A strategy [function $H \to \Delta(A)$ from histories to mixed strategies] that makes its opponent an offer that it cannot refuse.

Godfather (Littman and Stone, 2001)

- A strategy [function $H \to \Delta(A)$ from histories to mixed strategies] that makes its opponent an offer that it cannot refuse.
- Capitalises on the Folk theorem for repeated games with (not necessarily SGP) Nash equilibria.

Godfather (Littman and Stone, 2001)

- A strategy [function $H \to \Delta(A)$ from histories to mixed strategies] that makes its opponent an offer that it cannot refuse.
- Capitalises on the Folk theorem for repeated games with (not necessarily SGP) Nash equilibria.
- A pair of strategies (s_i, s_{-i}) is called a targetable pair if playing them results in each player getting more than the safety value (maxmin) and plays its half of the pair.

Godfather (Littman and Stone, 2001)

- A strategy [function $H \to \Delta(A)$ from histories to mixed strategies] that makes its opponent an offer that it cannot refuse.
- Capitalises on the Folk theorem for repeated games with (not necessarily SGP) Nash equilibria.
- A pair of strategies (s_i, s_{-i}) is called a targetable pair if playing them results in each player getting more than the safety value (maxmin) and plays its half of the pair.
- Godfather chooses a targetable pair.

Godfather (Littman and Stone, 2001)

- A strategy [function $H \rightarrow \Delta(A)$ from histories to mixed strategies] that makes its opponent an offer that it cannot refuse.

- Capitalises on the Folk theorem for repeated games with (not necessarily SGP) Nash equilibria.

- A pair of strategies (s_i, s_{-i}) is called a targetable pair if playing them results in each player getting more than the safety value (maxmin) and plays its half of the pair.

- Godfather chooses a targetable pair.

  1. If the opponent keeps playing its half of targetable pair in one stage, Godfather plays its half in the next stage.

Godfather (Littman and Stone, 2001)

- A strategy [function $H \to \Delta(A)$ from histories to mixed strategies] that makes its opponent an offer that it cannot refuse.
- Capitalises on the Folk theorem for repeated games with (not necessarily SGP) Nash equilibria.
- A pair of strategies (s_i, s_{-i}) is called a targetable pair if playing them results in each player getting more than the safety value (maxmin) and plays its half of the pair.
- Godfather chooses a targetable pair.

  1. If the opponent keeps playing its half of targetable pair in one stage, Godfather plays its half in the next stage.
  2. Otherwise it falls back forever to the (mixed) strategy that forces the opponent to achieve at most its safety value.

Godfather (Littman and Stone, 2001)

- A strategy [function $H \to \Delta(A)$ from histories to mixed strategies] that makes its opponent an offer that it cannot refuse.
- Capitalises on the Folk theorem for repeated games with (not necessarily SGP) Nash equilibria.
- A pair of strategies (s_i, s_{-i}) is called a targetable pair if playing them results in each player getting more than the safety value (maxmin) and plays its half of the pair.
- Godfather chooses a targetable pair.
  1. If the opponent keeps playing its half of targetable pair in one stage, Godfather plays its half in the next stage.
  2. Otherwise it falls back forever to the (mixed) strategy that forces the opponent to achieve at most its safety value.
- Godfather needs a memory of $k = 1$ (one round).

Folk theorem for NE in repeated games

Folk theorem for NE in repeated games

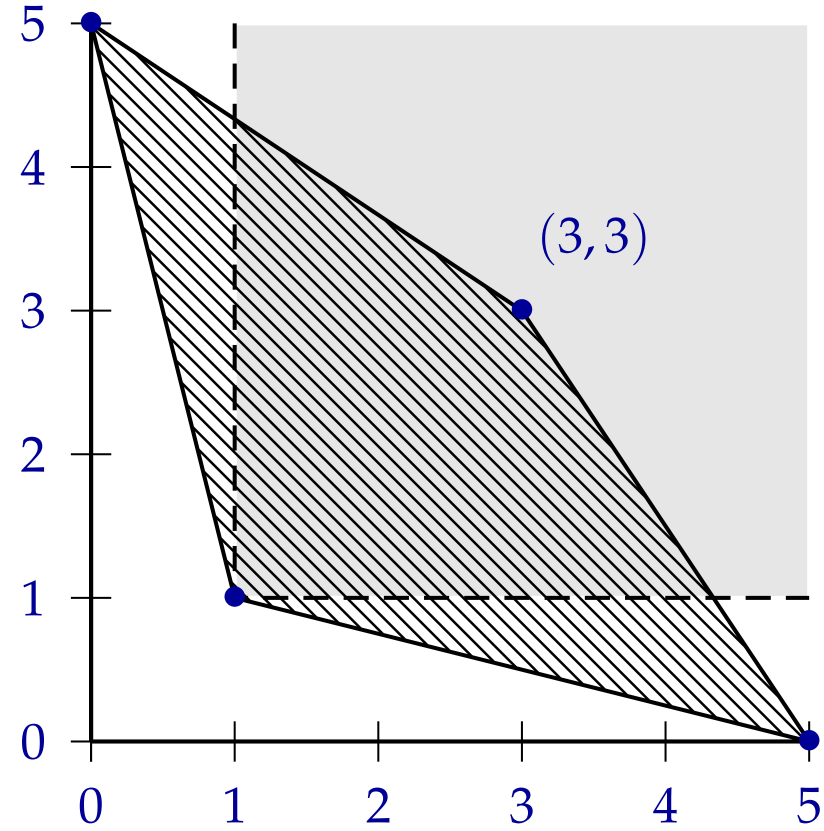
Folk theorem for NE in repeated games

- **Feasible payoffs** (striped): payoff combos that can be obtained by jointly repeating patterns of actions (more accurate: patterns of action profiles).

Folk theorem for NE in repeated games

- **Feasible payoffs** (striped): payoff combos that can be obtained by jointly repeating patterns of actions (more accurate: patterns of action profiles).
- **Enforceable payoffs** (shaded): no one goes below their minmax.

Theorem. If (x, y) is both feasible and enforceable, then (x, y) is the payoff in a Nash equilibrium of the infinitely repeated G with average payoffs.
Conversely, if (x, y) is the payoff in any Nash equilibrium of the infinitely repeated G with average payoffs, then (x, y) is enforceable.



Variations on Godfather with memory $k > 1$

(Taken from Chakraborty and Stone, 2008):

Variations on Godfather with memory $k > 1$

(Taken from Chakraborty and Stone, 2008):

- Godfather-lenient plays its part of a targetable pair if, within the last k actions, the opponent played its own half of the pair at least once.

  Otherwise execute threat. (But no longer forever.)

Variations on Godfather with memory $k > 1$

(Taken from Chakraborty and Stone, 2008):

- Godfather-lenient plays its part of a targetable pair if, within the last k actions, the opponent played its own half of the pair at least once.

  Otherwise execute threat. (But no longer forever.)

- Godfather-strict plays its part of a targetable pair if, within the last k actions, the opponent always played its own half of the pair.

Variations on Godfather with memory $k > 1$

(Taken from Chakraborty and Stone, 2008):

- Godfather-lenient plays its part of a targetable pair if, within the last k actions, the opponent played its own half of the pair at least once.

  Otherwise execute threat. (But no longer forever.)

- Godfather-strict plays its part of a targetable pair if, within the last k actions, the opponent always played its own half of the pair.

Godfather++ (Littman & Stone, 2005)

Godfather++ (Littman & Stone, 2005)

- The name “Godfather++” is due to Crandall (2005).

Godfather++ (Littman & Stone, 2005)

- The name "Godfather++" is due to Crandall (2005).
- Capitalises on the Folk theorem for repeated games with (not necessarily SGP) Nash equilibria.

Godfather++ (Littman & Stone, 2005)

- The name "Godfather++" is due to Crandall (2005).
- Capitalises on the Folk theorem for repeated games with (not necessarily SGP) Nash equilibria.
- Godfather++ a polynomial-time algorithm for constructing a finite state machine.

  This FSM represents a strategy which plays a Nash equilibrium for a repeated 2-player game with averaged payoffs.

Godfather++ (Littman & Stone, 2005)

- The name "Godfather++" is due to Crandall (2005).
- Capitalises on the Folk theorem for repeated games with (not necessarily SGP) Nash equilibria.
- Godfather++ a polynomial-time algorithm for constructing a finite state machine.

  This FSM represents a strategy which plays a Nash equilibrium for a repeated 2-player game with averaged payoffs.
-

Godfather++ (Littman & Stone, 2005)

- The name "Godfather++" is due to Crandall (2005).
- Capitalises on the Folk theorem for repeated games with (not necessarily SGP) Nash equilibria.
- Godfather++ a polynomial-time algorithm for constructing a finite state machine.

  This FSM represents a strategy which plays a Nash equilibrium for a repeated 2-player game with averaged payoffs.
- - Not for finite repeated games.

Godfather++ (Littman & Stone, 2005)

- The name "Godfather++" is due to Crandall (2005).
- Capitalises on the Folk theorem for repeated games with (not necessarily SGP) Nash equilibria.
- Godfather++ a polynomial-time algorithm for constructing a finite state machine.

  This FSM represents a strategy which plays a Nash equilibrium for a repeated 2-player game with averaged payoffs.
- • Not for finite repeated games.
  • Not for infinite repeated games with discounted payoffs.

Godfather++ (Littman & Stone, 2005)

- The name "Godfather++" is due to Crandall (2005).
- Capitalises on the Folk theorem for repeated games with (not necessarily SGP) Nash equilibria.
- Godfather++ a polynomial-time algorithm for constructing a finite state machine.

  This FSM represents a strategy which plays a Nash equilibrium for a repeated 2-player game with averaged payoffs.
- - Not for finite repeated games.
  - Not for infinite repeated games with discounted payoffs.
  - Not for n-player games, $n > 2$.

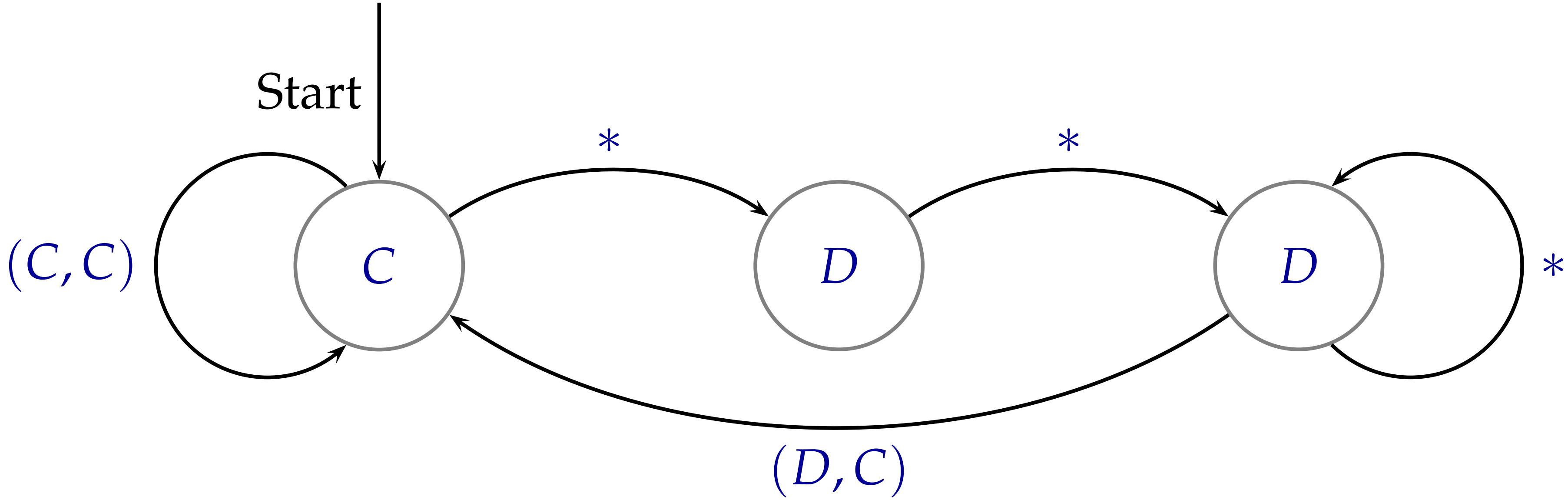
Godfather++ (Littman & Stone, 2005)

- The name "Godfather++" is due to Crandall (2005).
- Capitalises on the Folk theorem for repeated games with (not necessarily SGP) Nash equilibria.
- Godfather++ a polynomial-time algorithm for constructing a finite state machine.

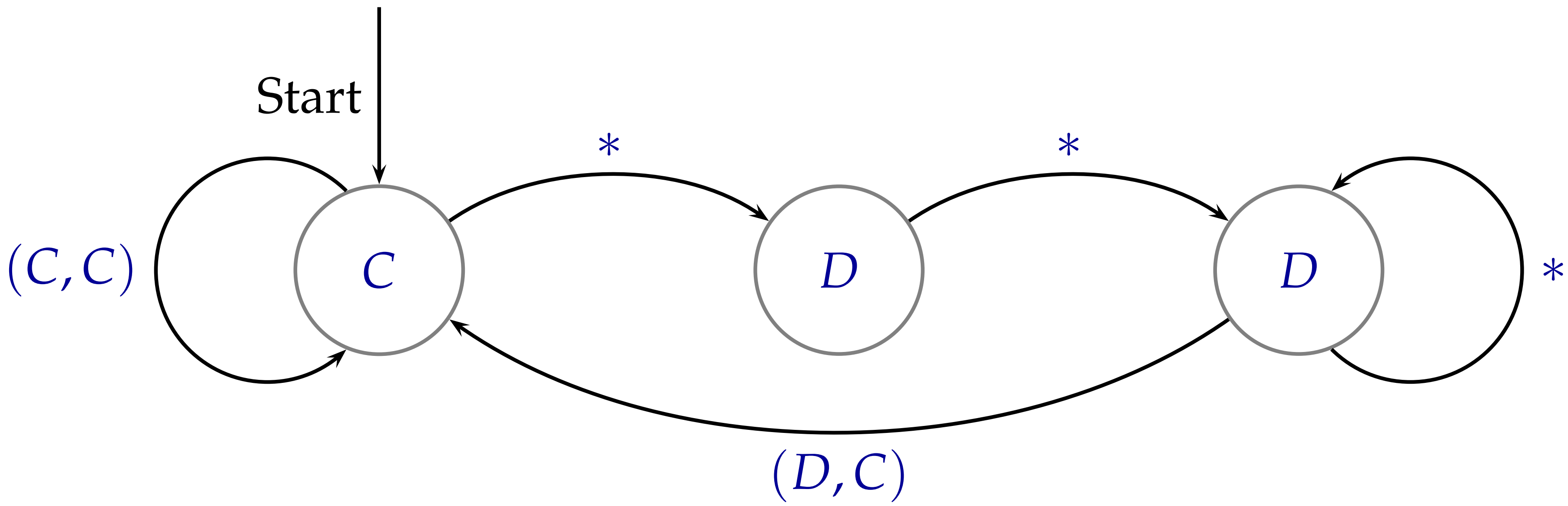
  This FSM represents a strategy which plays a Nash equilibrium for a repeated 2-player game with averaged payoffs.
-
  - Not for finite repeated games.
  - Not for infinite repeated games with discounted payoffs.
  - Not for n-player games, $n > 2$.

Michael L. Littman and Peter Stone (2005). "A polynomial-time Nash equilibrium algorithm for repeated games". In *Decision Support Systems* Vol. 39, pp. 55-66.

Finite machine for "two tits for tat"

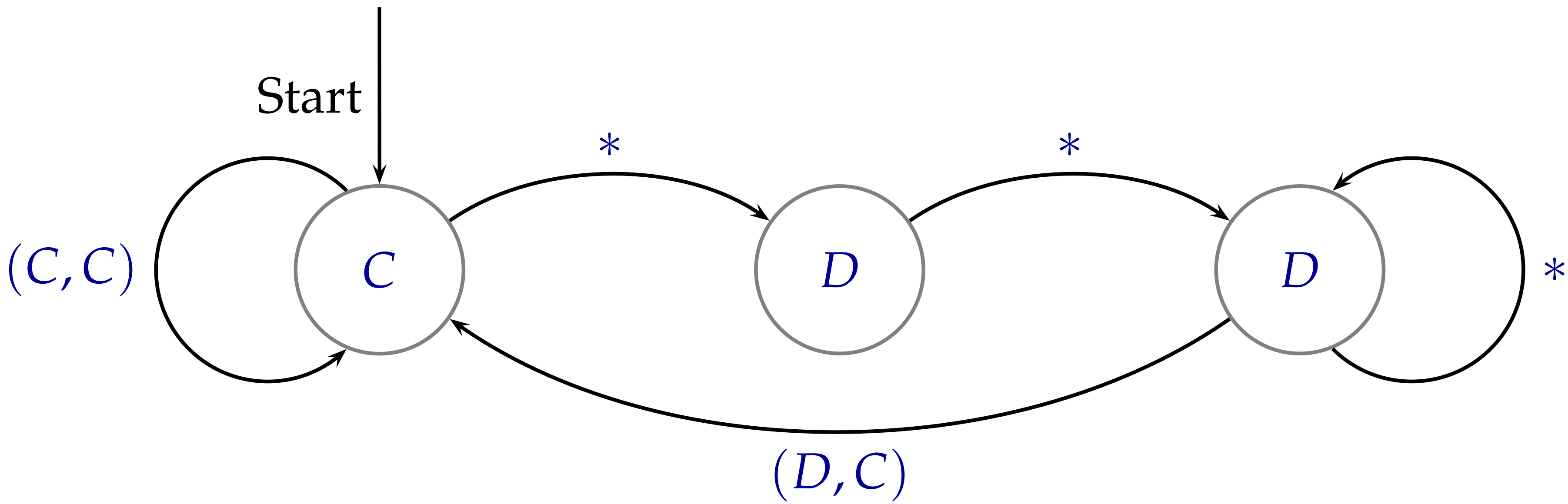


Finite machine for "two tits for tat"



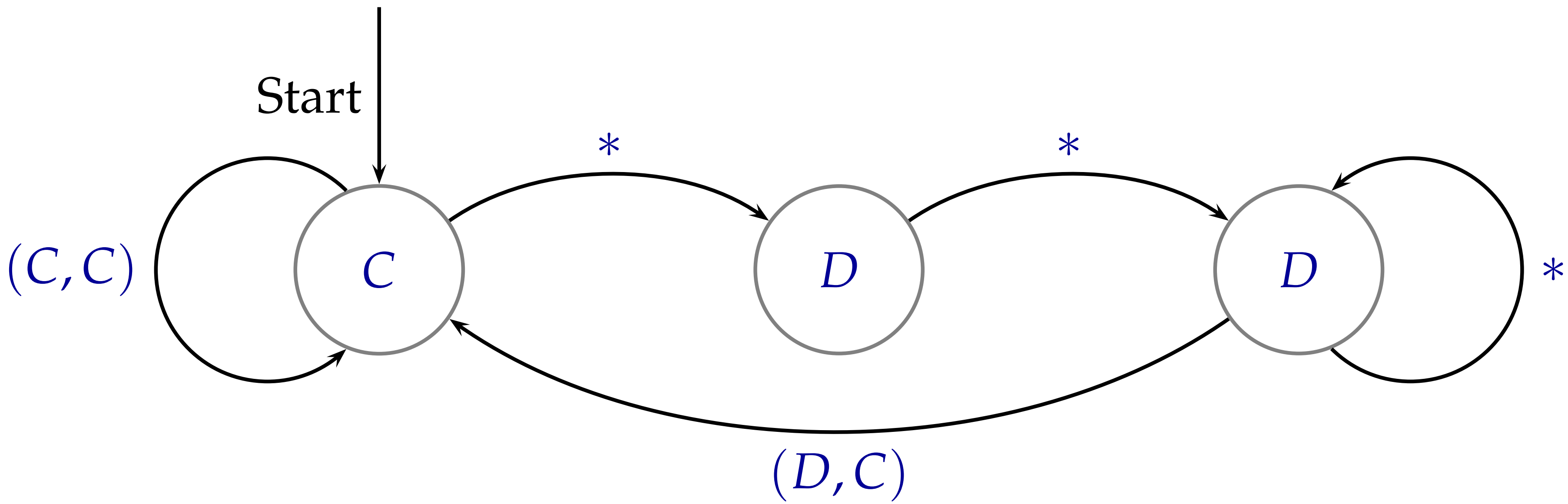
- Finite state machine for the **Prisoners' dilemma**.

Finite machine for "two tits for tat"



- Finite state machine for the **Prisoners' dilemma**.
- **Personal actions** determine states.

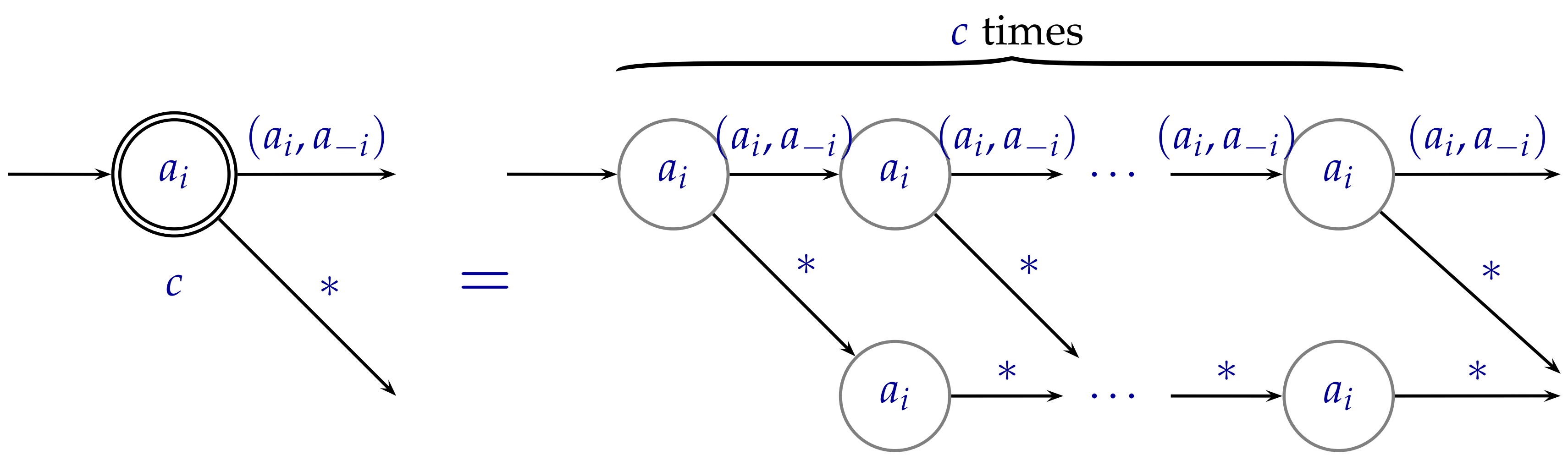
Finite machine for "two tits for tat"



- Finite state machine for the **Prisoners' dilemma**.
- **Personal actions** determine states.
- **Action profiles** determine transitions between states.

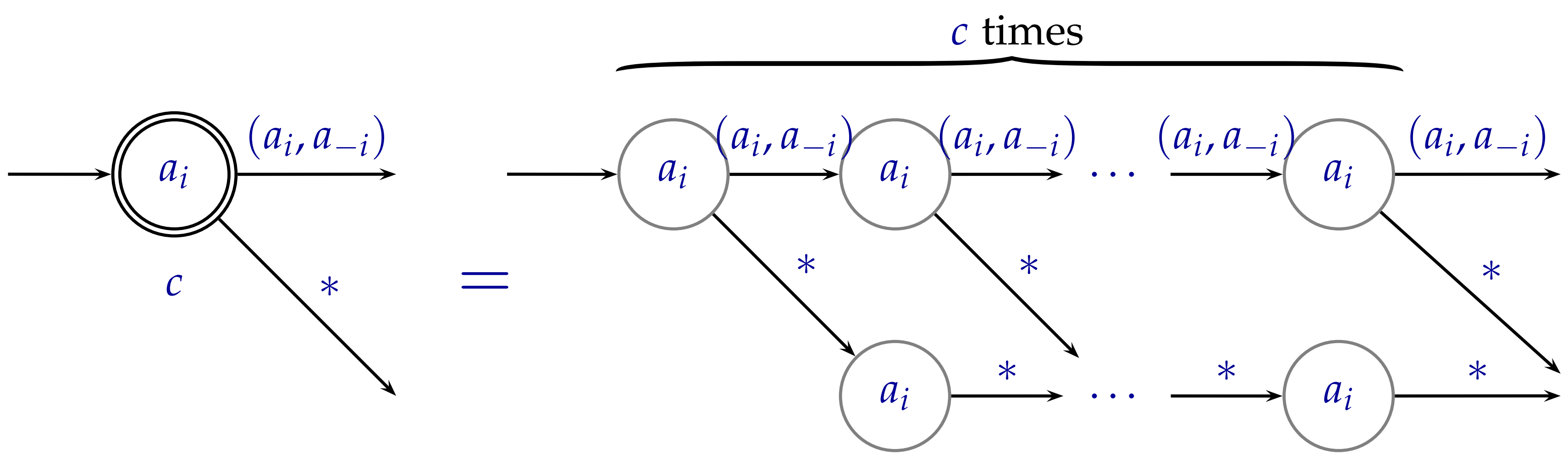
  The "$*$" represents an "else," in the sense of "all other action profiles".

The use of counting nodes



Upon entry:

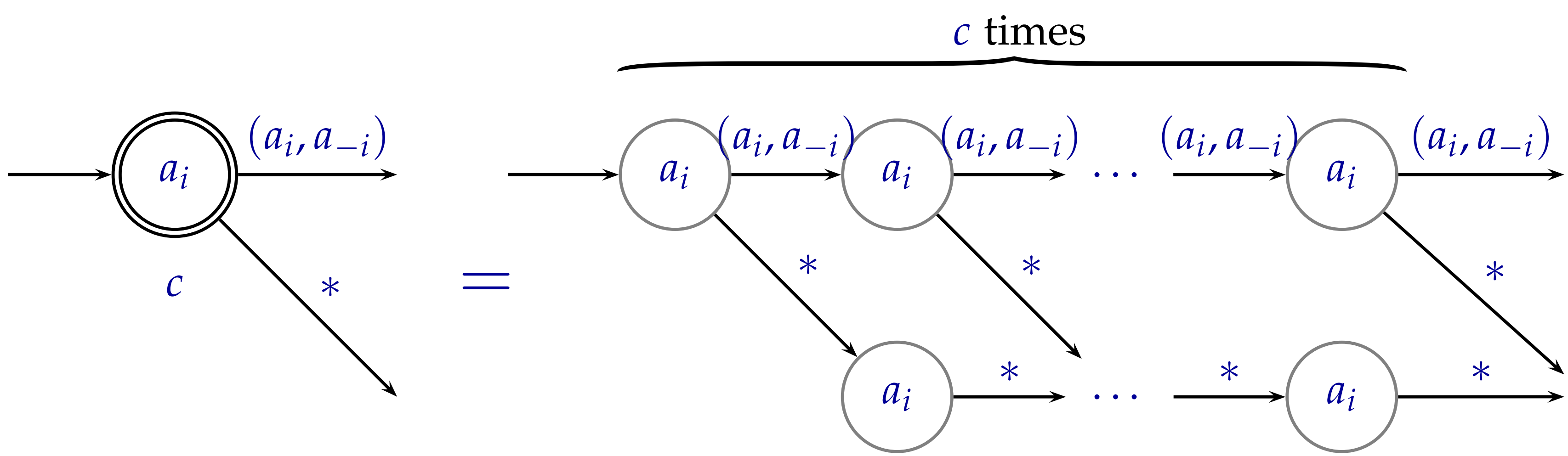
The use of counting nodes



Upon entry:

- If exactly c times action profile (a_i, a_{-i}) is played, then take exit above.

The use of counting nodes

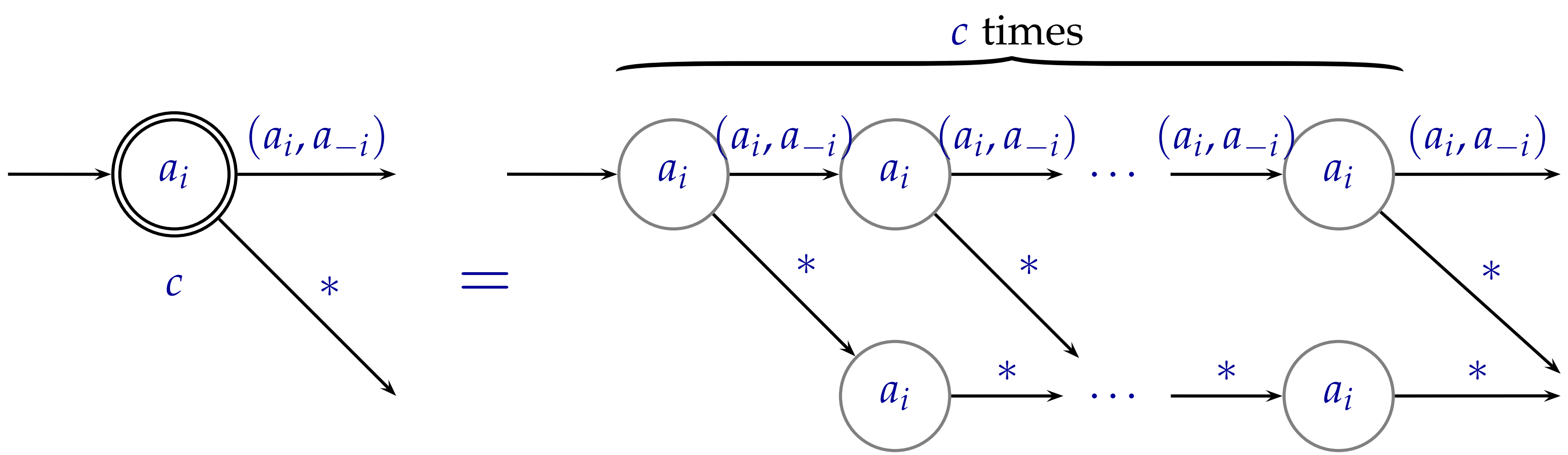


Upon entry:

- If exactly c times action profile (a_i, a_{-i}) is played, then take exit above.

- If column player deviates in round d, keep playing a_i for the remaining $c - (d + 1)$ rounds. Finally, exit below.

The use of counting nodes



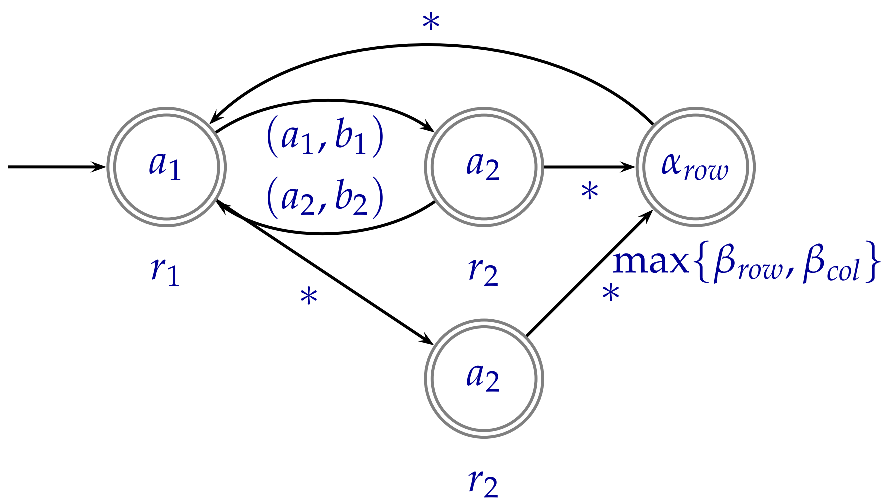
Upon entry:

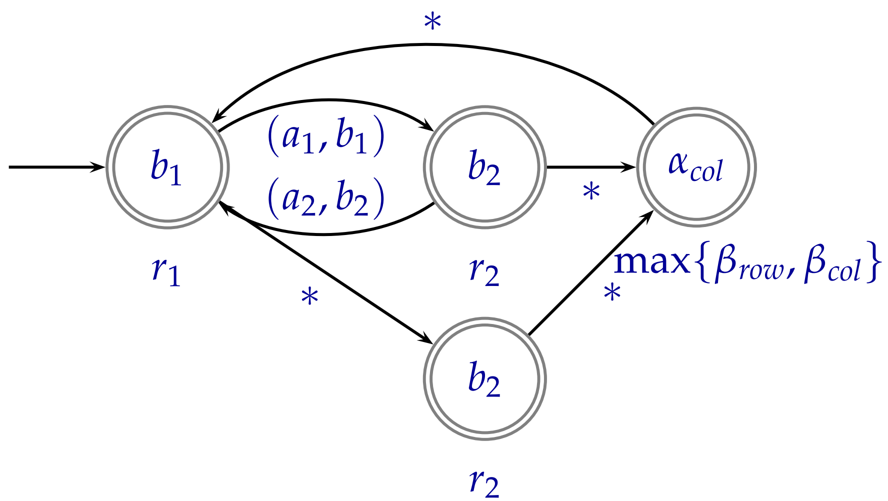
- If exactly c times action profile (a_i, a_{-i}) is played, then take exit above.

- If column player deviates in round d, keep playing a_i for the remaining $c - (d+1)$ rounds. Finally, exit below.

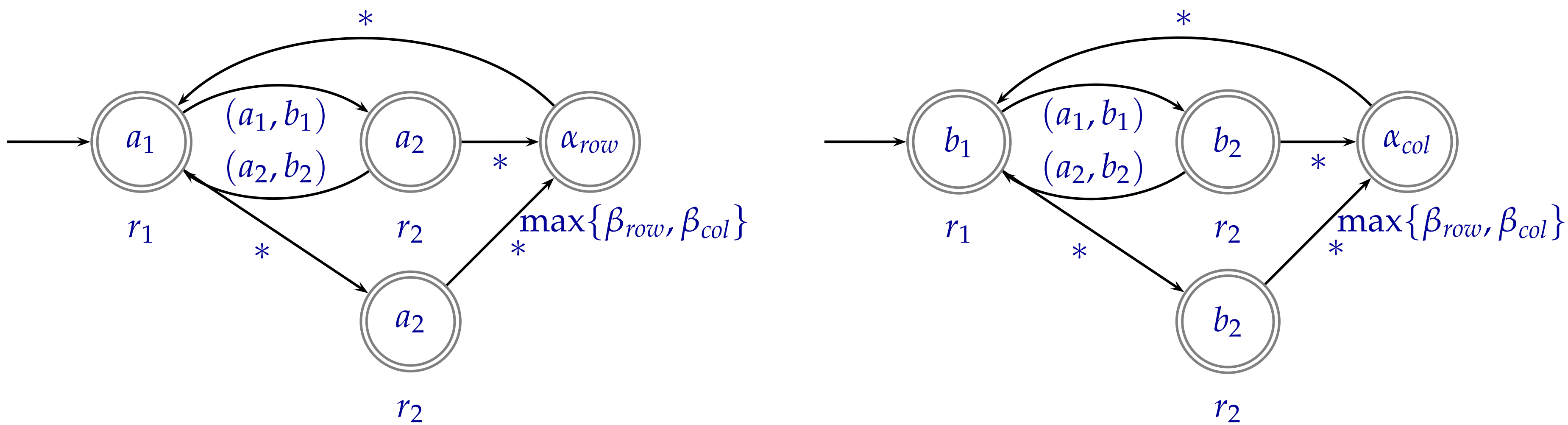
- Because integers up to c can be expressed in $\log c$ bits (roughly), size of finite machine is polynomial in $\log c$.

Pair of strategies that is a Nash equilibrium in a repeated game



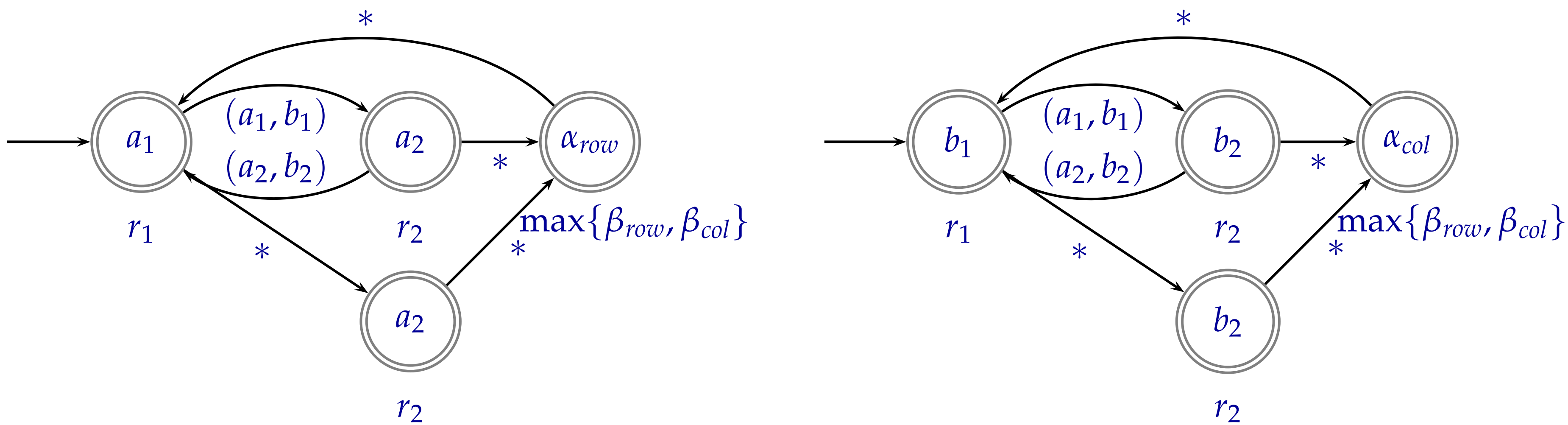


Pair of strategies that is a Nash equilibrium in a repeated game



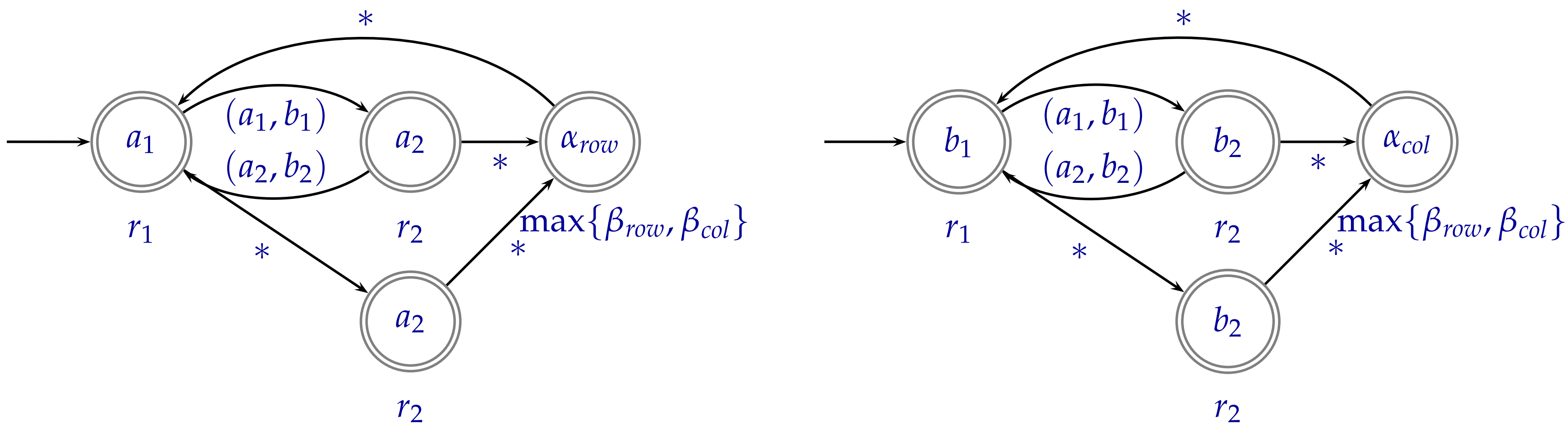
- Node a_1 and a_2 are the actions that row must play (in sync with col). First $r_1 \times a_1$, then $r_2 \times a_2$, then $r_1 \times a_1$, etc.

Pair of strategies that is a Nash equilibrium in a repeated game



- Node a_1 and a_2 are the actions that row must play (in sync with col). First $r_1 \times a_1$, then $r_2 \times a_2$, then $r_1 \times a_1$, etc.
- If opponent deviates, then retaliate with α_{row} for $\max\{\beta_{row}, \beta_{col}\}$ rounds.

Pair of strategies that is a Nash equilibrium in a repeated game



- Node a_1 and a_2 are the actions that row must play (in sync with col). First $r_1 \times a_1$, then $r_2 \times a_2$, then $r_1 \times a_1$, etc.
- If opponent deviates, then retaliate with α_{row} for $\max\{\beta_{row}, \beta_{col}\}$ rounds.
- The two automata *always* run in sync, no matter who deviates first. It can (easily) be deduced that, for each player, deviating at any node is detrimental $\Rightarrow$ Nash equilibrium in repeated game.

The devil and the details. . .

It should be that all parameters can be determined analytically, in polynomial time.

The devil and the details. . .

It should be that all parameters can be determined analytically, in polynomial time.

The devil and the details...

It should be that all parameters can be determined analytically, in polynomial time.

1. **The coordinated action profiles (a_1, b_1), (a_2, b_2) and their duration of play r_1, r_2.**

   Nash says: take strategy pair (s_1, s_2) that maximises the product of players' advantages. This pair can be obtained (or at least approximated) by playing convex

   $$\frac{r_1}{r_1 + r_2}(a_1, b_1) + \frac{r_2}{r_1 + r_2}(a_2, b_2)$$

   for r_1, r_2 not too large.

   Pair (s_1, s_2) is obtained by looping through $(A^2)^2$ (all pairs of pairs of actions).

The devil and the details...

It should be that all parameters can be determined analytically, in polynomial time.

1. **The coordinated action profiles (a_1, b_1), (a_2, b_2) and their duration of play r_1, r_2.**

   Nash says: take strategy pair (s_1, s_2) that maximises the product of players' advantages. This pair can be obtained (or at least approximated) by playing convex

   $$\frac{r_1}{r_1 + r_2}(a_1, b_1) + \frac{r_2}{r_1 + r_2}(a_2, b_2)$$

   for r_1, r_2 not too large.

   Pair (s_1, s_2) is obtained by looping through $(A^2)^2$ (all pairs of pairs of actions).

2. **The strategy and duration of punishment (α_{row}, α_{col} and β_{row}, β_{col}, respectively).**

The devil and the details…

It should be that all parameters can be determined analytically, in polynomial time.

1. **The coordinated action profiles (a_1, b_1), (a_2, b_2) and their duration of play r_1, r_2.**

   Nash says: take strategy pair (s_1, s_2) that maximises the product of players' advantages. This pair can be obtained (or at least approximated) by playing convex

   $$\frac{r_1}{r_1 + r_2}(a_1, b_1) + \frac{r_2}{r_1 + r_2}(a_2, b_2)$$

   for r_1, r_2 not too large.

   Pair (s_1, s_2) is obtained by looping through $(A^2)^2$ (all pairs of pairs of actions).

2. **The strategy and duration of punishment (α_{row}, α_{col} and β_{row}, β_{col}, respectively).**

The devil and the details...

It should be that all parameters can be determined analytically, in polynomial time.

1. **The coordinated action profiles (a_1, b_1), (a_2, b_2) and their duration of play r_1, r_2.**

   Nash says: take strategy pair (s_1, s_2) that maximises the product of players' advantages. This pair can be obtained (or at least approximated) by playing convex

   $$\frac{r_1}{r_1 + r_2}(a_1, b_1) + \frac{r_2}{r_1 + r_2}(a_2, b_2)$$

   for r_1, r_2 not too large.

   Pair (s_1, s_2) is obtained by looping through $(A^2)^2$ (all pairs of pairs of actions).

2. **The strategy and duration of punishment (α_{row}, α_{col} and β_{row}, β_{col}, respectively).**

   - α_{row} and α_{col} are the minmax strategies of the stage game.

The devil and the details…

It should be that all parameters can be determined analytically, in polynomial time.

1. **The coordinated action profiles (a_1, b_1), (a_2, b_2) and their duration of play r_1, r_2.**

   Nash says: take strategy pair (s_1, s_2) that maximises the product of players' advantages. This pair can be obtained (or at least approximated) by playing convex

   $$\frac{r_1}{r_1 + r_2}(a_1, b_1) + \frac{r_2}{r_1 + r_2}(a_2, b_2)$$

   for r_1, r_2 not too large.

   Pair (s_1, s_2) is obtained by looping through $(A^2)^2$ (all pairs of pairs of actions).

2. **The strategy and duration of punishment (α_{row}, α_{col} and β_{row}, β_{col}, respectively).**

- α_{row} and α_{col} are the minmax strategies of the stage game.
- β_{row} and β_{col} depend on turning points to "get even". These are determined by (i) the average payoff for cooperating (ii) upper bound on largest possible value for a single round of freeriding.

Part II:
Crandall & Goodrich (2005)