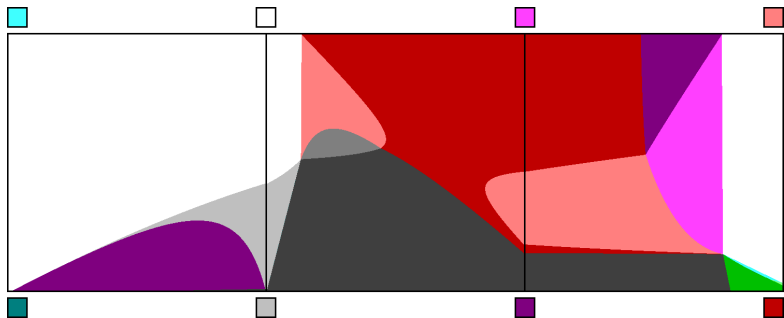


Solving the prisoner in memory-one strategies

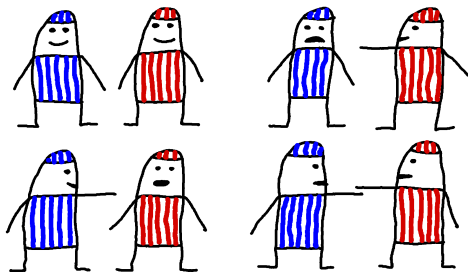
David Kruml¹

SSAOS 2016, Trojanovice



¹Supported by the project GAČR I 1923-N25.

One-round game



u / v	cooperate (C)	defect (D)
cooperate (C)	3 / 3	0 / 5
defect (D)	5 / 0	1 / 1

D dominates C : $5 > 3, 1 > 0$.

Iterated prisoner's dilemma

Players can react differently to every sequence of previous rounds (*history*) — much more strategies.

Popular simplifications — limited memory (automata, Markov strategies). Let us consider *memory-one strategies*. They react to the previous round situation. The expected pay-off is a *mean pay-off* for an infinitely iterated game.

Making the opponent cooperate is more profitable than just “stealing” some extra points by defection.

Strategies appreciating long run mutual cooperation are more successful. Concepts of revenge, forgiveness, temptation, etc. Applications in biology (evolution of altruism), economy (oligopoly), social and political studies, ethics (golden rule), etc.

Examples of memory-one strategies



AIC always cooperate



AID always defect



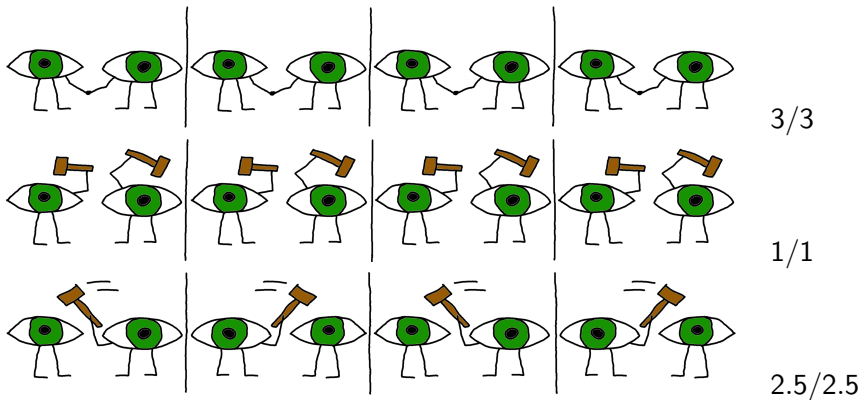
tit-for-tat imitate opponent's last move



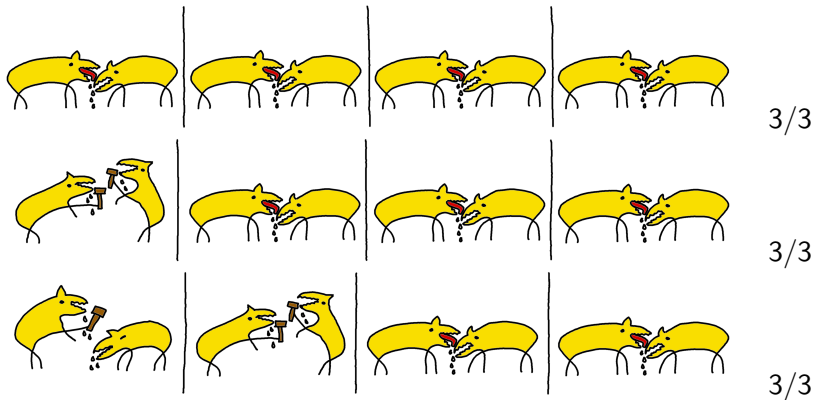
Pavlov win (CC, DC) — stay,
loose (CD, DD) — switch

⋮

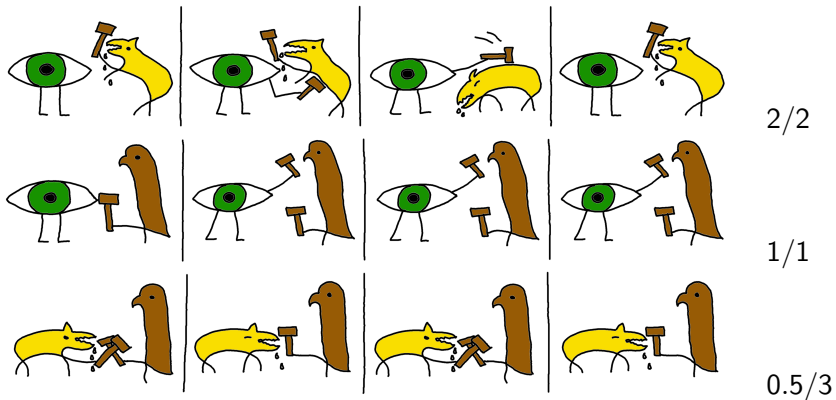
TFT vs. TFT









Pavlov vs. Pavlov



Other combats





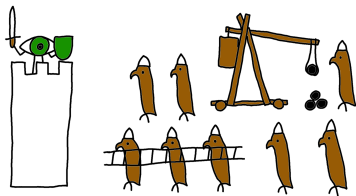
Tournaments

		AIC	AIID	TFT	pTFT	Pa	pPa	Σ
AIC		3	0	3	3	3	0	12
AIID		5	1	1	1	3	3	14
(friendly) TFT		3	1	3	2.5	3	2	14.5
probing TFT		3	1	2.5	1	2	2	11.5
(friendly) Pavlov		3	0.5	3	2	3	3	14.5
probing Pavlov		5	0.5	2	2	3	3	15.5

Evolutionary stability

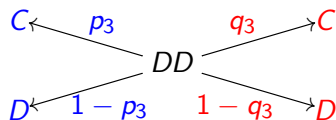
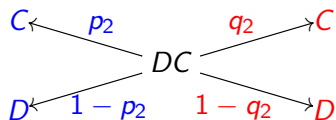
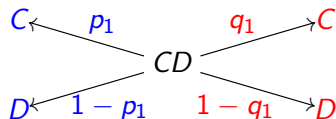
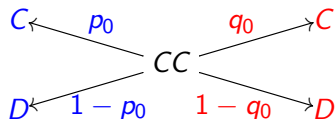
Population with x All D and y TFT. Simulated evolution — more successful strategy is awarded by a larger offspring.

		AllD	TFT	Σ
AllD		1	1	$x + y$
TFT		1	3	$x + 3y$



TFT invades AllD and TFT resists to AllD, for any ratio.

Probability memory-one strategies



Noise — probabilities restricted to $[e, 1 - e]$ for some small fixed $e > 0$. The induced Markov chain is ergodic, tends to a unique stationary vector, and the first round actions are irrelevant. The strategies are quadruples $p = [p_0, p_1, p_2, p_3]$, $q = [q_0, q_1, q_2, q_3]$.

Adjust your avatar!



p_0	$(CC \rightarrow C)$	niceness	$1 - p_0$	$(CC \rightarrow D)$	nastiness
p_1	$(CD \rightarrow C)$	gratuity	$1 - p_1$	$(CD \rightarrow D)$	retaliation
p_2	$(DC \rightarrow C)$	forgiveness	$1 - p_2$	$(DC \rightarrow D)$	impenitence
p_3	$(DD \rightarrow C)$	conciliatoryness	$1 - p_3$	$(DD \rightarrow D)$	cautiousness

Axelrod's recommendation: be nice + retaliate + forgive.

Notation for strategies



[1, 0, 1, 0]

TFT

ioio



[1, 1, 1, 1]

AIIC

iiii



[0, 0, 0, 0]

AIID

oooo



[1, 0, 0, 1]

Pavlov

iooi

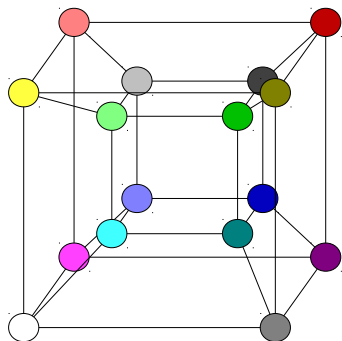


[1, 1/3, 1, 2/3] GTFT (gratuitous TFT)

ixiy

Noised versions: $[1 - e, e, 1 - e, e]$ is noised TFT, written as *ioio*.

Strategy space



White	<i>oooo</i>
Yellow	<i>oooi</i>
Pink	<i>ooio</i>
Light Red	<i>ooii</i>
Cyan	<i>oioo</i>
Light Green	<i>oioi</i>
Light Blue	<i>oioo</i>
Light Grey	<i>oiii</i>

Dark Grey	<i>iooo</i>
Olive	<i>iooi</i>
Purple	<i>ioio</i>
Red	<i>ioii</i>
Teal	<i>iioo</i>
Green	<i>iioi</i>
Blue	<i>iiio</i>
Dark Grey	<i>iiii</i>

16 corners (0-faces), 32 edges (1-faces), 24 squares (2-faces), 8 cubes (3-faces), 1 hypercube (4-face).

Notation: *?oi?* stands for 2-face $\text{conv}(ooio, ooii, ioio, ioii)$.

Who wins?

Non-noised TFT won the first Axelrod's tournament (against many sophisticated strategies) and has steady good results against any opponents (robust strategy).

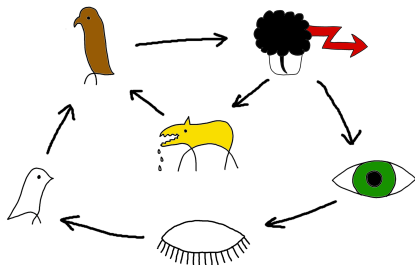
oooo (AllD) is an evolutionary stable strategy (ESS) in any noised version of IPD, but scores poorly.

iooi (Pavlov) and *iooo* (Grim trigger) can be ESS too for some game settings.

There is no universal answer which strategy is best.

Evolution dynamics

[Nowak & Sigmund 1993]: Simulated evolution of strategies.



p is a *best reply* to q if $u(p, q) \geq u(\bar{p}, q)$ for every \bar{p} .

(p, q) is a *Nash equilibrium* if p is BR to q and q is BR to p .

Nash equilibria are “stable islands” in the evolution drift.

Markov chain for two fixed strategies

Probability distribution $a = [a_0, a_1, a_2, a_3]$ of round n provides distribution aN of round $n + 1$ by transition matrix

$$N = \begin{pmatrix} p_0 q_0 & p_0(1 - q_0) & (1 - p_0)q_0 & (1 - p_0)(1 - q_0) \\ p_1 q_2 & p_1(1 - q_2) & (1 - p_1)q_2 & (1 - p_1)(1 - q_2) \\ p_2 q_1 & p_2(1 - q_1) & (1 - p_2)q_1 & (1 - p_2)(1 - q_1) \\ p_3 q_3 & p_3(1 - q_3) & (1 - p_3)q_3 & (1 - p_3)(1 - q_3) \end{pmatrix}$$





The stationary vector s satisfies $s = sN$, i. e. it is a normalized eigenvector for $\lambda = 1$.

s is a unique solution of $s(M - I) = [0, 0, 0, 1]$ where

$$M = \begin{pmatrix} p_0 q_0 - 1 & p_0 - 1 & q_0 - 1 \\ p_1 q_2 & p_1 - 1 & q_2 \\ p_2 q_1 & p_2 & q_1 - 1 \\ p_3 q_3 & p_3 & q_3 \end{pmatrix}.$$

Noised TFT vs. noised TFT

$$s = [0.25, 0.25, 0.25, 0.25]$$

CC		$0.25 \cdot 3 =$	0.75
CD		$0.25 \cdot 0 =$	0
DC		$0.25 \cdot 5 =$	1.25
DD		$0.25 \cdot 1 =$	0.25
			<hr/>
			2.25

Pay-off in Markov games

$w = [w_0, w_1, w_2, w_3] \dots$ the pay-off vector ($[3, 0, 5, 1]$).

Mean pay-off: $u = s_0 w_0 + s_1 w_1 + s_2 w_2 + s_3 w_3$.

s can be also calculated by Cramer's rule: $s_j = |M_j \mathbf{1}| / |M \mathbf{1}|$.

[Press & Dyson 2012] Using the Laplace expansion,

$$u = \frac{|M \ w|}{|M \ \mathbf{1}|}.$$

To find a best reply to q means to maximize u in variable p . So, let us derive it in parameters p_j .

Gradient of u

p_0 occupies only the first row of M , thus can be separated:

$$\begin{vmatrix} m_0 & w_0 \\ \bar{M} & \bar{w} \end{vmatrix} = p_0 \begin{vmatrix} m'_0 & 0 \\ \bar{M} & \bar{w} \end{vmatrix} + \begin{vmatrix} n_0 & w_0 \\ \bar{M} & \bar{w} \end{vmatrix}, \quad \begin{vmatrix} m_0 & 1 \\ \bar{M} & 1 \end{vmatrix} = p_0 \begin{vmatrix} m'_0 & 0 \\ \bar{M} & 1 \end{vmatrix} + \begin{vmatrix} n_0 & 1 \\ \bar{M} & 1 \end{vmatrix}$$

where m'_0 is a derivation of the first row m_0 , and n_0 its evaluation at $p_0 = 0$, \bar{M} , \bar{w} the rests of M , w .

This makes u a *linear fractional function* in p_0 :

$$u = \frac{\alpha p_0 + \beta}{\gamma p_0 + \delta} \quad u' = \frac{\alpha\delta - \beta\gamma}{(\gamma p_0 + \delta)^2}$$

The graph of u is a hyperbola, a denominator of u' is positive, and a nominator constant. Hence, u' is of constant sign, u is either strictly increasing, strictly decreasing, or constant, and acquires its maxima on the boundary or everywhere.

Best replies — comparison algorithm

Proposition

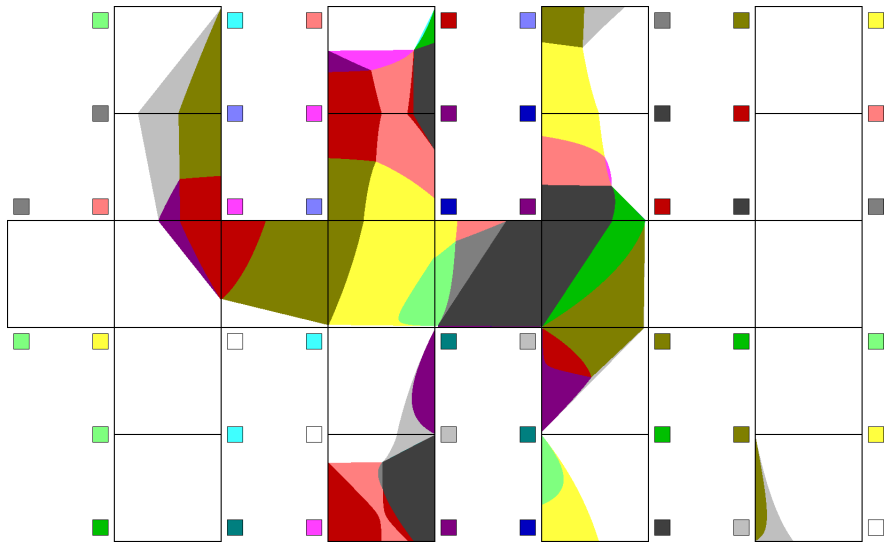
Let f be a face and q opponent's strategy.

If p is an inner point of f and a best reply to q , then any other point of f is also a best reply to q .

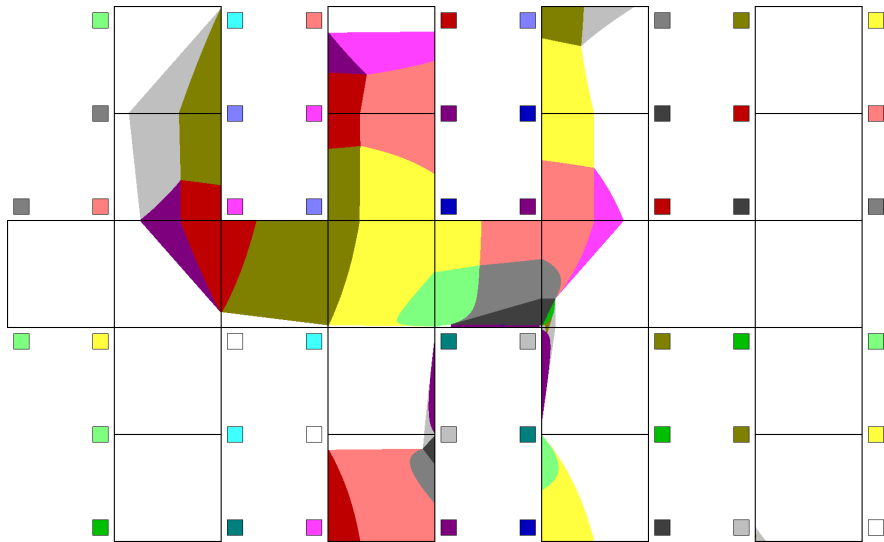
If all corners of f are best replies to q , then all points of f are also best replies.

The faces of best replies can be found by comparison of u at (finitely many) corner strategies.

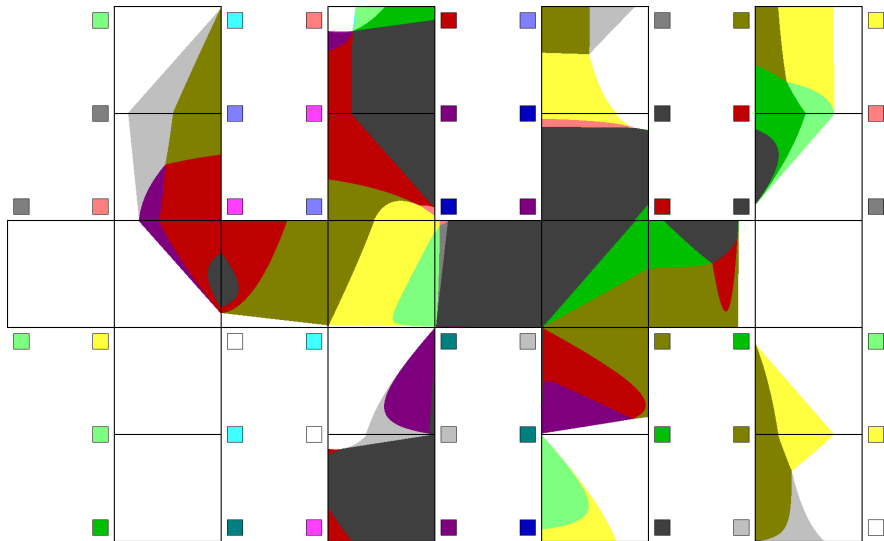
2-face net for $w = [3, 0, 5, 1]$, $e = 0.01$



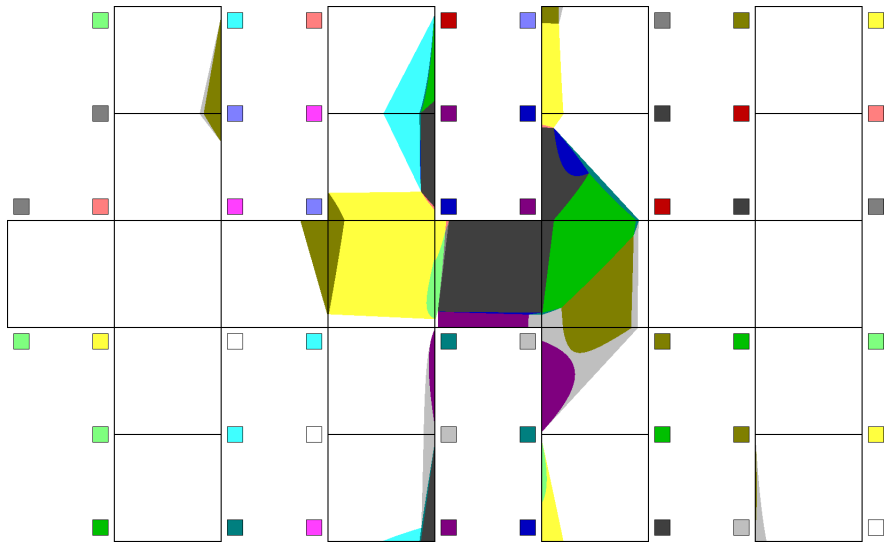
2-face net for $w = [2, 0, 9, 1]$, $e = 0.01$



2-face net for $w = [8, 0, 9, 1]$, $e = 0.01$



2-face net for $w = [8, 0, 9, 7]$, $e = 0.01$



Desnanot–Jacobi identity

[Desnanot 1819 (for $n \leq 7$), Jacobi 1841, Dodgson (Lewis Carroll) condensation 1866]:

$$\begin{vmatrix} a & m & b \\ v & A & w \\ c & n & d \end{vmatrix} \cdot |A| = \begin{vmatrix} a & m \\ v & A \end{vmatrix} \cdot \begin{vmatrix} A & w \\ n & d \end{vmatrix} - \begin{vmatrix} m & b \\ A & w \end{vmatrix} \cdot \begin{vmatrix} v & A \\ c & m \end{vmatrix}$$

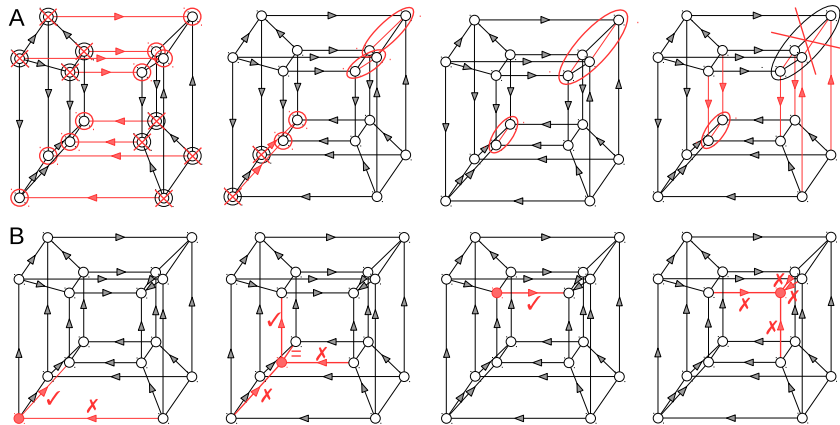
Application on $\frac{\partial u}{\partial p_j}$:

$$\frac{\partial u}{\partial p_j} = \frac{|\bar{M}|}{|M \ 1|} \cdot \frac{D_j}{|M \ 1|} \quad \text{where } D_j = \begin{vmatrix} m'_j & 0 & 0 \\ M & w & 1 \end{vmatrix}$$

and \bar{M} is M without j th row

D_j is the only factor responsible for the sign of $\frac{\partial u}{\partial p_j}$.

Sieve method, depth-first-search



Nash equilibria

Good candidates for strategies forming NE:

- ▶ corners of the hypercube,
- ▶ equalizers (next slide),
- ▶ critical points of u — boundary points of monochromatic regions.

The equalizers and critical points are solutions of one or more equations $D_j = 0$. The sets of best replies are higher-dimensional faces and can contain other critical points/equalizers → chance to find non-corner equilibria.

Equalizers

$$D_j = \begin{vmatrix} q_{\pi(j)} & 1 & 0 & 0 & 0 \\ p_0 q_0 - 1 & p_0 - 1 & q_0 - 1 & w_0 & 1 \\ p_1 q_2 & p_1 - 1 & q_2 & w_1 & 1 \\ p_2 q_1 & p_2 & q_1 - 1 & w_2 & 1 \\ p_3 q_3 & p_3 & q_3 & w_3 & 1 \end{vmatrix}$$

$$\text{for } \pi = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 2 & 1 & 3 \end{pmatrix}.$$

If the last three columns are linearly dependent then all $D_j = 0$ regardless on p . The player has a constant pay-off and every p is a best reply to q .

Such q is called *equalizer* [Boerlijst, Nowak, Sigmund 1997].

Equalizers form a plane, extremal points = intersections with 2-faces.

Two equalizers \rightarrow Nash equilibrium.

Critical points

D_j s differ only in the left upper corner ($q_{\pi(j)}$). If $D_j = D_k = 0$ then $q_{\pi(j)} = q_{\pi(k)}$ or ... (non-interesting cases).

D_j is quadratic in $q_{\pi(j)}$, linear in $q_k, k \neq \pi(j)$.

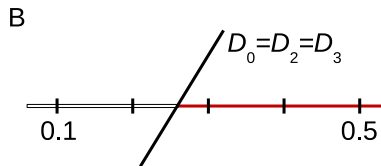
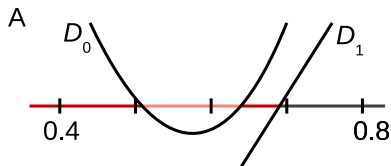
$D_j = 0$ is an unbounded quadric of “hyperbolic shape”.

Intersections of more quadrics lie on main diagonals of 2-, 3-, 4-faces.

Search of critical points is algorithmic.

We solve linear/quadratic equations in one variable!

Example: $?oio$ and $oo?o$ for $w = [3, 0, 5, 1]$, $e = 0.01$



A — three critical points, each provides a 1-face of best replies:

- ▶ Start in corner $ooio$, BR is $ioii$.
- ▶ Smallest root $e < x < 1 - e$ of some $D_j(ioii, xoio) = 0$ is $x = 0.502$ for $j = 0$, BR changes to $ooii$.
- ▶ Smallest root $x < y < 1 - e$ of some $D_j(ooii, yoio) = 0$ is $y = 0.742$ again for $j = 0$, BR changes back to $ioii$.
- ▶ Next root $y < z < 1 - e$ of some $D_j(ioii, zoio) = 0$ is $z = 0.796$ for $j = 1$, BR changes to $iiii$.
- ▶ No more roots $z < a < 1$ for $D_j(iiii, aoio) = 0$, the search is finished.

B — one critical point, BR changes from $oooo$ to $ioii$, 3-face of best replies.

Classification of Nash equilibria

Moving within the face of best replies and within the region “of the same colours” does not change pay-offs.

Nash equilibria which yield the same pay-offs are called *equivalent*.

Theorem

Every Nash equilibrium of a 2×2 game is equivalent to a situation formed by a pair of strategies from a finite set containing:

- ▶ *corners,*
- ▶ *extremal equalizers,*
- ▶ *and critical points on edges and main diagonals of faces.*

Example: $w = [3, 0, 5, 1]$, $e = 0.01$ |

critical point strategies		
strat.	value	b. r.
<i>ooxo</i>	$x = 0.266345$	<i>?o??</i>
<i>ooix</i>	$x = 0.387302$	<i>ioi?</i>
<i>ooiy</i>	$y = 0.586000$	<i>?o?o</i>
<i>oxxo</i>	$x = 0.216122$	<i>?oo?</i>
<i>oxio</i>	$x = 0.417338$	<i>io?i</i>
<i>oxix</i>	$x = 0.388532$	<i>io??</i>
<i>oixo</i>	$x = 0.027563$	<i>?oo?</i>
<i>oixx</i>	$x = 0.399779$	<i>ioo?</i>
<i>oiiy</i>	$y = 0.776061$	<i>?ooo</i>
<i>xooo</i>	$x = 0.645085$	<i>?ooo</i>
<i>xoxo</i>	$x = 0.263133$	<i>oo??</i>
<i>yoyo</i>	$y = 0.512488$	<i>??ii</i>
<i>xoio</i>	$x = 0.502042$	<i>?oii</i>
<i>yoio</i>	$y = 0.741645$	<i>?oii</i>
<i>zoio</i>	$z = 0.796101$	<i>i?ii</i>

strat.	value	b. r.
<i>xoix</i>	$x = 0.404639$	<i>?oi?</i>
<i>xxio</i>	$x = 0.449288$	<i>?o?i</i>
<i>xxix</i>	$x = 0.400732$	<i>?o??</i>
<i>xiiio</i>	$x = 0.334563$	<i>?ooi</i>
<i>xiix</i>	$x = 0.383853$	<i>?oo?</i>
<i>ioox</i>	$x = 0.010335$	<i>ioo?</i>
<i>iooy</i>	$y = 0.952823$	<i>ioo?</i>
<i>iooz</i>	$z = 0.955572$	<i>?ooo</i>
<i>ioxo</i>	$x = 0.010294$	<i>i?oo</i>
<i>ioyo</i>	$y = 0.010303$	<i>ii??</i>
<i>ioxx</i>	$x = 0.010317$	<i>i?o?</i>
<i>ioyy</i>	$y = 0.964430$	<i>i?o?</i>
<i>ioix</i>	$x = 0.658763$	<i>ii?i</i>
<i>ioiy</i>	$y = 0.964875$	<i>iiio?</i>
<i>ioiz</i>	$z = 0.965323$	<i>??oo</i>
<i>ixoo</i>	$x = 0.010309$	<i>io?o</i>

Example: $w = [3, 0, 5, 1]$, $e = 0.01$ II

critical point strategies		
strat.	value	b. r.
<i>iyoo</i>	$y = 0.969889$	<i>io?o</i>
<i>izoo</i>	$z = 0.969899$	<i>?ooo</i>
<i>ixxo</i>	$x = 0.010294$	<i>i??o</i>
<i>iyyo</i>	$y = 0.804296$	<i>o??i</i>
<i>ixxx</i>	$x = 0.010318$	<i>i???</i>
<i>iyyy</i>	$y = 0.482182$	<i>i???</i>
<i>ixio</i>	$x = 0.334257$	<i>??ii</i>
<i>iyio</i>	$y = 0.786849$	<i>oo?i</i>
<i>ixix</i>	$x = 0.328870$	<i>??ii</i>
<i>iyiy</i>	$y = 0.591202$	<i>oo??</i>
<i>iixx</i>	$x = 0.015127$	<i>ooo?</i>
<i>iyyo</i>	$y = 0.015247$	<i>o?oi</i>
<i>iizo</i>	$z = 0.655894$	<i>o?oi</i>
<i>iiix</i>	$x = 0.535109$	<i>ooo?</i>

extremal equalizers		
strat.	values	
<i>xoyo</i>	$x = 0.510000$	$y = 0.260000$
<i>xoiy</i>	$x = 0.802000$	$y = 0.594000$
<i>ixyo</i>	$x = 0.970000$	$y = 0.020000$
<i>ixiy</i>	$x = 0.323333$	$y = 0.656667$

Example: $w = [3, 0, 5, 1]$, $e = 0.01$ III

Nash equilibria without pairs of equalizers			
$oooo : oooo$	$iooo : iooo$		
$ooxo : ooxo$	$xooo : xooo$	$xoix : xoix$	$ioox : ioox$
$iooy : iooy$	$ixxo : ixxo$	$ixxx : ixxx$	$iyyy : iyyy$
$ooxo : xoxo$	$ooiy : xoxo$	$ioox : iooy$	$ioxo : ixoo$
$ioxo : iyoo$	$ixxx : iyyy$		
$ooxo : xoyo$	$ooxo : xoiy$	$ooiy : xoyo$	$xoix : xoiy$
$xxix : xoyo$	$xxix : xoiy$	$ixxo : ixyo$	$ixxx : ixyo$
$ixxx : ixiy$	$iyyy : ixyo$	$iyyy : ixiy$	

By Theorem, the list contains “essentially all” Nash equilibria.

Conclusion

- ▶ Theory works for any iterated 2×2 game.
- ▶ u is strictly monotone in each variable, hence it typically achieve maxima on the boundary. (“Rigorous” strategies prevail “infirm” strategies.)
- ▶ For calculating best replies, only corner strategies must be inspected.
- ▶ Critical points demarcating “monochromatic” regions of best replies can be found by a search on edges and diagonals. Only linear or quadratic equations must be solved.
- ▶ There is a finite set of equivalence classes of Nash equilibria. Their representatives arise from corners, extremal equalizers, and critical points.
- ▶ **The algorithms are direct and bypass dynamical modelling.**

Perspectives

- ▶ Comprehensive discussion of solvability of $D_j = 0$ w. r. t. game parameters \rightarrow ultimate classification of NE in memory-one IPD.
- ▶ Multi-player version, more actions for players ($m \times n$ games), memory-two strategies \rightarrow much more states, large determinants, need more effective methods.
- ▶ Study of polymorphic populations.



Thank you for your attention!