Cooperation in the Prisoner's Dilemma with Anonymous Random Matching

Glenn Ellison

RES 1994

Introduction

- **Folk Theorem**: "cooperative" outcome can be sustained in a sequentail equilibrium of a repeated prisoner's dilemma
  - In every period, each player knows his opponent (unique) and what his opponent did before

$1, 1$	$-l, 1+g$
$1+g, -l$	$0, 0$

- However, the result are not applicable to models of social games where a large population of players are randomly matched
  - player has limited information about other player's action
  - players probably cannot identify his opponent and thus cannot punish a certain deviator

Introduction

- To which extent Folk Theorem-type results may be obtained in such random-matching games?
  - Kandori (1992) and Harrington (1991) introduced the idea of **"contagious" punishment:** *you start to cheat once you see cheating*
  - Kandori showed that in the case of *no information processing*
    - If l is big enough, cooperation is sustainable in a sequentail equilibrium for suffiently patient players with any fixed population size
  - Kandori also argued that such an equilibrium is **fragile** in the sense that a bit noise would cause it to break down
- This paper built on Kandori's arguments
  - Introducing public randomization to adjust the severity of punishments
  - Cooperation is thus sustainable for general payoffs
  - Robust and approximately efficienct with noise
  - Extension of the result to a model without public randomization

The Model

- M players indexed by $\{1, 2, 3, ..., M\}$ where $M \geq 4$ is an even number
- In each period $t \in \{1, 2, 3, ...\}$ the players are randomly matched into pairs with player i facing player $o_i(t)$
- The pairings are independent over time and uniform: $\text{Prob}\{o_i(t) = j | h_{t-1}\} = \frac{1}{M-1}, \forall j \neq i$
- The stage game is the prisoner's dilemma shown below with positive g and nonnegative l

	C	D
C	$1, 1$	$-l, 1+g$
D	$1+g, -l$	$0, 0$

The Model

- All players have common discount factor $\delta \in (0, 1)$
  - Later this assumption would be relaxed
- Before players choose their actions in period t, they observe a **public** random variable q_t
  - q_t is drawn independently over time from $U[0, 1]$
  - Later the assumption of public randomization would be relaxed

Proposition 1

Theorem

$\exists \underline{\delta} < 1$ such that $\forall \delta \in [\underline{\delta}, 1)$ there is a sequential equilibrium $s^*(\delta)$ of this random-matching repeated game with public randomizations, where all players play C in every period along the equilibrium path.

- The strategies $s^*(\delta)$ are as follows
  - Phase I
    - Play C in period t
    - If (C, C) is the outcome for matched players i and j, both play according to phase I in period $t+1$
    - Otherwise, in period $t+1$both play according to phase II if $q_{t+1} \leq q(\delta)$ and according to phase I if $q_{t+1} > q(\delta)$
  - Phase II
    - Play D in period t
    - In period $t+1$ play according to phase II if $q_{t+1} \leq q(\delta)$ and according to phase I if $q_{t+1} > q(\delta)$
  - In period 1, all players play according to phase I

Proof of Propostion 1

- Let $f(k, \delta, q)$ be player i's continuation payoff from period t on when all players are playing the strategies above, and player i and $k-1$ others are playing according to phase II
- The continuation payoffs must satisfy two constraints derived from players not having a profitable single-period deviation
  - No profitable deviation in phase I:
    $(1-\delta)g \leq \delta q(\delta)(1 - f(2, \delta, q(\delta))) \quad (1)$
  - No profitable deviation in phase II (*facing a phase I player*):
    $(1-\delta)g \geq \delta q(\delta) E_j[f(j, \delta, q(\delta)) - f(j+1, \delta, q(\delta))] \quad (2)$
    - expectation reflects player i's beliefs over the number of players who will play according to phase II at $t+1$
    - it suffices to show it holds pointwise:
      $(1-\delta)g \geq \delta q(\delta)[f(j, \delta, q(\delta)) - f(j+1, \delta, q(\delta))] \;\; \forall j \geq 3$

Proof of Proposition 1

Lemma

$f(k, \delta, q)$ is convex in k for $k \geq 1$, i.e.
$f(k, \delta, q(\delta)) - f(k+1, \delta, q(\delta)) \geq f(k+s, \delta, q(\delta)) - f(k+s+1, \delta, q(\delta))$ $\forall s \geq 1$

- The proof of this lemma depends on the fact that the group of phase I players in a future period shrinks as more players play phase II today
  - $f(k, \delta, q) - f(k+1, \delta, q) = \sum_{t=0}^{\infty}(1-\delta)q^t\delta^t(1+g)\Pr\{\omega \in \Omega | o_1(t, \omega) \in C(t, k, \omega) \cap D(t, \omega)\}$
  - $C(t, k, \omega) \subseteq C(t, k+s, \omega)$
- Another fact we need to notice is that when (1) holds with equality, a player in phase I is exactly indifferent between playing C and D in a certain period
  - $(1-\delta)g = \delta q(\delta)(1 - f(2, \delta, q(\delta))) \iff (1-\delta)g = \delta q(\delta)(f(1, \delta, q(\delta)) - f(2, \delta, q(\delta)))$ (3)

Proof of Proposition 1

- Now what we need to show is that there exists $\underline{\delta}$ and $q(\delta)$ such that $\forall\delta \in [\underline{\delta}, 1)$, (1) and (3) both holds with equality
  - If $q(\delta) = 1$, punishments are infinite and all players would eventually be infected with probability 1 if someone already started to play D: $\lim_{\delta \longrightarrow 1} f(2, \delta, 1) = 0$
  - Thus $\lim_{\delta\to 1} \frac{\delta}{1-\delta}(1 - f(2, \delta, 1)) = \infty$ and $\lim_{\delta\to 0} \frac{\delta}{1-\delta}(1 - f(2, \delta, 1)) = 0$
  - By continuity, $\exists\underline{\delta} \in (0, 1)$ so that $\frac{\underline{\delta}}{1-\underline{\delta}}(1 - f(2, \underline{\delta}, 1)) = g$: for $\underline{\delta}$ and $q(\underline{\delta}) = 1$, (1) holds with equality and thus (3) holds with equality
  - Note that $\frac{\delta q}{1-\delta}(f(k, \delta, q)) - f(k + 1, \delta, q)) = \sum_{t=0}^{\infty}(q\delta)^{t+1}(1 + g) \Pr\{\omega \in \Omega | o_1(t, \omega) \in C(t, k, \omega) \cap D(t, \omega)\}$, RHS only depends on $q\delta$
  - Thus if we define $q(\delta) = \underline{\delta}/\delta$ for all $\delta \in [\underline{\delta}, 1)$, then (3) holds with equality for all such δ and $q(\delta)$
  - Then (1) holds with equality for all such δ and $q(\delta)$ and (2) also holds by convexity of f

Stability and Efficiency

- This equilibrium obviously satisfies the property of **global stability**: *after any finite history, the continuation payoffs of the players eventually return to the cooperative level (with probability 1)*
  - due to the introduction of public randomizations
- What if we introduce noise ε?
  - In contrast to Kandori's equilibrium, such sequantial equilibrium is also **robust to little noise**
  - Moreover, this equilibrium is **approximately efficient with little noise**

Proposition 2

Theorem

$\exists \underline{\delta}' < 1$ and a set of strategy profiles $s^*(\delta)$ for $\delta \in [\underline{\delta}', 1)$ of the random-matching game with the following three properties:
1. In the game with discount factor δ, $s^*(\delta)$ is a sequential equilibrium with all players playing C on the path in every period.
2. Define $s^*(\delta, \varepsilon)$ to be the strategy which at each history assigns probability ε to D and probability $1 - \varepsilon$ to the action given by $s^*(\delta)$. Then $\exists \bar{\varepsilon}$ such that $\forall \varepsilon < \bar{\varepsilon}$ $s^*(\delta, \varepsilon)$ is a sequential equilibrium of the perturbed game where all players are required to play D with probability at least ε at eachi history.
3. For u_i defined to player is expected per period payoff,
$\lim_{\varepsilon \longrightarrow 0} \lim_{\delta \longrightarrow 1} u_i(s^*(\delta, \varepsilon)) = 1.$

Outline of Proof

- $s^*(\delta)$ can be taken to have the same form as in Proposition 1, but with a slightly larger probablity $q'(\delta)$ of continuing in a punishment phase
- The basic idea is that the continuation payoff function f can indeed be shown as **strictly** convex in k
  - Note that generally $C(t, k, \omega) \subset C(t, k + s, \omega)$
  - Strict convexity allows us to pick a slightly larger $q'(\delta) = \underline{\delta}'/\delta$ to let the two constraints hold with strict inequality
  - Thus the equilibrium could endure noise ε

Heterogeneity in Time Preferences

- Up to now all the results require the assumption that all players share the same discount factor δ
  - Indeed, the equilibrium $s^*(\delta)$ depends on δ because we need to define $q(\delta)$
  - This seems not plausible when players have heterogeneous time preferences
- Indeed, a strategy profile s^* with similar form as before but independent of discount factor can still be a sequantial equilibrium.

Proposition 3

Theorem

There exists a strategy profile s^* and a constant $\underline{\delta}'' < 1$ such that $\forall \delta \in [\underline{\delta}'', 1)$, s^* is a sequential equilibrium of the repeated game and all players play C in every period on the path of s^*.

- Define $q''(\delta) \equiv q'' = \lim_{\delta \to 1} q'(\delta)$ $(= \lim_{\delta \to 1} \underline{\delta}'/\delta = \underline{\delta}')$ and $\underline{\delta}'' = \underline{\delta}/q''$
  - $\delta \geq \underline{\delta}'' \implies \delta q'' \geq \underline{\delta} = \underline{\delta} q(\underline{\delta}) \implies \frac{\delta q''}{1-\delta}(f(1, \delta, q'')) - f(2, \delta, q'')) \geq \frac{\underline{\delta}}{1-\underline{\delta}}(f(1, \underline{\delta}, 1)) - f(2, \underline{\delta}, 1)) = g$
  - $\delta < 1 \implies \delta q'' < q'' = \underline{\delta}' = \underline{\delta}' q'(\underline{\delta}') \implies \frac{\delta q''}{1-\delta}(f(2, \delta, q'')) - f(3, \delta, q'')) < \frac{\underline{\delta}'}{1-\underline{\delta}'}(f(2, \underline{\delta}', 1)) - f(3, \underline{\delta}', 1)) < g$
- with convexity of f the proof is finished

Proposition 4

- Public randomizations are playing two critical roles here
  - A coordination device so that all players can **simultaneously** return to cooperation at the end of a punishment phase
    - simultaneity is important because all players only slightly prefer cooperating when all others are doing so
  - To adjust the expected length and hence the severity of punishments
    - punishments are not so severe that no one is willing to carry them out
- Without public randomizations, can we still find a sequantial equilibrium to sustain cooperation and endure little noise?

Theorem

The results of Proposition 2 still hold in a model where no public randomizations are available.

Outline of Proof

- Basically, we need to find a sequential equilibrium with $q \equiv 1$
- Note that for Proposition 2, we have $q'(\underline{\delta}') = 1$ and $\underline{\delta}'$ for the two constraints to hold with strictly inequality
  - By continuity we know that $\exists \delta_1 > \underline{\delta}'$ and the two constraints still hold for any $\delta \in [\underline{\delta}', \delta_1]$ and $q \equiv 1$
  - The following lemma will then help us to finish the proof

Lemma

Let $G(\delta)$ be any repeated game of complete information, and suppose that there is a non-empty interval (δ_0, δ_1) such that $G(\delta)$ has a sequential equilibrium $s^*(\delta)$ with outcome a for all $\delta \in (\delta_0, \delta_1)$. Then $\exists \underline{\delta} < 1$ such that $\forall \delta \in (\underline{\delta}, 1)$ we can also define a strategy profile $s^{**}(\delta)$ which is also a sequantial equilibrium of $G(\delta)$ with outcome a.

- the constructed equilibrium uses **infinite periodic punishments**
- global stability will **not hold** in this case although approximate efficiency is still available

Conclusions

- "Contagious" punishments lead to a break down of cooperation, but the convexity of the breakdown process can be exploited
- Stability and limiting efficiency with noise are achievable with public randomizations
- Cooperation is also possible with heterogeneity in time preferences or without public randomizations
- With a stage game not having a dominant strategy equilibrium, whether these results could be further extended remains interesting

Learning, Local Interaction, and Coordination

Glenn Ellison

Econometrica 1993

Introduction

- Game theoretic models all too often have multiple equilibria

	A	B
A	$2,2$	$0,0$
B	$0,0$	$1,1$

- Why we should expect players to coordinate on a particular equilibrium
  - Whether there is any reason to believe that one equilibrium is more likely than the other
- Foster and Young(1990) and Kandori, Mailath and Rob(1993) derived strong predictions on the **evolution** of play over time
  - how players **learn** their opponents' play and **adjust** their strategies over time
- KMR(1993) showed that in the long run limit, players will achieve coordination on the particular "**risk dominant**" equilibrium
  - (A, A) in the example above

Introduction

- This paper built on KMR's work while
  - The behavioral assumptions incorporate **noise** and myopic responses by **boundedly rational** players
  - The **rate** at which each dynamic process converges is considered
    - In reality it is important whether the evolutionary forces would be felt within a reasonable time horizon
  - The **nature of the interations** within a population plays a crucial determinant of play
    - KMR used uniform matching rule while two extreme cases described as **uniform** and **local** are considered here

Repeated Coordination Games

- A large population of N players
- A repeated coordination game played in periods $t = 1, 2, 3, ...$
  - $a - d > b - c \Longrightarrow (A, A)$ is "risk dominant" equilibrium

	A	B
A	a, a	c, d
B	d, c	b, b

- In each period t, player i chooses an action $a_{it} \in \{A, B\}$ and his payoff is $u_i(a_{it}, a_{-i,t}) = \sum_{j \neq i} \pi_{ij} g(a_{it}, a_{jt})$
  - payoffs g are those of the 2×2 coordination game above
  - π_{ij} represents tthe probability that player i and j are matched in a given period
    - independent of t as the matching rule is time consistent

Bounded Rationality and Noise

- Boundedly rational players: $a_{it} \in \arg\max_{a_i} u_i(a_i, a_{-i,t-1})$
  - player i is reacting to the distribution of play in period $t-1$, not to the action of his matched opponent
    - fairly naive in predicting how his potential opponents would play in period t
- Disturbed by noise
  - With probability $1-2\varepsilon$ player i plays according to the rule above with probability
  - With probability 2ε player i chooses an action equally at random

Local and Uniform Matching Rules

- Uniform matching rule: $\pi_{ij} = \frac{1}{N-1} \quad \forall j \neq i$
  - With this rule, a myopic player will choose his period t strategy considering only the fraction of the population playing each strategy at time $t-1$
- "Local" matching: each player is likely to be matched only with a small fixed subset of the population
  - $2k$-neighbour matching (players are thought to be spatially distributed around a circle)
    - $\pi_{ij} = \frac{1}{2k-1} I\{i - j \equiv \pm 1, \pm 2, ..., \pm k \pmod{N}\}$
  - Probability assigned to a match is declining with distance
    - $\pi_{ij} = \begin{cases} \frac{3}{\pi^2}\frac{1}{d^2} & \text{for } d = min\{|i-j|, N-|i-j|\} \neq \frac{N}{2} \\ 1 - \frac{3}{\pi^2} \sum\limits_{|i-j| \neq N/2} \frac{1}{d^2} & \text{otherwise} \end{cases}$

Modelling Dynamics

- Assume that at some point in the past, arbitrary historical factors determined the **initial** strategies of the players
  - the behavior rules then generate a dynamic system which describes the evolution of player's strategy over time
- With uniform matching
  - Let q_i be the fraction of player i's opponents who player A in period $t-1$
  - Player i will play A in period t iff $q_i \geq q^* \equiv \frac{b-c}{(a-d)+(b-c)} < \frac{1}{2}$
  - The state of the system is denoted as a N-tuple $s_t \in S = \{A, B\}^N$, and $A(s_t)$ the total number of players playing A at t
    - The cutoff of player's response above becomes $A(s_t) > \lceil q^*(N-1) \rceil$
  - Without noise, there are two steady states, $\overrightarrow{A}$ and $\overrightarrow{B}$, with nearby states jumping to them
  - With noise ε, the transitions are governed by a Markov process
    - once play approaches either equilibrium it will likely remain nearby for a **long** period of time

Modelling Dynamics

- With local $2k$-neighbour matching (set $q^* = \frac{1}{3}$ and $k = 4$)
  - The cutoff becomes whether the number of your 8 neighbours playing A exceed 3
  - Without noise, there are at least two steady states, $\overrightarrow{A}$ and $\overrightarrow{B}$
    - both have a nontrivial attractive basin, but that of $\overrightarrow{A}$ is bigger than that of $\overrightarrow{B}$
    - with **four adjacent** players playing A at a time, the dynamic process will eventually goes to $\overrightarrow{A}$
  - With noise, the differing sizes of these attracitive basins cause **relatively rapid** convergence to $\overrightarrow{A}$
    - starting from $\overrightarrow{B}$, it is far more likely to see 4 adjacent distrubances than $\lceil (N-1)/3 \rceil$ simultaneous ones when N is large

Further Notations

- We view the time t strategy profiles as the states s_t of a Markov process
- The time t probability distribution over the states is represented by an 1×2^N vector v_t
- The evolution fo the process is governed by $v_{t+1} = v_t P(\varepsilon)$
  - $P(\varepsilon)$ is the transition matrix with $p_{ij}(\varepsilon) = \Pr\{s_{t+1} = j | s_t = i\}$
  - Write $P^u(\varepsilon)$ for uniform matching and $P^{2k}(\varepsilon)$ for local matching
- $P(\varepsilon)$ is strictly positive if $\varepsilon > 0 \Rightarrow \exists! \ \mu(\varepsilon)$ such that $\mu(\varepsilon) = \mu(\varepsilon)P(\varepsilon)$
  - Let $\mu_s(\varepsilon)$ denote the probability assigned to state s by distribution $\mu(\varepsilon)$
- Use O-approximations for the asymptotic behavior of $\mu(\varepsilon)$ as $\varepsilon \longrightarrow 0$
  - $f(x) = O(g(x))$ $(x \rightarrow 0)$ if $\exists C, c > 0$ such that $cg(x) \leq f(x) \leq Cg(x)$ for sufficiently small x

Converging to Risk Dominant Equilibrium

Theorem

For sufficiently large N we have:
(a) $\lim_{\epsilon\to 0} \mu^{u}_{\overrightarrow{A}}(\varepsilon) = 1$, $\lim_{\epsilon\to 0} \mu^{2k}_{\overrightarrow{A}}(\varepsilon) = 1$;
(b) $\mu^{u}_{\overrightarrow{B}}(\varepsilon) = O(\varepsilon^{N-2\lceil q*(N-1)\rceil+1})$, $\mu^{2}_{\overrightarrow{B}}(\varepsilon) = \begin{cases} O(\varepsilon^{N-2}) & \text{for } N \text{ even} \\ O(\varepsilon^{N-1}) & \text{for } N \text{ odd} \end{cases}$

- The proof does not rely on the fact that the matching distribution has finite support
  - the matching rule with declining probability also works, even with $N \to \infty$
- The matching rule can not be too concentrated
  - If $\pi_{ij} > 1 - q^*$ then the probability of the cycle where i and j alternatively play (A, B) and (B, A)
- The long-run outcome may differ between the two matching rules when we move beyond 2×2 games.

Rates of Convergence

- Theorem 1 implies that if the coordination games are repeated enough times we expect to see the risk dominant equilibrium played almost all the time
- Whether this "eventually" is relevant depends on the rate of convergence
- Let ρ be an arbitrary initial state $\Rightarrow \mu(\varepsilon) = \lim_{t\to\infty} \rho P(\varepsilon)^t$
- Define $\| \mu - \nu \| \equiv \max_{s\in S} |\mu_s - \nu_s|$
- Define $r^u(\varepsilon) = \sup_{\rho\in\Delta} \limsup_{t\to\infty} \| \rho P^u(\varepsilon)^t - \mu^u(\varepsilon) \|^{1/t}$ and $r^2(\varepsilon) = \sup_{\rho\in\Delta} \limsup_{t\to\infty} \| \rho P^2(\varepsilon)^t - \mu^2(\varepsilon) \|^{1/t}$

Theorem

Assume $\lceil q^*(N-1) \rceil < N/2$, as $\varepsilon \to 0$ we have:
$1 - r^u(\varepsilon) = O(\varepsilon^{\lceil q^*(N-1)\rceil})$, $1 - r^2(\varepsilon) = O(\varepsilon)$.

Rates of Convergence

- Loosely speaking, $\| \rho P^u(\varepsilon)^t - \mu^u(\varepsilon) \| = O(r^t)$ for some $r < 1$
  - convergence is approximately at an exponential rate
- $r^u(\varepsilon)$ is much closer to 1 than $r^2(\varepsilon)$ for small ε, so the rate of convergence with uniform matching is much slower
- An alternate measure: $W(N, \varepsilon, \alpha) = E(\min\{t | A(s_t) \geq (1-\alpha)N\} | s_0 = \overrightarrow{B})$
  - $W(N, \varepsilon, \alpha)$ is the expected waiting time until at least $1 - \alpha$ of the players play A given that eveyone starts off playing B

Theorem

For ε sufficiently small we have:
$W^u(N, \varepsilon, \alpha) = O(\sqrt{N} e^{((q^* - \varepsilon)/\varepsilon(1-\varepsilon))N})$, $W^{2k}(N, \varepsilon, \alpha) = O(1)$

Different Matching Rules

	$W^{2k}(N, \varepsilon, \alpha)$		
	$\varepsilon = 0.025$	$\varepsilon = 0.05$	$\varepsilon = 0.1$
$k = 1$	11	8	6
$k = 2$	44	23	12
$k = 3$	93	25	11
$k = 4$	522	45	11

- For small ε, evolution is faster for more concentrated matching rules
- For large ε, evolution can be faster for less concentrated matching rules
- The assumption of players located around the circle is crucial
  - This implies a great **overlap** of the groups of neighbours
  - With less overlap(lattice of more dimensions), the evolution may be slower

Heterogeneity

- The players are assumed to have heterogeneous tastes $u_i(A,A)$ and $u_i(B,B)$ with *lognormal distributions*
  - (A,A) is still better: $u_i(A,A) \overset{D}{\sim} (17/7)u_i(B,B)$

	$W^{2k}(N,\varepsilon,\alpha)$		
$Var(u_i(B,B))$	$\varepsilon = 0.025$	$\varepsilon = 0.5$	$\varepsilon = 0.1$
0	522	45	11
0.1	75	19	9
0.2	28	14	7

- Heterogeneity increases the rate of convergence (especially when ε is small)
  - Stable clusters for players with great utility from (A,A) is smaller
- When evolution is already rapid for a homogeneous population, heterogeneity only has limited effect

Conclusion

- Boundedly rational players' myopic adjustments creat evolutionary forces which may select among the equilibria
- The nature of the matching rule helps us weight historical factors and evolutionary forces
  - With uniform matching among a large population play will reflect arbitrary historical factors for a long period of time
  - With local matching evolutionary forces may be felt early in the game