ゲ[illegible]と
進[illegible]ス

第4[illegible]学習

B4　近松京介

第4章　相互依存的な学習

■協調的関係

複数の均衡解 ➡ 優劣関係が存在

but　多くの人が劣勢な戦略

➡ 一人ひとりにとって同じように劣勢な戦略を選択するのが合理的

■相補的関係

多数の主体の選択に分散 ➡ 一人一人に利益

but　一人一人が合理的にふるまう

➡ 全員同じ戦略 ➡ 望ましい均衡解ではない

このような悪循環から脱出する学習法は？？

協調的関係と相補的関係で臨まれる学習法は違うのか？？

（1）最適反応学習

他の主体と相互依存関係 ⟺ 自己利益最大化

so　相手戦略を知る必要性！

仮想ゲーム：お互いに相手戦略を予想して行われる相互作用

仮想ゲームが繰り返し行われることを前提

⇒ 相手戦略に逐次適応しながら自らの利得を最大にする戦略をとる

（2）自然淘汰による学習

三章の通り（優れた性質をもつ戦略が生き残る）

（3）強化学習

人間・動物

like　過去にうまくいった行動

dislike　過去にうまくいかなかった行動

基準

戦略

このような方法に基づき，確率的に戦略を決定する学習法

※あくまでも自分の利得を基準に戦略を決定する

（4）模倣

将来を完全に見通して，合理的な計算に基づき戦略を決定 ➡ 困難

自分よりも多く利得を獲得した人の戦略を採用して適応 ➡ 有効

4.1.2　学習形態による分類

(1)個人学習

個人的に獲得する利得を基準に，優れた戦略を主体が求めること

最適反応学習　強化学習

(2)社会(集団)学習

社会全体で，優れている戦略を採用する主体が生き残りるとして扱う

自然淘汰に基づく学習

(3)共同学習

お互いに相手の優れたものを取り入れること

※相互主義的な姿勢が必要

※メタレベルからの関係性の追求が重要になる

（1）戦略の学習

各主体が自らの利得を最大にする戦略を探索すること

（2）メタ戦略の学習

相互依存関係を長期的に維持

→ 相手戦略との組み合わせに基づき戦略を決定

メタ戦略

お互いの戦略の組み合わせの関数として表される戦略決定のルール

各主体が相互作用を繰り返しながら，優れたメタ戦略を追及すること

■タカハトゲームで考える

xの戦略 \ yの戦略	S_1 (タカ派)	S_2 (ハト派)
S_1 (タカ派)	(v-c)/2 ; (v-c)/2	v ; 0
S_2 (ハト派)	0 ; v	v/2 ; v/2

N人の集団

n人が戦略S_1

xの割合で戦略S_1

ある主体が他のすべての主体と相互作用する場合の総利得

$$\begin{aligned}\overline{U}(S_1) &= (n-1)(v-c)/2 + (N-n-1)v, \\ \overline{U}(S_2) &= (N-n-1)v/2\end{aligned} \tag{4.1}$$

ある主体が一人の主体と相互作用する場合の平均利得

$$\begin{aligned}U(S_1) &= \overline{U}(S_1)/N \cong x(v-c)/2 + (1-x)v, \\ U(S_2) &= \overline{U}(S_2)/N = (1-x)v/2\end{aligned} \tag{4.2}$$

集団全体での一人あたりの平均利得

$$xU(S_1) + (1-x)U(S_2) = (v - cx^2)/2 \tag{4.3}$$

$x = 0$のとき，平均利得は最大値$v/2$をとる

個人の立場に立つと…

$$\begin{aligned} U(S_1) &= \overline{U}(S_1)/N \cong x(v-c)/2 + (1-x)v, \\ U(S_2) &= \overline{U}(S_2)/N = (1-x)v/2 \end{aligned} \tag{4.2}$$

$$\begin{array}{lll} (1) & U(S_1) \geqq U(S_2) \quad \text{ならば}, & \text{戦略 } S_1 \\ (2) & U(S_1) < U(S_2) \quad \text{ならば}, & \text{戦略 } S_2 \end{array} \tag{4.5}$$

$$\begin{array}{lll} (1) & x \leqq (v/c) \quad \text{ならば}, & \text{戦略 } S_1 \\ (2) & x > (v/c) \quad \text{ならば}, & \text{戦略 } S_2 \end{array} \tag{4.6}$$

4.2　最適反応学習と均衡解の選択性 10

仮定

すべての主体が最適反応学習を行う
集団の戦略分布を各主体は共通に知ることができる

$x(t)$がv/c以下のときは，次の時点$t+1$では，全員がタカ派戦略S_1

次の時点$t+2$では，全員がハト派戦略S_2

繰り返す

各主体の平均利得

$$((v-c)/2+v/2)/2 = v/2 - c/4$$

4.2　最適反応学習と均衡解の選択性 11

■相補ゲーム

表 4.1　相補ゲームの利得行列

自分の戦略＼相手の戦略	S_1	S_2
S_1	0, 0	1, 1
S_2	1, 1	0, 0

(i)

$$(S_1, S_2) \to (S_1, S_2) \to (S_1, S_2) \to \cdots$$

(ii) 悪循環

$$(S_1, S_1) \to (S_2, S_2) \to (S_1, S_1) \to \cdots$$

■協調ゲーム

表 4.2　協調ゲームの利得行列
($\alpha_i > 0,\ \beta_i > 0,\ i = A, B$)

A の戦略＼B の戦略	S_1	S_2
S_1	α_A, α_B	0, 0
S_2	0, 0	β_A, β_B

(i)

$$(S_1, S_1) \to (S_1, S_1) \to (S_1, S_1) \to \cdots$$

(ii) 悪循環

$$(S_1, S_2) \to (S_2, S_1) \to (S_1, S_1) \to \cdots$$

■混合戦略における最適反応学習

協調ゲーム

$$U_A(x,y) = \alpha_A xy + \beta_A(1-x)(1-y)$$

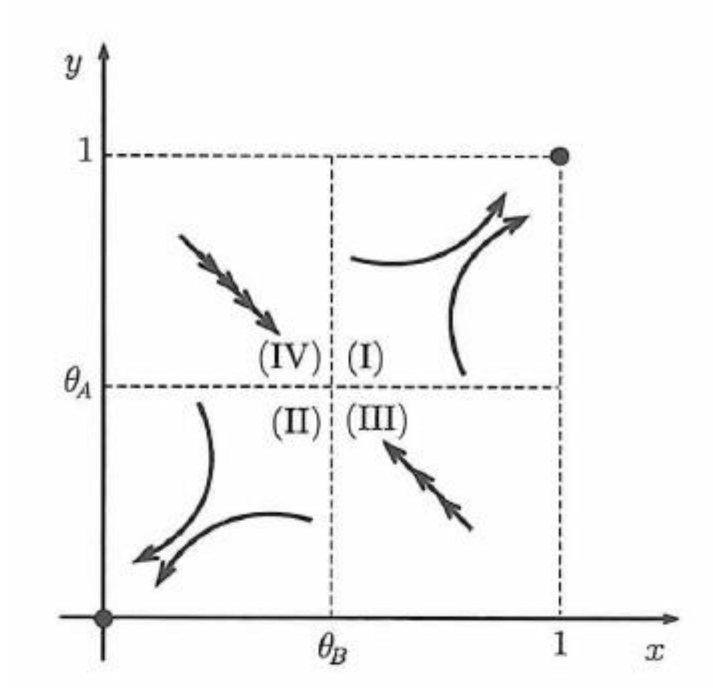
$$U_{\mathrm{B}}(x,y) = \alpha_{\mathrm{B}} xy + \beta_{\mathrm{B}}(1-x)(1-y)$$

$$\phi_A(y) = \arg\underset{x\in[0,1]}{Max}\, U_A(x,y)$$

$$\phi_B(x) = \arg\underset{y\in[0,1]}{Max}\, U_B(x,y)$$

$$\dot{x}(t) = \phi_A(y) - x(t)$$

$$\dot{y}(t) = \phi_B(x) - y(t)$$



混合戦略に拡大しても最適反応学習のもつ特徴は変わらない

■不確実性を考慮した最適反応学習

協調ゲーム

主体Aと主体Bがとる純粋戦略の組を次の状態変数で表す

$$s_1 = (S_1, S_1), \qquad s_2 = (S_1, S_2),$$
$$s_3 = (S_2, S_1), \qquad s_4 = (S_2, S_2)$$

普通は… $s_1 \to s_1 \to s_1 \to$

$s_2 \to s_3 \to s_2 \to$

$s_4 \to s_4 \to s_4 \to$

最適戦略がS_1のとき，確率εで戦略S_2を選択すると仮定

$$\begin{array}{cc} & \begin{array}{cccc} s_1 & s_2 & s_3 & s_4 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \end{array} & P = \begin{bmatrix} (1-\varepsilon)^2, & (1-\varepsilon)\varepsilon, & (1-\varepsilon)\varepsilon, & \varepsilon^2 \\ (1-\varepsilon)\varepsilon, & \varepsilon^2, & (1-\varepsilon)^2, & (1-\varepsilon)\varepsilon \\ (1-\varepsilon)\varepsilon, & (1-\varepsilon)^2, & \varepsilon^2, & (1-\varepsilon)\varepsilon \\ \varepsilon^2, & (1-\varepsilon)\varepsilon, & (1-\varepsilon)\varepsilon, & (1-\varepsilon)^2 \end{bmatrix} \end{array}$$

4.2　最適反応学習と均衡解の選択性 14

両者戦略の組が，時点tにおいて，状態s_1, s_2, s_3, s_4のいずれかにある確率

$$\boldsymbol{\lambda}(t) = (\lambda_1(t), \lambda_2(t), \lambda_3(t), \lambda_4(t))$$

次点t+1において四つの状態のいずれかにある確率は次式

$$\boldsymbol{\lambda}(t+1) = \boldsymbol{\lambda}(t)P$$

このダイナミクスの均衡解は次の方程式で求まる

$$\overline{\boldsymbol{\lambda}} = \overline{\boldsymbol{\lambda}}P \quad \Rightarrow \quad \bar{\lambda} = (E - P)^{-1}$$

この方程式より $\overline{\boldsymbol{\lambda}} = (1/4, 1/4, 1/4, 1/4)$ と求まる

協調ゲームのお互いに同じ純粋戦略を選択することで利得が得られる状態とはかけ離れている

協調ゲームと相補ゲーム → 純粋戦略による二つの均衡解 / 混合戦略による一つの均衡解

戦略的な不確実性

どの均衡解が果たして実現されるのかが不確定

ナッシュ均衡解の洗練

複数の均衡が気が存在する中で優れた均衡解が選択されるための条件を明らかにすること

自然淘汰と均衡解の選択性

自然淘汰プロセス＋不確実性 ➡ 均衡解の選択性はどうなるか

戦略S_1をパレート優越戦略，S_2をリスク優越戦略とする

■不確実性を考えない場合

戦略の初期分布によりどちらの純粋戦略になるか決まる

■不確実性を考える場合

戦略の初期分布によらず全員リスク優越戦略を選択する

逆淘汰

劣性な戦略によって有意な戦略が退けられてしまうこと

相対主義

他の主体の存在を自分と同じ立場に立って認識すること

功利主義

自分さえよければいい！という考え方

■二人の各主体が同じ混合戦略xをとるときの期待利得

$$\begin{aligned} U(\boldsymbol{x},\boldsymbol{x}) &= \boldsymbol{x}U(\boldsymbol{e}_1,\boldsymbol{x}) + (1-\boldsymbol{x})U(\boldsymbol{e}_2,\boldsymbol{x}) \\ &= (a+d-b-c)x^2 + (b+c-2d)x + d \end{aligned}$$

■ジレンマゲーム

$$\text{(i)} \quad c > a > b > d, \qquad \text{(ii)} \quad 2a > b + c$$

以上の条件により，$U(x, x)$は$x \in [0,1]$で増加関数

$X = 1$で最大値をとり，このことから相対主義に基づく学習によって双方とも協調戦略S_1を選択 → ジレンマ解決

■協調ゲーム

レプリケータダイナミクスより，パレート優越戦略が選択される可能性が高い

■相補ゲーム

$U(x,x)$が最大値をとるのは$x=1/2$のとき

相対主義に基づく学習では，利得パラメータに依存することなく戦略S_1とS_2を選択する主体が半分ずつに分かれる

次のような戦略と利得行列の場合を考える

戦略 S_1：強固な態度をとる，　戦略 S_2：譲歩的な態度をとる

Bの戦略 \ Aの戦略	S_1 (強気)	S_2 (譲歩)
S_1 (強気)	0 0	1 α
S_2 (譲歩)	α 1	0 0

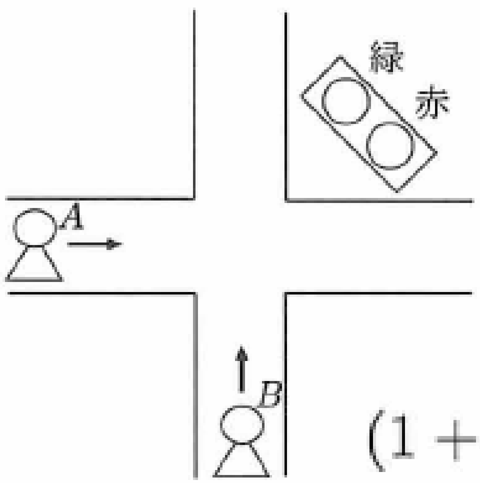
	平均利得
レプリケータダイナミクス	$2\alpha/(\alpha+1)^2$
相対主義	$(1+\alpha)/4$

譲り合い学習

利得を獲得できなかった人にゆずるために次回の戦略を変更

利得を獲得できなかった主体はランダムに戦略を選択

共通の信号などに関連がつけられて決定される戦略を，相関戦略



主体 A の期待利得は $1-p(1-\alpha)$

主体 B の期待利得は $1+p(1-\alpha)$

お互いの期待利得は等しく，$(1+\alpha)/2$

$$(1+\alpha)/2 > \alpha/(1+\alpha)$$

相関戦略による期待利得＞ナッシュ均衡での利得

■二人の主体間での譲り合いに基づく戦略決定ルール

戦略 S_1：前進する (go), 戦略 S_2：停止する (stop)

$\omega_i(t)=1$：利得 1 を獲得, $\omega_i(t)=0$：利得はゼロ

$$
\begin{aligned}
&(1) \quad (\omega_i(t)=1)\wedge(a_i(t)=1) \Longrightarrow a_i(t+1)=0\\
&(2) \quad (\omega_i(t)=1)\wedge(a_i(t)=0) \Longrightarrow a_i(t+1)=1\\
&(3) \quad (\omega_i(t)=0)\wedge(a_i(t)=1) \Longrightarrow a_i(t+1)=RND(x)\\
&(4) \quad (\omega_i(t)=0)\wedge(a_i(t)=0) \Longrightarrow a_i(t+1)=RND(y)
\end{aligned}
$$

$$s_1=(S_1,S_1),\ s_2=(S_1,S_2),\ s_3=(S_2,S_1),\ s_4=(S_2,S_2),$$

$$
P=\begin{bmatrix}
y^2 & y(1-y) & y(1-y) & y(1-y)^2\\
0 & 0 & 1 & 0\\
0 & 1 & 0 & 0\\
(1-x)^2 & x(1-x) & x(1-x) & x^2
\end{bmatrix}
$$

$$\boldsymbol{\lambda}(t) = (\lambda_1(t), \lambda_2(t), \lambda_3(t), \lambda_4(t))$$

$$\boldsymbol{\lambda}(t+1) = \boldsymbol{\lambda}(t)P$$

$$\overline{\boldsymbol{\lambda}} = \overline{\boldsymbol{\lambda}}P$$

$$\overline{\boldsymbol{\lambda}} = (0, 1/2, 1/2, 0)$$

状態s2=(S1,S2)と状態s3=(S2,S1)が等確率0.5で生起

等しく平均利得 $(1+\alpha)/2$

少数派ゲーム

各主体にとって，多くの人がとる戦略とは異なることが望ましい相互依存問題

主体の利得の与え方が二通り

$$(1)\quad u_i(t) = -a_i(t)\,\mathrm{sgn}(A(t))$$

$$(2)\quad u_i(t) = -a_i(t)A(t)/N$$

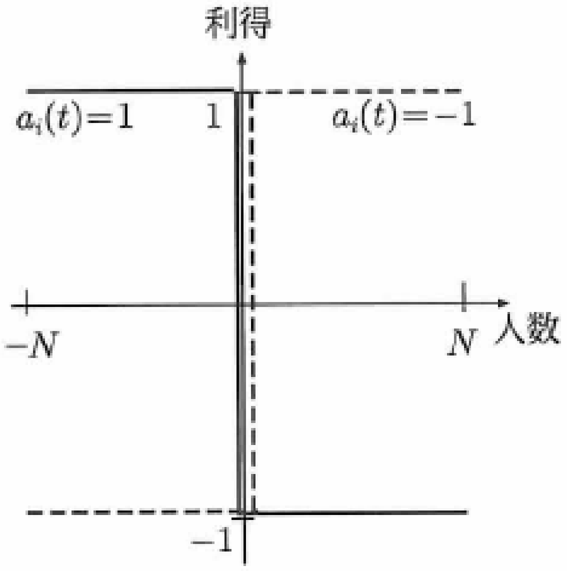


図 **4.3**　少数派ゲームの利得関数 (1)

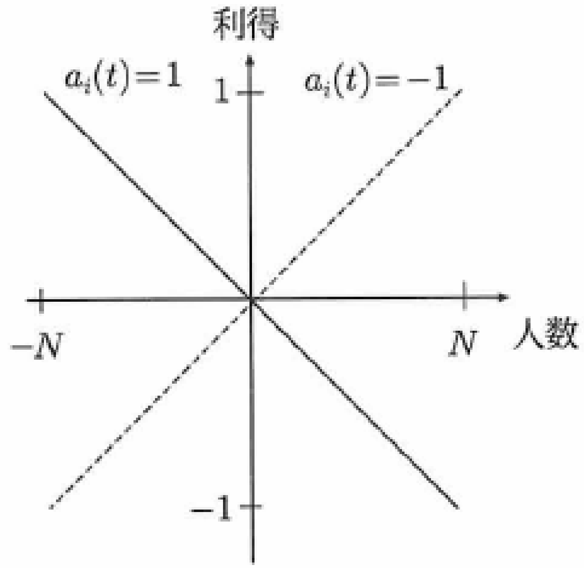


図 **4.4**　少数派ゲームの利得関数 (2)

この利得行列で考えてみる

表 4.4　少数派ゲームの利得行列

自分の戦略 ＼ 相手の戦略	S_1	S_2
S_1	-1 / -1	1 / 1
S_2	1 / 1	-1 / -1

$$\overline{U}_i(S_1) = -n + N - n = -A(t), \quad \overline{U}_i(S_2) = n - (N - n) = A(t) \quad (4.37)$$

$$U(S_1) = \overline{U}(S_1)/N = -A(t)/N, \quad U(S_2) = \overline{U}(S_2)/N = A(t)/N \quad (4.38)$$

(2)　$u_i(t) = -a_i(t)A(t)/N$　に一致

多人数による小人数ゲームの解はこの利得行列の均衡解と一致

仮定

バーに行きたいと考える人：N人

バーの収容人員：Nc人

定員を超える人がバーに押しかけた場合には誰一人満足することはできない

戦略 S_1：バーに行く (go)　　戦略 S_2：家にいる (stay)

バーに行く人数は，$A(t) = \sum_{1 \leqq i \leqq N} a_i(t)$

状態 ω_1：バーは混雑している　$(A(t) > N_c)$,

状態 ω_2：バーは空いている　$(A(t) \leqq N_c)$

戦略S_1を選択する人数:$N_1(t)$
戦略S_2を選択する人数:$N_2(t)$

戦略S_1を選択する人の割合:$p(t) = N_1(t)/N$

バーの人数から決まる定員率:$\theta = N_c/N$

主体iの利得関数$u_i(t)$を二つの方式で与える

〈方式 1〉

(1)　$a_i(t) = 1,\ p(t) \leqq \theta$　(バーに行きバーは空いている)　$u_i(t) = 1$
(2)　$a_i(t) = 1,\ p(t) > \theta$　(バーに行きバーは混雑)　$u_i(t) = 0$
(3)　$a_i(t) = 0,\ p(t) \leqq \theta$　(家にいてバーは空いている)　$u_i(t) = 0$
(4)　$a_i(t) = 0,\ p(t) > \theta$　(家にいてバーは混雑)　$u_i(t) = 1$

〈方式 2〉

(1)　$a_i(t) = 1$ のとき，　$u_i(t) = 2\theta(1 - p(t))$
(2)　$a_i(t) = 0$ のとき，　$u_i(t) = 2(1 - \theta)p(t)$

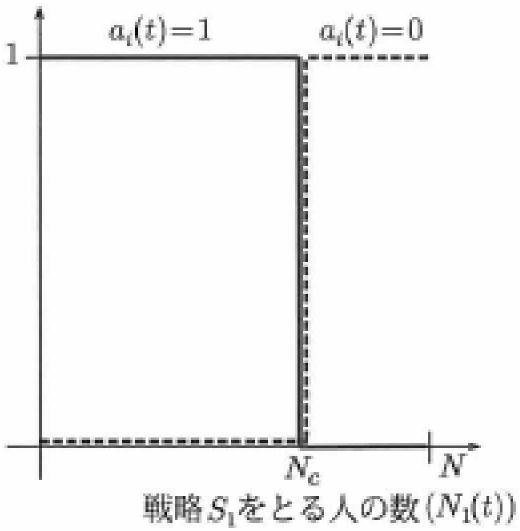


図 4.5 非対称少数派ゲームの利得関数 (1)

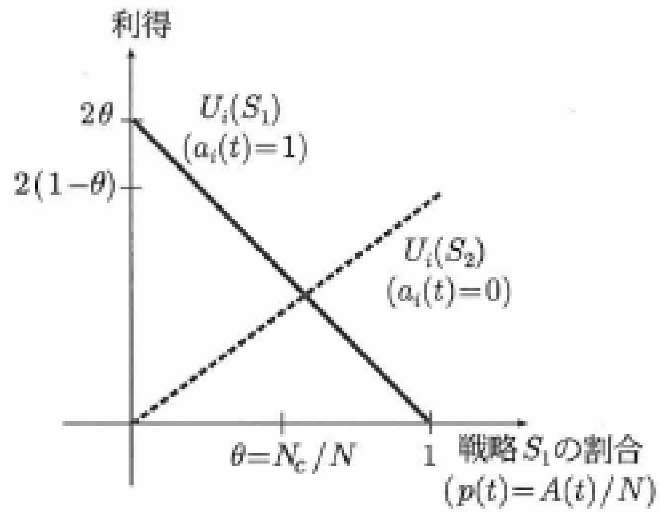


図 4.6 非対称少数派ゲームの利得関数 (2)

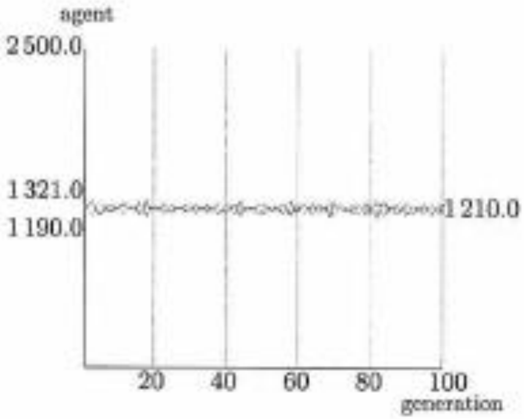
表 4.4 少数派ゲームの利得行列

相手の戦略 / 自分の戦略	S_1	S_2
S_1	-1 -1	1 1
S_2	1 1	-1 -1

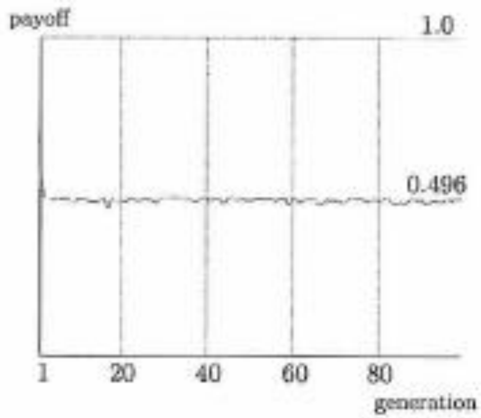
表 4.5 非対称な少数派ゲームの利得行列

相手の戦略 / 自分の戦略	S_1	S_2
S_1	0 0	$2(1-\theta)$ 2θ
S_2	2θ $2(1-\theta)$	0 0

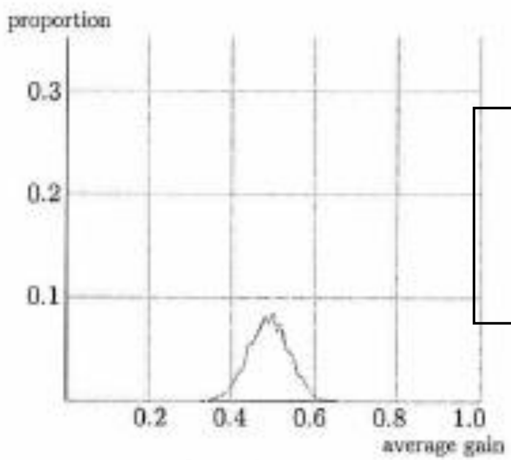
■対称な少数派ゲーム



(a)　各戦略を選択する主体数の推移



(b)　平均利得の推移



(c)　各主体の利得分析

利得は約0.35～0.65の範囲で広く分布

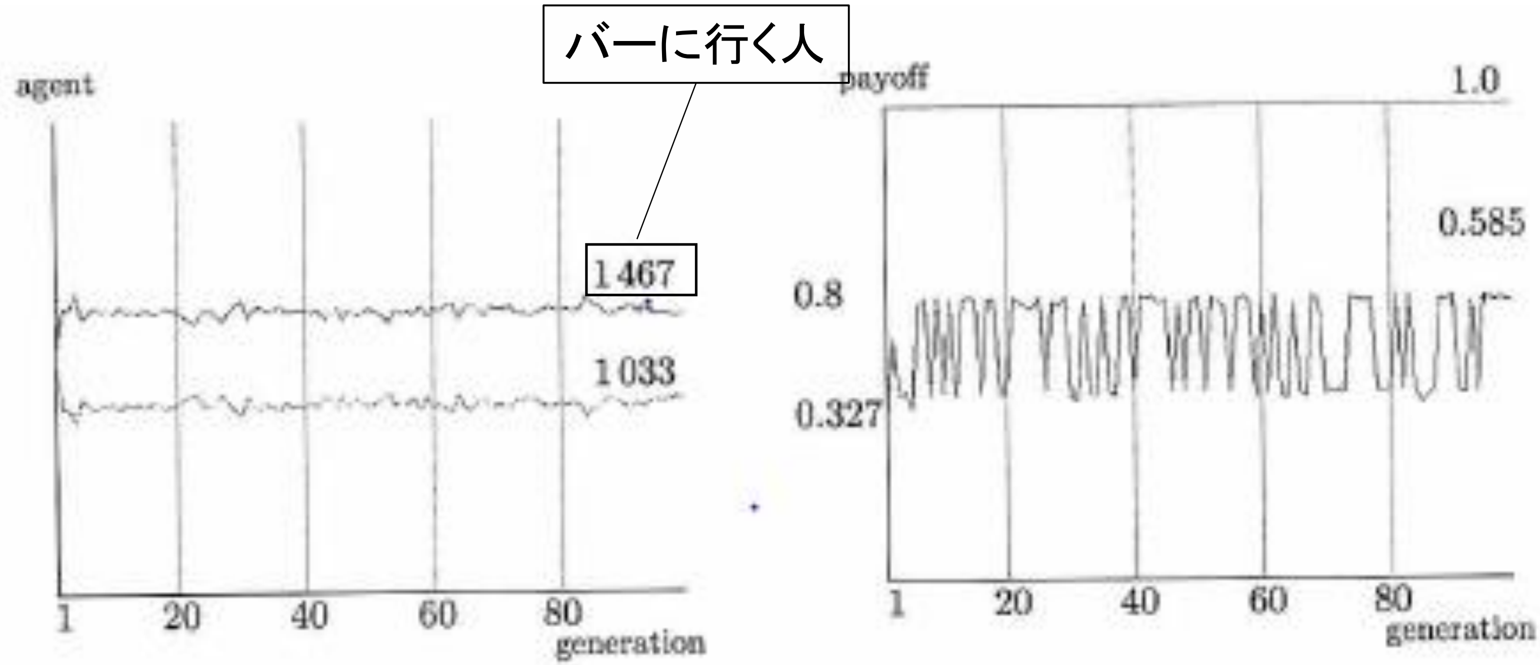
少数派に長くとどまることで多くの利得を獲得

多数派であった期間が長く利得の少ない主体

図 4.7　ナッシュ均衡戦略の下での集合行為 (定員率 $\theta = 0.5$)

ナッシュ均衡戦略では
不公平な集合行為

■非対称な少数派ゲーム（θ＝0.6）



(a)　各戦略を選択する主体数の推移　　(b)　平均利得の推移

図 **4.8**　ナッシュ均衡戦略の下での集合行為 (定員率 $\theta = 0.6$)

一人当たりの平均利得は対称な少数派ゲームと異なり，$1 - \theta = 0.4$と$\theta = 0.6$の間で変動

■非対称な少数派ゲーム（θ＝0.8）

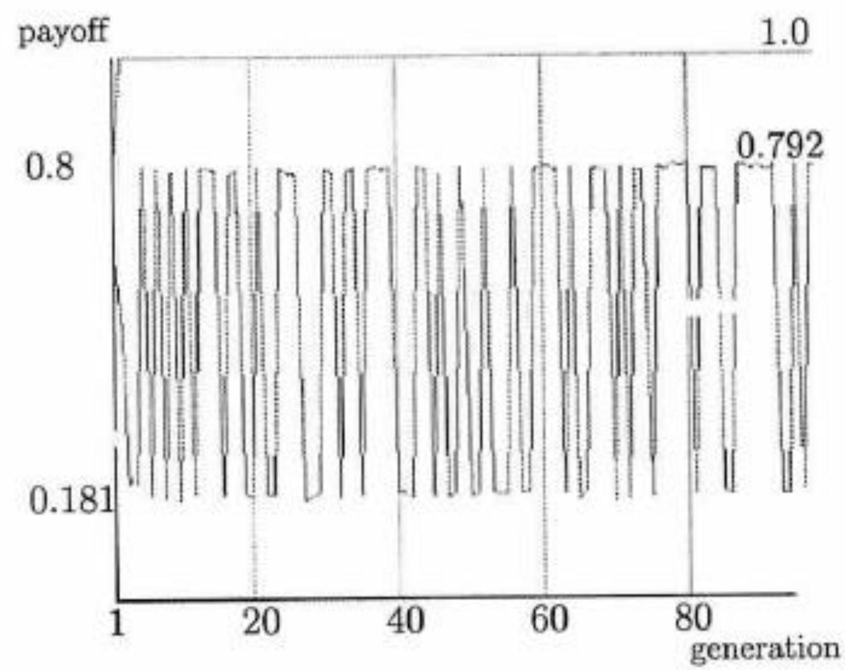


図 4.9　ナッシュ均衡戦略の下での平均利得の推移 (定員率 $\theta = 0.8$)

一人当たりの平均利得は対称な少数派ゲームと異なり，$1-\theta = 0.2$と$\theta = 0.8$の間で変動

4.5.2　譲り合い学習の下での集合行為

■二人の主体間での譲り合いに基づく戦略決定ルール

$$
\begin{array}{ll}
(1) & (\omega_i(t)=1)\wedge(a_i(t)=1)\Longrightarrow a_i(t+1)=0\\
(2) & (\omega_i(t)=1)\wedge(a_i(t)=0)\Longrightarrow a_i(t+1)=1\\
(3) & (\omega_i(t)=0)\wedge(a_i(t)=1)\Longrightarrow a_i(t+1)=RND(x)\\
(4) & (\omega_i(t)=0)\wedge(a_i(t)=0)\Longrightarrow a_i(t+1)=RND(y)
\end{array}
$$

拡張

■多人数の間の譲り合いルール

〈少数派で利得を獲得できた人〉

利得を獲得できたすべてが，つぎの時点で戦略を変更しても収容人員の制約条件を満たす場合には，全員戦略を変更する．一方で，そのことで定員上の制約条件を満たさなくなる場合には，一部の人が戦略を変更し，残りの人は再び同じ戦略を選択する．

〈多数派で利得を獲得できなかった人〉

利得を獲得できなかった主体すべてが，つぎの時点で戦略を変更しても収容人員をオーバーしない場合には，全員戦略を変更する．一方で，つぎの時点で

(1)　対称な少数派ゲーム ($\theta = 0.5$)

時点 t で，バーに行った人が定員をオーバーしたとする．すなわち，

〔ケース 1〕

$$N_2(t) < N_c < N_1(t) \tag{4.47}$$

$$N_1(t+1) = xN_1(t) + N_2(t)$$

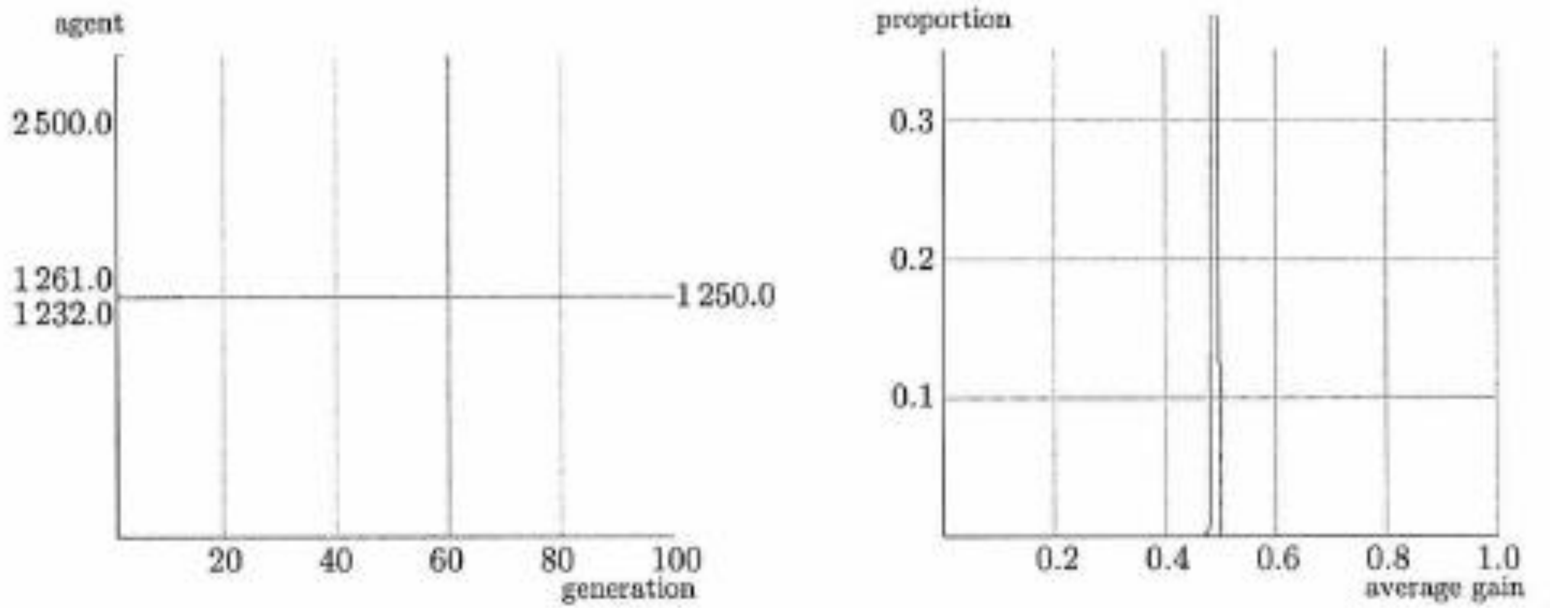
$$x = (N_c - N_2(t))/N_1(t)$$

〔ケース 2〕

$$N_1(t) < N_c < N_2(t) \tag{4.50}$$

$$N_1(t+1) = yN_2(t)$$

$$y = N_c/N_2(t)$$



(a)　各戦略を選択する主体数の推移　　(b)　各主体の利得分布

図 4.10　譲り合いの下での集合行為 (定員率 $\theta = 0.5$)

(1)　$N_c < N_2(t)$ のとき

(ア)　時点 t で戦略 S_1 を選択した人は，つぎの時点 $t+1$ では戦略を変更し，戦略 S_2 を選択する．

(イ)　戦略 S_2 を選択した人で，割合 $y = N_c/N_2(t)$ に相当する人が，つぎの時点 $t+1$ で戦略 S_1 に変更し，残りの人は同じ戦略 S_2 を繰り返す．

(2)　$N_c > N_2(t)$ のとき

(ア)　戦略 S_2 を選択した人は，つぎの時点 $t+1$ では戦略を変更し，戦略 S_1 を選択する．

(イ)　時点 t で戦略 S_1 を選択した人は，割合 $x = \{N_c - N_2(t)\}/N_1(t)$ に相当する人がつぎの時点 $t+1$ で戦略 S_2 に変更し，残りの人は同じ戦略 S_1 を繰り返す．

少数派ゲームの非対称性が増すと，集合行為が理想的な形態で自己組織化されるのは困難になる