

不平等な調整ゲームにおける動的シグナリングの進化

Coordination by Evolved Dynamic Signaling in a Game with Unequal Equilibrium Payoffs

高野雅典 有田隆也
Masanori Takano Takaya Arita

名古屋大学大学院情報科学研究科
Graduate School of Information Science, Nagoya University

Communication is necessary for coordination among agents. We construct an agent based model where agents transmit continuous signals each other, and conduct evolutionary simulations in a game with unequal equilibrium payoffs. The dynamics of each other's signaling are analyzed.

1. はじめに

複数の均衡を持つ調整ゲーム (Coordination Game) では、プレイヤーが全員でどの均衡を選択するかが焦点となる。プレイヤーが協力してある均衡を選択するためには、互いの選択に関する情報を共有する必要がある。情報共有のための一つの方法は意思決定の前にコミュニケーションすること (Pre-play Communication) である。進化ゲーム理論において、複数の均衡から最もゲームに参加する全プレイヤーの利得の高くなるものを選択するための Pre-play Communication に関して議論されてきた (例えば [Robson 90, Warneryd 91, Matsui 91, Kim 95])。

また、進化や学習などの適応機構を持つプレイヤーの集団においてコミュニケーションが成立するためには、情報を示すために発信されるシグナルとその発信者の意図・状態の対応関係を獲得することが必要である [Steels 03]。そのような適応機構を持ったエージェントまたはロボットを用いた研究において、明示的な意味を持たないシグナルの送受信から、シグナルの意味の獲得と協力タスクの遂行としての動的な行動の調整が研究されてきた [柴田 99, Miller 04, Marocco 06, Lorincz 06]。それらでは関わるエージェントにとって調整された行動の組が全員に平等であるという特別な状況を扱っている。

しかし、交渉という状況を考えると、例えば男女の争いゲームのような、両者の最良の行動の組は異なるが、両者の行動の選択を調整しなければ両者にとって不利益という状況のように、関わるプレイヤーにとって調整された行動の組に不平等さがある場合も一般的に考えられる。このとき、プレイヤーの利害は完全には一致していないので、シグナルと発信者の意図・状態の対応関係の獲得はより難しくなると予想される。

表 1: 対称化した「男女の争い」ゲーム

	A	B
A	0, 0	1, α
B	α , 1	0, 0

本稿では不平等な調整状況の最も単純な場合として男女の争いゲーム (表 1) を考える。表 1 ではプレイヤーの戦略と利得の組が対称になるように示している。 $0.0 < \alpha < 1.0$ の場合、このゲームでは両者が異なった戦略を選択しなければ共に利得を得ら

れないが、その場合でも戦略 B を選択すると相手より利得が低くなる。したがって、自分が戦略 A を選択し、相手が B を選択する場合は最適である。この不平等さは α が大きいほど弱くなり、1 になるとなくなるので、 α はゲームの平等度を表すといえる。以降ではゲームの結果が両プレイヤーが利得が得られる戦略の組 (AB または BA) になることをゲームが成功するという。

まず、コミュニケーションをしない場合について述べる。男女の争いゲームにおける進化的に安定な戦略 (ESS) [Smith 73] は戦略 s を選択する確率 $P(s)$ が $(P(A), P(B)) = (1/(\alpha+1), \alpha/(\alpha+1))$ の戦略である [石原 02]。その ESS 同士の対戦の平均利得は $\alpha/(1+\alpha)$ であり、ゲームの成功率は $2\alpha/(1+\alpha)^2$ である。ゲームの前にコミュニケーションをしない場合では $\alpha = 1.0$ (不平等でない場合) でもゲームの成功率は 0.5 である。これはゲームに参加している 2 者が共に相手がどちらの戦略を選択するかわからないためであり、事前にコミュニケーションを行うことでゲームの成功率は高くなると考えられる [岡田 97]。ただし、前述したように $\alpha < 1.0$ のときゲームが成功する戦略の組は不平等な利得をプレイヤーに与える。そのとき、ゲームに参加するプレイヤーにとって都合のいい結果 (自分が A, 相手が B) は互いに異なり、交渉によって自分に都合のいい結果を導くことが重要になる。

本研究では、このような不平等な調整状況における動的なコミュニケーションの進化的獲得とそのコミュニケーションによる交渉のダイナミクスに焦点を当てる [高野 07]。行動の前に相互にシグナルを送受信し、その後同時に行動をするというモデルを設計し、男女の争いゲームを用いた進化シミュレーションを行った。そして、獲得された調整行動はどのようにしてゲームが成功するような行動の組を形成していくか、そのダイナミクスについて解析した。

2. モデル

一回の対戦でエージェントは 1 対 1 で 1 回だけゲームを行う。その対戦では、まずシグナルの送受信を繰り返し行った後、戦略を決定しゲームをする。ゲームの前にシグナルの送受信を行う期間を以下では交渉期間という。集団内でゲームの総当たり対戦を行い、その対戦の平均得点をそのエージェントの適応度とした。そしてそれに基づき進化させた。

エージェントは状態として、直接戦略を表す $a \in \mathcal{R}$ と戦略とは直接関係無いが相手に観測されるシグナル $\phi \in \mathcal{R}$ 、内部状態 $I \in [-0.5, 0.5]$ の 3 つを持つ。 a は交渉期間の後、その値がそのエージェントの戦略 s となる。本稿では 2×2 ゲームを用いるので、 $a \leq 0$ の場合は $s = A$ 、 $a > 0$ の場合には $s = B$ を選択するとした (表 1)。エージェントの制御にはニューラルネットワー

連絡先: 高野雅典, 名古屋大学大学院情報科学研究科, 名古屋市千種区不老町, takano@alife.cs.is.nagoya-u.ac.jp

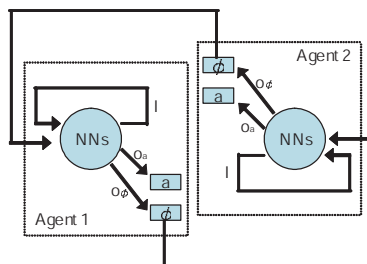


図 1: エージェントの構成

クを用いた。

ニューラルネットワークの入力は相手のシグナル ϕ_{other} と自分の内部状態 I_{self} , 出力は o_a, o_ϕ, o_I である (出力の範囲はすべて $[-0.5, 0.5]$)。 I_{self} は前の時刻のニューラルネットワークの出力のうちの一つ o_I である。エージェントの入出力の関係を図 1 に示す。これら出力によりエージェントは自分の状態 a, ϕ, I を次のように変更する。

$$a(t+1) = a(t) + o_a d \quad (1)$$

$$\phi(t+1) = \phi(t) + o_\phi d \quad (2)$$

$$I(t+1) = o_I \quad (3)$$

t の範囲 (交渉期間) は T ステップとし, d は $d \ll 1$ の定数とする。上記の状態の更新を $t = 1$ から T まで繰り返し, $a(T)$ の値を前述のように戦略 s として評価しゲームを行う。

3. シミュレーション

3.1 設定

ニューラルネットワークの重みを実数値の遺伝的アルゴリズムを用いて進化させた。エージェントの初期状態 ($t = 0$) では $a(0), \phi(0), I(0)$ を 0.0 とし, 交渉期間 T は 200, d は 0.05 とした。ニューラルネットワークは 3 層であり, 入力ニューロンは 2 つ (ϕ_{other}, I_{self}), 中間ニューロンは 4 つ, 出力ニューロンは 3 つ (o_a, o_ϕ, o_I) とした。各層にバイアスニューロンがあるので重みの数は 31 となる。遺伝的操作として, 突然変異は変異率 0.05 で平均 0.0, 標準偏差 0.5 の正規分布に従って与え, 交叉は交叉率 0.8 の二点交叉, 選択はルーレット選択を用いた。遺伝子長はニューラルネットワークの重みの数になる。以上の方法で個体数 100 で 8000 世代実行した。利得行列は表 1 の α ($[0.0, 1.0]$) を 0.1 刻みで変え進化を行った。

3.2 シグナリングの効果の基本的傾向

進化の結果得られた集団において, どのような相互作用をしているかについて示す。各 α について進化を 16 試行を行った。特に断らなければ, 以降の節で示している値は 16 試行の平均である。

まず, 進化の結果のおおまかな傾向として, 各 α についてゲームの成功率と戦略に関する集団の分布について示す。以下で示している値は世代 7000 から 8000 の平均値である。

図 2 に本モデルの進化によって得られた集団 (以下コミュニケーション集団) と ESS で構成される集団 (以下 ESS 集団) の α ごとのゲームが成功した割合を示す。コミュニケーション集団は $\alpha = 1.0$ でほぼ割合 1 でゲームが成功し, α が小さいほどゲームが成功した割合は小さくなった。ここで, ESS 集団の成功率と比較すると, $\alpha \geq 0.5$ においてコミュニケーション集団は ESS

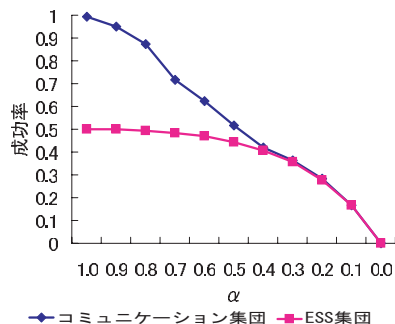


図 2: ゲームが成功した割合

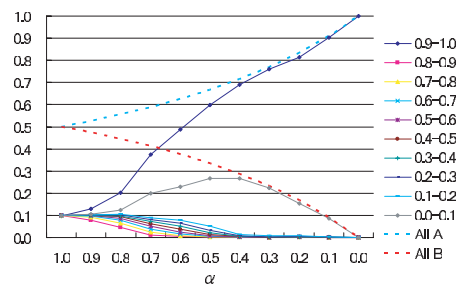


図 3: 戦略 A の選択割合別のエージェントの頻度

集団を上回り, $\alpha < 0.5$ では α が小さくなるほどコミュニケーション集団は ESS 集団に漸近していることがわかる。

図 3 に, 各 α において集団で総当たり対戦をした場合に戦略 A を選択した割合をエージェントごとに求め, その頻度を求めたものを示す。加えて, 純粋戦略の平衡点の集団 (以下純粋戦略集団) の A と B の割合も示す。純粋戦略の平衡点は A の割合 $= 1/(\alpha + 1)$, B の割合 $= \alpha/(\alpha + 1)$ である。コミュニケーション集団は $\alpha = 1.0$ で頻度は一様になった。 $0.5 \leq \alpha \leq 1.0$ では, α が小さいほどほとんど一方の戦略しか選択しないエージェントの頻度 (0.0 - 0.1, 0.9 - 1.0) が多かった。さらに α が小さくなり, $0.0 \leq \alpha < 0.5$ ではほとんど A を選択するエージェントの頻度 (0.9 - 1.0) が多かった。ここで, 純粋戦略集団の各戦略の割合と比較すると, $\alpha < 0.5$ では α が小さくなるほど 0.0 - 0.1, 0.9 - 1.0 は純粋戦略集団の各戦略の割合に漸近した。

以上より, 進化によって α が大きいとコミュニケーションによる調整が獲得されるが, 0.5 より小さくなるとコミュニケーションが無効化し, 交渉しない 2 つの純粋戦略の平衡点の状態と同様の集団に漸近していると考えられる。

典型的な進化の過程を図 4 示す。図で示しているのは $\alpha = 0.9, 0.5, 0.1$ における, コミュニケーション集団の各世代のゲームが成功した割合と, 対応する ESS 集団の成功率である。 $\alpha = 0.9$ の成功率は, 早い世代から ESS 集団を大きく上回った。コミュニケーションが成立している状態が維持されていると考えられる。 $\alpha = 0.5$ の成功率は, ESS 集団と同程度の場合と上回っている場合の 2 つの場合を行き来するような不安定な挙動を示した。これはコミュニケーションが成立している状態も成立していない状態も, 安定して維持されていないと考えられる。 $\alpha = 0.1$ の成功率は, 早い世代から成功率が ESS 集団と同程度になった。コミュニケーションが成立していない状態が維持されたと考えられる。前述したように $\alpha = 0.5$ を境に進化の結果は大きく異なった

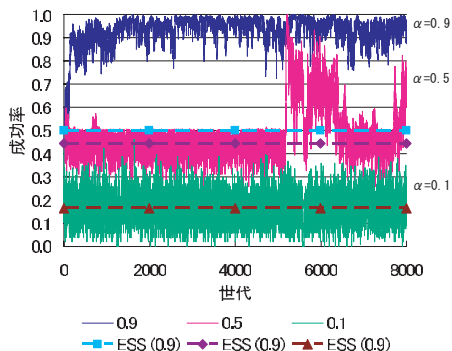


図 4: 進化における成功率の変化 ($\alpha = 0.9, 0.5, 0.1$) と対応する ESS の成功率

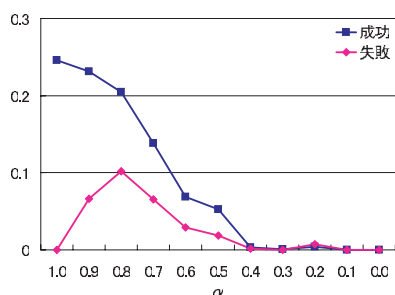


図 5: $t = T$ における ϕ と s の相互情報量

ものになったが、その $\alpha = 0.5$ 付近の進化的な挙動は不安定になった。

3.3 シグナルに関する相互情報量の分析

前節で示されたようなコミュニケーションによる調整が、どのように行われているか、相互情報量を用いてシグナルの持つ意味について調べる。

まず、交渉期間後 ($t = T$) の $\phi(T)$ と選択した戦略 s の相互情報量について調べた。対象とした集団は各進化の試行によって得られた最終世代である。相互情報量は各個体の全対戦の $\phi(T)$ と s に関して求め、それを全個体で平均した。 $\phi(T)$ は -6.0 から 6.0 の範囲で幅 0.12 に区切って離散化し求めた。図 5 に各 α について相互情報量を示す。 $\phi(T)$ による情報とゲームの成功率の関係性を調べるために、同図ではゲームが成功した場合 (AB または BA) と失敗した場合 (AA または BB) で分けて示している。成功した場合には α が小さくなると相互情報量が減少した。特に、 $\alpha < 0.5$ では成功・失敗にかかわらず相互情報量はほぼ 0 になった。これは、 α が小さくなると純粋戦略のようなエージェントが集団内に多くなるためであると思われる。 $\alpha \geq 0.5$ では、ゲームが成功した場合は失敗した場合に比べ相互情報量が高くなった。以上より、 $\alpha \geq 0.5$ ではゲームが成功する場合、最終的なシグナルの値 $\phi(T)$ にそのエージェントの選択した s に関する情報が示されているといえる。

つぎに、エージェントの $\phi(t)$ と選択した戦略 s の相互情報量が交渉期間の間どのように変化し、前述のように $\phi(T)$ が戦略 s の情報を示すようになるかを調べた。図 6 に相互情報量の変化を示す。相互情報量がある点 t で増加しているのがわかる。ここから、エージェントには相手に自分の戦略の情報を示し始める t

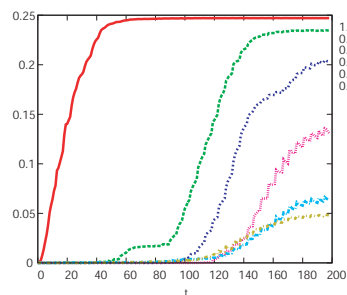


図 6: 相互情報量の時間変化

に関する閾値 t' があると考えられる。そして、 t' は α が小さいほど遅くなった。

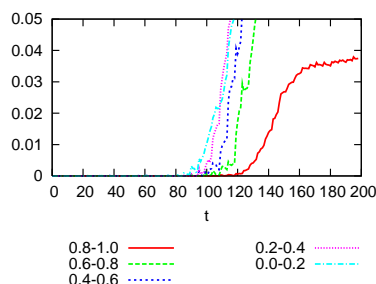
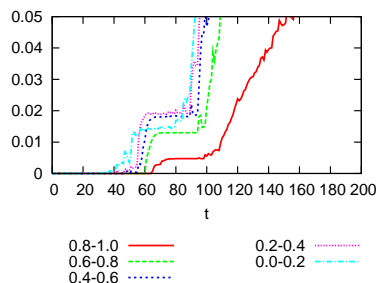


図 7: 戦略 A を選択した割合別の相互情報量の時間変化 ($\alpha = 0.9$ (上), $\alpha = 0.8$ (下))

図 7 に交渉期間中に相互情報量が 0 から増加したときの变化を示す。ここでは典型的な例として $\alpha = 0.9, 0.8$ に関してのみ示している。ただし、同図は戦略 A を選択した割合 ($0.0 - 0.2$, $0.2 - 0.4$, $0.4 - 0.6$, $0.6 - 0.8$, $0.8 - 1.0$) ごとに平均して示している。 A を選択する割合の高いエージェントほど閾値 t' が大きくなり、集団内に t' に関して多様性があることがわかる。 α が小さいと A を選択する割合の分布は偏る (図 3) ためこの多様性は減少する。このことから、自分の戦略に関する情報を相手に示すことは、自分が B を選択することを相手に示すことであると考えられる。

$\alpha = 1.0$ では、ゲームが成功した場合のエージェントの利得は等しく、選択された戦略が異なればよい。閾値 t' が集団中で多様性を持っていればよいと考えられる。しかし、 $\alpha < 1.0$ で

は、ゲームが成功した場合のエージェントの利得は異なるため、閾値 t' による区別に α による利得差が関わってくる。そのため、 t' は大きくなる方向へ進化したと考えられる。しかし、 t' が大きいエージェント同士では、戦略の分担が間に合わずゲームは失敗しやすい。そのため、閾値 t' のそれほど大きくないエージェントも集団中に一定割合で存在して進化したと考えられる。その結果、 α が小さくなるほど、純粋戦略が増加し、相手次第という側面はなくなるため、 $\phi(t)$ の意味はなくなったと思われる。

3.4 行動の調整方法の分析

以下では、どのように2体のエージェントが ϕ, a を変化させ、戦略の選択を調整しているかを、具体的な場合を用いて説明する。図8に α が0.9におけるゲームが成功したエージェントの組の $\phi(t)$ と $a(t)$ の典型的な軌道を示す。初期状態は $(\phi(0), a(0)) = (0.0, 0.0)$ のため、原点から軌道は始まる。同図の $\phi_{i1}(t)$ と $a_1(t)$ はエージェント1, $\phi_{i2}(t)$ と $a_2(t)$ はエージェント2, $\phi_{i3}(t)$ と $a_3(t)$ はエージェント3の状態をあらわす。同図では2つ軌道の $a(t)$ の符合が $t = T$ で異なることがゲームが成功したことをあらわす。

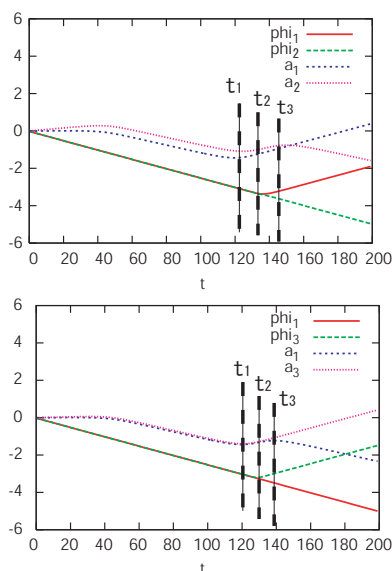


図8: ゲームが成功した場合の2体のエージェントの $\phi(t)$ と $a(t)$ の典型的な軌道 ($\alpha = 0.9$)

両図共、以下に示すことを行っていると考えられる。最初の段階でエージェントは $a(t)$ は A を選択する方向へ動かす。しかし、両者が A を選択するとゲームが失敗するため、図中の t_1 において $a(t)$ の速度を B 方向へ変えはじめる。 t_2 において、一方のエージェントがその変更を $\phi(t)$ の速度を逆方向へ変えることで示す。 t_3 において、もう一方のエージェントがそれに反応し、 t_1 で変えた $a(t)$ の速度を A 方向へ変える。ここで、 t_2 において $\phi(t)$ の速度を変えたのは、上の図ではエージェント1であるが、下の図ではエージェント3である。つまり、エージェントの持つ閾値が異なることによって相手によって A を選択する場合、 B を選択する場合が変わり、 $\phi(t)$ による調整がなされている。

4. まとめ

本論文では、動的シグナリングの中に意味を進化的に獲得する現象における不平等性の役割を検討するために、同時に行動する前にシグナルを一定期間相互に送るモデルを設計し、男女の争

いゲームを用いて進化シミュレーションを行った。このゲームでは、両者が同じ戦略をすると両者の利得は0になるため、一方だけが戦略を変更する必要があるが、同時に変更してしまうことを避けるために、自分が戦略を変更したことを相手に伝えることで利得が増加する。実験の結果、自分の選択しようとしている戦略を変更したことを相手に伝える役割をシグナルが持つように進化したことが示された。利得の差が小さい、つまり相互作用が平等に近いならば、シグナルを送るタイミングの多様性のために、このコミュニケーションシステムは機能する。しかし、獲得する利得の差が開いて平等度が小さくなっていくと、タイミングを遅らせる選択圧が強まり、同時に、戦略の変更が間に合わなくなるリスクが増加する。そして、コミュニケーションは崩壊し、相手によらない純粋戦略の共存という分極化したすみ分け社会に等しくなってしまう。

参考文献

- [Kim 95] Kim, Y.-G. and Sobel, J.: An Evolutionary Approach to Pre-play Communication, *Econometrica*, Vol. 63, No. 5, pp. 1181–1193 (1995).
- [Lorincz 06] Lorincz, A., Gyenes, V., Kiszlinger, M., and Szita, I.: Mind model seems necessary for the emergence of communication, *arXiv:q-bio/0611012* (2006).
- [Marocco 06] Marocco, D. and Nolfi, S.: Self-Organization of Communication in Evolving Robots, in Rocha, L. M. ed., *Artificial Life X*, pp. 178–184, MIT Press (2006).
- [Matsui 91] Matsui, A.: Cheap-talk and cooperation in a society, *Journal of Economic Theory*, Vol. 54, No. 2, pp. 245–258 (1991).
- [Miller 04] Miller, J. H. and Moser, S.: Communication and coordination, *Complexity*, Vol. 9, No. 5, pp. 31–40 (2004).
- [Robson 90] Robson, A. J.: Efficiency in evolutionary games: Darwin, nash and the secret handshake, *Journal of Theoretical Biology*, Vol. 144, No. 3, pp. 379–396 (1990).
- [Smith 73] Smith, M. J. and Price, G. R.: The Logic of Animal Conflict, *Nature*, Vol. 246, No. 5427, pp. 15–18 (1973).
- [Steels 03] Steels, L.: Evolving grounded communication for robots, *Trends in Cognitive Sciences*, Vol. 7, No. 7, pp. 308–312 (2003).
- [Warneryd 91] Warneryd, K.: Evolutionary stability in unanimity games with cheap talk, *Economics Letters*, Vol. 36, No. 4, pp. 375–378 (1991).
- [岡田 97] 岡田 章: ゲーム理論, 有斐閣 (1997).
- [高野 07] 高野 雅典, 有田 隆也: 不平等ゲームにおいて進化した動的なシグナリングによる行動の調整, 人工知能学会研究会資料, SIG-KBS-A604, pp. 77–82 (2007).
- [柴田 99] 柴田 克成, 伊藤 宏司: 利害の衝突回避のための交渉コミュニケーションの学習と個性の発現-リカレントニューラルネットを用いたダイナミックコミュニケーションの学習-, 計測自動制御学会論文誌, Vol. 35, No. 11, pp. 1346–1354 (1999).
- [石原 02] 石原 英樹, 金井 雅之: 進化的意思決定, 朝倉書店 (2002).