ORにおける人工知能システムの有効性について

林田　智弘，西﨑　一郎

複数の意思決定者の利得が，それぞれの意思決定や行動の組み合わせにより決まる状況をゲーム的状況とよび，個人合理性に基づく均衡理論や，全体合理性に着目した解析的な研究が行われている．また，確率的行動選択モデルや学習理論に基づく行動選択モデルなどが提案されている．これらの妥当性を検討するための被験者実験が数多く報告されており，その多くはこれらの行動選択モデルを支持する結果を得ているが，一部のゲームモデルにおいては単一の行動選択モデルだけでは説明できないような結果が得られている．本稿では，このようなゲーム的状況に対する，人工適応型エージェントを用いたシミュレーション分析の有効性について議論し，OR 分野への適用例として展開型ゲームにおける行動分析に関する研究を紹介する．

キーワード：人工適応型エージェント，シミュレーション分析，展開型ゲーム

1. はじめに

意思決定者が複数存在する状況で，それぞれの利得が自分と他の意思決定者の行動に依存して決定するような状況をゲーム的状況という．ゲーム理論では，意思決定者をプレイヤー，それぞれの意思決定を戦略と呼ぶ．

ゲーム的状況はプレイヤーの戦略と利得構造によりさまざまに分類される．本稿では，各プレイヤーがそれぞれ戦略を決定し，すべてのプレイヤーの戦略の組み合わせにより利得が決定される非協力ゲームについて述べる．$N=\{1,2,\ldots,n\}$ をプレイヤー集合としたとき，n 人で行われる非協力ゲームに対して，プレイヤー $i\in N$ の純戦略集合を $S_i=\{s_{i1},s_{i2},\ldots,s_{ii_m}\}$，$S=S_1\times S_2\times\cdots\times S_n$ としたとき，すべてのプレイヤーの戦略の組 $\boldsymbol{s}=(s_1,s_2,\ldots,s_n)\in S\ (s_i\in S_i)$ を戦略プロファイルという．プレイヤー i の戦略 s_i に対して，i 以外のすべてのプレイヤーの戦略の組を $\boldsymbol{s}_{-i}$ とする．$\pi_i:S\to\mathbb{R}$ をプレイヤー i の利得関数としたとき，式 (1) を満たす戦略プロファイル $\boldsymbol{s}^*=(s_i^*,\boldsymbol{s}^*_{-i})$ をナッシュ均衡という．

$$\pi_i(\boldsymbol{s}^*)\geq\pi_i(s_i,\boldsymbol{s}^*_{-i}),\forall s_i\in S_i,\forall i\in N \qquad (1)$$

式 (1) より，戦略プロファイル $\boldsymbol{s}^*$ がナッシュ均衡である場合，すべてのプレイヤーの戦略は他のプレイヤーの戦略の組に対する最適応答となっている．ナッシュ均衡は，非協力ゲームに対する強力な解概念であることが知られており，多くのゲームモデルにおける人間の行動を説明することに成功している．しかし一方で，ナッシュ均衡だけでは説明できない結果を得ている被験者実験も多く報告されている [3, 4, 7, 8, 9, 17, 20]．ナッシュ均衡は，非協力ゲームにおいてプレイヤーがいかに行動すべきかを示す規範的な解概念といえ，プレイヤーが厳密に利得最大化のための合理的な意思決定をするという仮定のもと導出されている．しかし，上述のような被験者実験の結果が得られていることから，人間は必ずしもこのような意思決定を行うとは限らない．すなわち，1 回のゲームで得られる利得だけではなく，他のプレイヤーの利得やこれまでの累積利得，過去の経験など，ゲームの結果に対する複数の基準に基づく満足度（効用）により行動を決定していると考えられる．

ゲーム的状況における人間の意思決定については，ゲームの利得を含めた複数の基準で構成される効用関数や，人間の試行錯誤的な意思決定構造などを考慮した分析が行われている．本稿では特に，人工知能システムの一種である人工適応型エージェントを用いたシミュレーションにより，ゲーム的状況に対する被験者実験での被験者の行動分析を行うことに着目する．

本稿は，非協力ゲームについて述べたあと，人工知能システムの設計に必要な機械学習について簡単に説明する．最後に，非協力ゲームの一種である展開型のゲームに関する研究を紹介する．

はやしだ　ともひろ，にしざき　いちろう
広島大学大学院工学研究院
〒 739–8527 広島県東広島市鏡山 1–4–1

2. 非協力ゲームと人工知能システム

2.1 均衡概念と人間の行動

すべてのプレイヤーが同じ戦略をとるときに高い利得を得ることができる非協力ゲームを協調ゲームという．2 人プレイヤーの協調ゲームの利得表の例を表 1 に示す．表 1 では，プレイヤー 1 の戦略が列，プレイヤー 2 の戦略が行で示されており，各戦略プロファイルに対する利得の組 (a, b) は，プレイヤー 1 の利得が a，プレイヤー 2 の利得が b であることを表している．

表 1 2 人非協力ゲームの例（協調ゲーム）

		プレイヤー 1	
		$s_{11} = A$	$s_{12} = B$
プレイヤー 2	$s_{21} = A$	$(10, 10)$	$(3, 1)$
	$s_{22} = B$	$(1, 3)$	$(7, 7)$

式 (1) より，表 1 に示される協調ゲームでは，2 種類の戦略プロファイル $(A, A), (B, B)$ がナッシュ均衡となるが，どちらの均衡が実現するかを予測することは難しい．このような問題は均衡選択問題と呼ばれる．この非協力ゲームに対する均衡は，リスク優越均衡と利得優越均衡の 2 種類に分類される．表 1 に示される協調ゲームでは，(B, B) がリスク優越均衡，(A, A) が利得優越均衡となる．プレイヤーが意思決定において小さな確率でエラーを起こすことを想定した繰り返しゲームを考え，確率的安定とよばれる解概念を用いた場合，協調ゲームではリスク優越均衡に収束することが示され [15, 22]，これは多くの均衡選択問題に対する有効な解概念であることが知られている．さらに，非協力ゲームにおける解概念として，ナッシュ均衡や確率的安定などの他に，プレイヤー間の暗黙の合意形成を考慮した解概念である関連均衡 [1, 2]，展開型のゲームにおけるサブゲーム完全均衡などが提案されている．しかし，均衡理論だけでは説明できないような被験者の行動も報告されている [3, 20].

人間は自らの行動の結果をいくつかの基準により複合的に評価し，試行錯誤的に意思決定や行動戦略を学習していると考えられる．被験者のこのような意思決定構造が，ゲームの利得に基づいた解概念だけでは被験者の行動を十分に説明できない原因であると考えられる．試行錯誤的な意思決定構造を考慮した行動分析を行うために，上述のような人間の意思決定構造を模倣することのできるエージェントを用いた，マルチエージェントベースシミュレーションについて考える．このような適応能力を有するエージェントを一般に人工適応型エージェントと呼ぶが，本稿では簡単に「エージェント」と呼ぶ．

2.2 エージェント

コンピュータ上に設計され，自律的に意思決定，行動および学習を行うような機能単位はエージェントと呼ばれ，人間の思考パターンや意思決定，行動規則を模倣できる人工知能システムを用いて構築される．コンピュータ上に複数のエージェントが存在する人工社会では，エージェント間の複雑な相互作用関係により，多種多様のグローバルな性質や秩序が創発される．各エージェントは周囲の環境や他のエージェントの挙動などを知覚し，与えられた目的を達成するために自らの行動ルールや戦略を更新していく．

人間は周囲の環境情報などに基づいて意思決定すると考えられ，エージェントは If-then 形式の意思決定ルールを保持することでこのような意思決定構造を模倣する．すなわち，エージェントは周囲の情報を知覚してこれをシステムへの入力とし，対応するシステムからの出力に基づいて行動を決定する．エージェントは，意思決定ルールを特徴づけるいくつかのパラメータの適切な設定により適切に学習を行い，試行錯誤的に意思決定するエージェントを考えたとき，機械学習の枠組みを用いることで適切な学習が可能となる．エージェントの学習方法は，学習目標となる教師信号が事前に与えられる「教師あり学習」と，教師信号が与えられない「教師なし学習」の 2 種類に分類される．強化学習，クラシファイアシステム，遺伝的アルゴリズム，ニューラルネットワークなどの機械学習の枠組みを適用することで，エージェントの意思決定および学習機構を構築できる．

2.3 シミュレーション分析

ニューラルネットワークを用いたエージェントの意思決定について述べる．

多入力 1 出力の関係を持つ，ニューロンと呼ばれるユニットを複数結合したものをニューラルネット

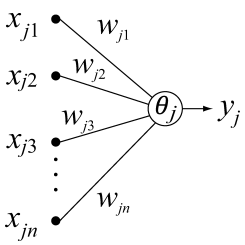

図 1 ニューロン j

ワークと呼び，人間の脳内での情報処理の仕組みを単純化した工学モデルである．ニューロン j への入力を $x_{j1}, x_{j2}, \ldots, x_{jn}$ としたとき，出力は $y_j = f(\sum_i x_{ji} w_{ji} - \theta_j)$ で与えられる (図 1) .

このとき，w_{ij} をニューロン i, j 間の結合重み，θ_j をニューロン j の閾値，f はニューロンの伝達関数という．入力層，中間層，出力層の 3 層から構成されるフィードフォワード型ニューラルネットワークの例を図 2 に示す.

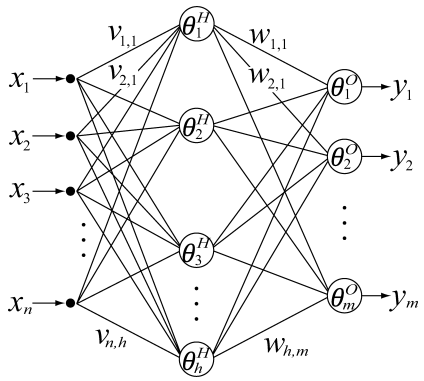

図 2　フィードフォワード型ニューラルネットワーク

伝達関数をシグモイド関数 $f(z) = \frac{1}{1+\exp(-\gamma z)}$ とし，中間層のノード数とシグモイド関数のゲイン γ および，重み $\boldsymbol{v}, \boldsymbol{w}$，閾値 $\boldsymbol{\theta}$ の値を適切に設定すれば，フィードフォワード型ニューラルネットワークは，n 入力 m 出力の任意の関数に対する高い近似能力を持つことが知られている．教師あり学習であれば，誤差逆伝播法によりニューラルネットワークの入出力関係と教師信号の誤差を最小化することで $\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{\theta}$ の適切な値を学習することが可能である [19].

エージェントの得る外部情報をニューラルネットワークに入力情報として与え，それに対応した出力に基づいて意思決定する．ニューラルネットワークは複数の入力情報を一括して処理することができ，人間の意思決定構造をエージェントに模倣させることができる.

表 1 に示されるような同時手番の協調ゲームのように，複数の均衡が存在するゲーム的状況を考えると，各プレイヤーが利得を最大化するような戦略を教師信号として個々に与えることは難しい．このため，ゲーム的状況において，エージェントはゲームの結果得られた利得などの情報に基づいて行動を事後評価する，教師なし学習を行うことが適切である．ニューラルネットワークの教師なし学習の方法として，図 3 に示されるような重みや閾値により構成される遺伝子を考え，これを遺伝的アルゴリズムにより学習する手法が提案されている [16].

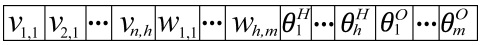

図 3　結合重みと閾値で構成される遺伝子

われわれはこれまでに，上記のようなニューラルネットワークと遺伝的アルゴリズムを組み合わせた意思決定および学習機構を持つエージェントを用いて，ネットワーク形成に関する被験者実験における被験者の行動分析を行っている．さらに，それ以外の機械学習の枠組みを用いたシミュレーションモデルも構築している．たとえば，寡占的市場における販売者と消費者の 2 種類のエージェントを考え，クラシファイアシステムの一種である XCS (eXtended Classifier System) [21] を用いて，販売者エージェントに市場戦略を学習させる人工市場モデルを提案している．他の行動分析を含めた詳細については，文献 [12, 13, 14, 16] を参照していただければ幸いである.

次節では，逐次手番の非協力ゲームの一種である最終提案ゲームに対して，マルチエージェントシミュレーションモデルを適用した研究について概説する.

3. 展開型ゲームにおける行動分析

複数のプレイヤーがあらかじめ決められた順番に従って意思決定する逐次手番ゲームを，ゲームの木と呼ばれるグラフの形式で表したものを展開型ゲームといい，サブゲーム完全均衡と呼ばれる均衡解により多くの被験者実験の結果が説明できることが報告されている．しかし，展開型ゲームの一種である最終提案ゲームに関する被験者実験 [10, 18] により，一部の展開型ゲームにおいては，被験者の行動は必ずしもサブゲーム完全均衡とは一致しないことが示されている.

人工知能システムのひとつである人工適応型エージェントを用いたシミュレーション分析によって，n 人プレイヤーの展開型ゲームにおける行動分析を行うことが可能である．ここでは，適用例として，最終提案ゲームに対する被験者の行動分析について簡単に述べる．最終提案ゲームは，2 人のプレイヤーが一定額の金銭を分割するゲームであり，被験者実験の結果はサブゲーム完全均衡とは一致しないことが報告されている．この理由として，被験者は均衡理論で仮定されているような利得最大化のための合理的な意思決定を行うので

はないことが考えられる．この結果に対して，Duffy and Feltovich [6] は，情報量がゲームに与える影響に関して強化学習のモデルにより分析した．また，Fehr and Schmidt [5] は，人間の公平性の概念を取り入れた効用関数を提案し，Gomes and Zauner [10] は，2人の得る利得に関する不公平感を考慮した解析的な分析を行った．

本稿では，利得の公平性と学習とを考慮したシミュレーションモデルについて述べる．エージェントはニューラルネットワークに基づいた意思決定を行うが，ニューラルネットワークに関するパラメータの初期値は乱数が与えられるため，最初に誤差逆伝播法により利得などのゲーム構造をニューラルネットワークに事前に学習させる．次に，エージェント間でゲームを行い，その結果得られた利得を用いた遺伝的アルゴリズムによりニューラルネットワークを進化させる．

3.1 最終提案ゲーム

最終提案ゲームでは，2人のプレイヤーをそれぞれ先手，後手と呼ぶ．P ドルを分割する最終提案ゲームでは，まず先手が $\pi \in \{0, 1, \ldots, P-1\}$ を提案し，後手はこれを受諾するか拒否するかを決める．後手が受諾した場合，先手と後手の利得は $(\pi, P-\pi)$ となり，拒否した場合，2人のプレイヤーの利得は0となる．最終提案ゲームのサブゲーム完全均衡は，先手が $(P-1)$ ドルを要求し，後手はその要求を受け入れ1ドル得ることである．$P = 10$ とした被験者実験 [10] の結果，平均提案額は6.04ドルであった．横軸を先手の提案額，縦軸を提案者の割合と受諾比率としたグラフを図4に示す．

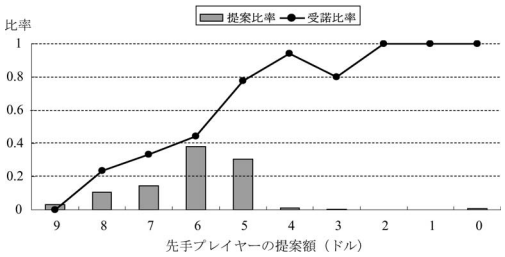

図 4 被験者実験の結果 [10]

3.2 シミュレーションモデル

エージェントの初期個体群として，先手と後手それぞれ N 体ずつをランダムに生成する．ここで，エージェントは図2に示されるような3層フィードフォワード型ニューラルネットワークを用いて意思決定を行う．最初に各エージェントは誤差逆伝播法によりゲームの構造を事前学習し，各期ごとにランダムに決定される N 個のペアでゲームを行う．ここで，各ゲームにおけるエージェントの意思決定はニューラルネットワークの出力に基づくものとし，遺伝的アルゴリズムを用いて学習を行い次世代の個体群を生成し，これを規定回数繰り返す．図5にシミュレーションモデルの概略を示す．

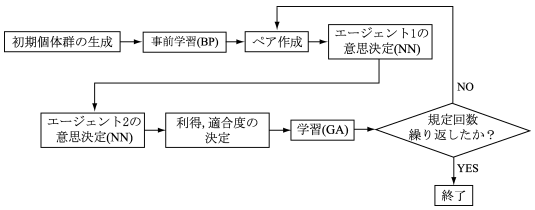

図 5 シミュレーションモデル

先手，後手エージェントの直前期における戦略と利得，直前期までの平均利得と平均提案額，受諾比率を先手エージェントのニューラルネットワークの入力値として与える．出力層のノード数は10とし，各ノードからの出力値を各戦略 $0, 1, \ldots, 9$ に対する優先度と解釈し，先手エージェントは最も優先度の大きい戦略を採用する．後手エージェントには，先手エージェントに与える情報に加えてゲーム相手となる先手エージェントの提案額をニューラルネットワークの入力値として与える．後手エージェントのニューラルネットワークの出力層のノード数を2とすると，それぞれ受諾と拒否に対する優先度と解釈され，出力値の大きいほうの戦略を後手エージェントの戦略とする．

プレイヤーは不公平な利得分割に対して不効用を得るものと考え，プレイヤー i の効用関数は，ゲームにより得られる利得と，自分とゲーム相手の得る利得の不公平性から生じるペナルティにより構成され，i がゲームにより得る利得を π_i，相手プレイヤーの利得を π_j としたとき，

$$f_i(\pi_i, \pi_j) = \pi_i - \alpha_i \max\{\pi_j - \pi_i, 0\} - \beta_i \max\{\pi_i - \pi_j, 0\} \tag{2}$$

により与えられるものとする [5]．第2項は相手プレイヤーの利得のほうが大きい場合，第3項は自分の利得のほうが大きい場合のペナルティを表している．不公平な利得分割は2種類に分類できるが，プレイヤーは得られる利得の差分に比例して不効用を得るとすると，α_i, β_i はそれぞれに対する重みである．式 (2) で与えられる効用関数を用いて，パラメータ α_i, β_i を変動さ

せたシミュレーション実験を行い，不公平な利得配分に対する被験者の感じる不効用の大きさの，ゲームへの影響を検証する．なお，各エージェントは遺伝的アルゴリズムに基づいてニューラルネットワークの重みと閾値を更新し，自らの意思決定機構であるニューラルネットワークを進化させる．ただし，遺伝的アルゴリズムに用いられる適合度は，式 (2) で与えられる効用値とする．

3.3 シミュレーション分析

本研究では，被験者実験 [6] の結果を説明可能な不公平性からのペナルティに対する重み α_i, β_i を同定するとともに，各パラメータの変動実験を行うことで，その影響に関する分析を行う．以降，先手をプレイヤー 1，後手をプレイヤー 2 とする．

3.3.1 シミュレーション実験

α_2, β_1 は，それぞれ $\pi_1 > \pi_2$ の場合の，不公平な利得分割に対するペナルティの重みである．これらのパラメータはゲームの結果に対する影響が大きいと考えられるため，被験者実験の結果と一致するような α_2, β_1 の値を同定する．逆に，α_1, β_2 は，それぞれ $\pi_1 < \pi_2$ の場合の，不公平な利得分割に対するペナルティに対する重みであることから，ゲームの構造上プレイヤーの行動に与える影響が小さいと考えられるため，$\alpha_1 = \beta_2 = 0$ とする．

3.3.2 シミュレーション結果

$\alpha_2 = 1.55$, $\beta_1 = 0.5$ としたときのシミュレーション実験の結果を図 6 に示す．

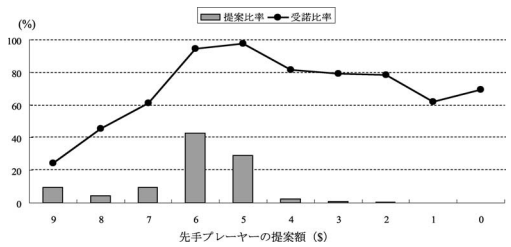

図 6 シミュレーション結果

図 6 より，平均提案額が 5.99 ドルであり，被験者実験の結果と近い値を得た．$\alpha_2 = 1.55$, $\beta_1 = 0.5$ であるので，後手が公平性を強く求める傾向があるといえる．これに比べて先手は後手よりも公平な利得配分を強く意識しているというよりもむしろ，後手に受諾されやすいような提案を行っていると考えられる．

詳しくは，文献 [11] を参照していただければ幸いである．

4. おわりに

本稿では，人工知能システムの一種である，人工適応型エージェントを用いたシミュレーションシステムについて簡単に紹介し，複数の意思決定者（プレイヤー）がそれぞれ独立して意思決定するようなゲーム的状況に対して，この人工知能システムを応用した研究を概説した．

本稿で紹介したシミュレーションモデルは，第 3 節で述べたような展開型のゲーム以外にもさまざまなゲームモデルへ適用可能であり，多くの研究成果が報告されている．これらの研究では，さまざまなゲーム的状況に関する被験者実験において，均衡理論などの数理的な解析手法では説明できないような被験者の行動を人工知能システムを用いて分析している．

近年のコンピュータ技術の急速な発展や，新しいシミュレーション手法の登場により，被験者実験のような限定的な環境における人間の行動分析だけではなく，現実の複雑な社会現象の分析などに対しても人工知能システムの適用範囲が広がることが期待される．

参考文献

[1] R. J. Aumann: Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, **1** (1974) 67–96.

[2] R. J. Aumann: Correlated equilibrium as an expression of bayesian rationality. *Econometrica*, **55** (1987) 1–18.

[3] S. Berninghaus, K.-M. Ehrhart and C. Keser: Coordination games: Recent experimental results. *Working Paper* (1997) 97–129.

[4] S. K. Berninghaus, K. M. Ehrhart, M. Ott and B. Vogt: Evolution of networks—an experimental analysis. *Journal of Evolutionary Economics*, **17** (2007) 317–347.

[5] E. Fehr and K. M. Schmidt: A theory of fairness, competition and cooperation. *Quarterly Journal Economics*, **114** (1999) 817–868.

[6] J. Duffy and N. Feltovich: Does observation of others affect learning in strategic envirnments? An experimental study. *International Journal of Game Theory*, **28** (1999) 131–140.

[7] J. Duffy and E. Hopkins: Learning, information, and sorting in market entry games: theory and evidence. *Games and Economic Behavior*, **51** (2005) 31–62.

[8] I. Erev and A. Rapoport: Coordination, “magic,” and reinforcement learning in a market entry game. *Games and Economic Behavior*, **23** (1998) 146–175.

[9] J. K. Goeree, C. A. Holt and T. R. Palfrey: Quantal response equilibrium and overbidding in private value auctions. *Journal of Economic Theory*, **104** (2002) 247–272.

[10] M. C. Gomes and K. G. Zauner, Ultimatum bargaining behavior in Israel, Japan, Slovenia, and United

States: a social utility analysis. *Games and Economic Behavior*, **34** (2001) 238–269.
[11] T. Hayashida, I. Nishizaki and H. Katagiri: Artificial adaptive agent model characterized by learning and fairness in the ultimatum games. *Journal of Telecommunications and Information Technology*, **2007/4** (2007) 36–44.
[12] 林田智弘, 西崎一郎, 片桐英樹: 社会的評判を考慮したネットワーク形成に関するエージェントベースシミュレーション分析. 日本経営システム学会誌, **25** (2009) 21–32.
[13] T. Hayashida, I. Nishizaki and H. Katagiri: Network formation and social reputation: a theoretical model and simulation analysis. *International Journal of Knowledge Engineering and Soft Data Paradigms*, **2** (2010) 349–377.
[14] 片桐英樹, 西崎一郎, 林田智弘: 寡占的競合市場における企業の商品戦略に関するエージェントベースシミュレーション分析—クラシファイアシステムに基づく人工適応型エージェントモデルの提案—. 日本経営システム学会誌, **27** (2010) 33–41.
[15] M. Kandori, G. J. Mailath and R. Rob: Mutation and long run equilibria in games. *Econometrica*, **61** (1993) 29–56.
[16] I. Nishizaki: A general framework of agent-based simulation for analyzing behavior of players in games. *Journal of Telecommunications and Information Technology*, **2007/4** (2007) 28–35.
[17] A. Rapoport, D. A. Seale and E. Winter: Coordination and learning behavior in large groups with asymmetric players. *Games and Economic Behavior*, **39** (2002) 111–136.
[18] A. Roth, V. Prasnikar, M. Okuno-Fujiwara and S. Zamir: Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study. *American Economic Review*, **81** (1991) 1068–1095.
[19] D. E. Rumelhart, G. E. Hinton and R. J. Williams: Learning representations by back-propagating errors. *Nature*, **323** (1986) 533–536.
[20] J. B. van Huyck, R. C. Battalio and R. O. Beli: Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, **80** (1990) 234–249.
[21] S. Wilson: Classifier fitness based on accuracy. *Evolutionary Computation*, **2** (1995) 149–175.
[22] H. P. Young: *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, Princeton University Press (1993).

