

マルチエージェント実験経済学の方法的基礎

川越 敏司

1. はじめに

経済学やゲーム理論の諸命題を検証するために実験室において被験者に意思決定させてデータを集め分析する実験経済学と、コンピュータ・プログラムとして実装されたエージェントが互いに学習を通じて相互作用する場を研究するマルチエージェント・システム研究との交流が最近盛んに行われています。これをマルチエージェント実験経済学と総称するようです。本稿では、マルチエージェント実験経済学研究の方法論的基礎について実験経済学者の視点で述べてみたいと思います。なお、実験経済学全般については Friedman and Sunder[4]を参照してください。

2. 構成することによる理解

おそらく、マルチエージェント実験経済学にとって、R. Axelrodが行った囚人のジレンマ繰り返しゲームの研究 (Axelrod[1])ほど広い範囲に影響を与えた研究はないと思います。ある意味でこうした研究のスタイルを確立した感があります。ここでいうスタイルとは、(1)動学ゲームにおけるエージェントの戦略を事前の行動プランとしてコンピュータ・プログラムに表現し、(2)そうしたプログラムを学習・進化させることによって、(3)モデルや戦略を具体的に構成することで複雑な経済現象を理解する、というものです。本稿ではこうした研究の上記3つのキーポイントについて、その意義と問題点を実験経済学の立場から検討していきたいと思います。

さて、囚人のジレンマゲームとは表1のような2人戦略形ゲームです。この囚人のジレンマ繰り返しゲームの実験においては、2人のプレイヤーが互いに協力を選ぶという、1回限りのゲームにおいてはナッシュ均衡ではない結果が実現することがこれまでも多くの

表1 囚人のジレンマゲーム

1/2	協力	裏切り
協力	3,3	0,4
裏切り	4,0	1,1

実験で知られています。ところで、この事実の理論的基礎となるフォーク定理を証明するには、実行可能で個人合理性を満たす任意の利得の組が無限繰り返しゲームのナッシュ均衡になることを、具体的な戦略を構成することで示す必要があります。こうした戦略の中で、初めは協力を選び、相手が一度でも裏切りを選ぶと2度と協力を選ばないというトリガー戦略がよく知られています。Axelrodは彼の主催するコンピュータ・プログラム同士の総当り戦で参加者に具体的に戦略を構成させることを通じて、ほかにも集団的安定な戦略が存在することを発見したわけです。その1つが、初めは協力を選び、それ以降は前回相手が選んだ手を選ぶというしっぺ返し (Tit-for-Tat) 戦略です¹。このしっぺ返し戦略を用いることで、本来つねに裏切ることが最善である有限繰り返しゲームにおいても、最後の数回を除いて協力を達成できることが理論的に示される (Kreps et al.[10])など、数々の有益な結果が生み出されています。このように、具体的に構成することによって理解を進めていくことがマルチエージェント実験経済学のひとつの重要な標語になったのはAxelrodの研究以後なのではないかと思います。

Axelrodはその後の著作でさらにKISS (Keep It Simple, Stupid) 原理というものを提唱しています (Axelrod[2])。これは、マルチエージェント実験経済学が取り扱う複雑な現象を探求する場合、あくまでもモデルの前提は単純であるべきだという主張です。それは、モデルから複雑で驚くような結果が生じた場

かわごえ としじ
公立はこだて未来大学 システム情報科学部
〒041-8655 函館市亀田中野町 116-2

¹ 残念ながら、この戦略を含めて囚人のジレンマ繰り返しゲームには進化的安定戦略は存在しないことが Boyd and Lorberbaum[3]などによって示されている。

合には、モデルの前提が単純な方が原因究明がしやすく、前提が複雑なモデルより複雑な現象が創発するプロセスの理解に役立つはずだからです。この点をさらに具体的な事例で考えてみましょう。

マルチエージェント実験経済学では、株式市場や外国為替市場のような市場取引がよく研究されています。市場取引については、企業や消費者の合理的行動から需要・供給関数が導かれ、さらにこの方程式系から市場均衡解の存在が導かれ、最後にそれがパレート最適であることが理論経済学の教科書には解析的に示されています。一般均衡理論と呼ばれるこの精緻な理論は非線形の相互作用を扱った複雑なものです。ところが、市場均衡のパレート最適性については、非常に単純なモデルによっても示すことができるのです。Gode and Sunder[6]は、次のような知性ゼロの取引者(zero intelligence traders)を前提しても市場均衡のパレート最適性が達成可能であることを示したのです。知性ゼロの取引者には売り手と買い手がいて、仮に1種類の財を取引しているものとします。売り手*i*はそれぞれ財の生産に費用 $c_i > 0$ を費やしており、 c_i から市場で許されるある上限価格 $M > c_i$ までの間の一様乱数で価格を決めます。また、買い手*j*には予算制約があり $v_j > 0$ 以上は支払えないので、0 から v_j までの間の一様乱数で価格を決めます。このようなエージェント同士が取引をするというモデルをシミュレーションすると、パレート最適な取引に限りなく近い結果が実現されるのです。このモデルでは、売り手も買い手も損をしない範囲でランダムに価格を決めているに過ぎないのにパレート最適な取引を達成できるのです。

このように、知性ゼロの取引者モデルは、一般均衡理論のようにエージェントの合理的選択に関する複雑な前提なしに、市場のパレート最適性を説明できることを示したのです。かつて経済学者のアダム・スミスやハイエクが述べたように、市場という制度においては、エージェントがただ価格だけを頼りに行動するだけでもパレート最適な市場均衡へ向かう性質をもっていることをこれほど単純なモデルで示した例はないと思います。この研究はまさに、市場のパレート最適性がエージェントの合理性とは無関係であることを実際に構成することによって示したという意味で、マルチエージェント実験経済学における重要な成果であると思います。

3. プログラムされた戦略

展開形ゲームや繰り返しゲームのような動学ゲームに関する実験を行うにあたっては実験デザインに2通りのアプローチが存在します。特に名前はないので、暫定的に逐次選択法とプラン選択法と呼んでおきます(Selten et al.[13]はプラン選択法を戦略選択法と呼んでいます)。逐次選択法では、ゲームの経過に従って、各プレイヤーは自分の手番になったときに初めて、可能な選択肢の中から手を選択することになります。プラン選択法では、ゲームをプレイする前にどの手を選択するつもりなのか、あらゆる自分の手番について事前にプランを決め、そのプランに従ってゲームをプレイしていくことになります。

動学ゲームにおいてはこの2つの間の区別が大事です。たとえば、逐次選択法では、各プレイヤーは他のプレイヤーの選択に逐次的に対応していくので、ゲームツリー上に実際のプレイでは到達されないノードが存在する場合があります。特に、一部のプレイヤーが均衡経路から逸脱したために実験では均衡が達成されなかった場合、均衡経路から逸脱したプレイヤーが均衡経路に従ってプレイしていれば均衡が達成されたのか、あるいは仮にそのプレイヤーが均衡経路から逸脱しなくてもその後手番になるプレイヤーの誰かが均衡経路から逸脱してしまうのか、逐次選択法による実験では判断ができません。その点、プラン選択法では事前にあらゆる手番でどういった選択を行うつもりなのか明らかですので、手番が先のプレイヤーが均衡経路から逸脱して均衡が達成されなかった場合に、そのプレイヤーさえ均衡経路に従って選択していれば均衡が達成されたのか否かは、各プレイヤーの選択したプランを検討すれば明確になるわけです。図1に示されたむかでゲームを使って具体的に考えてみましょう。

2人のプレイヤーが交互にCかSの選択を行います。各ノードの番号はプレイヤーの番号です。終端ノードの2組の数字はプレイヤーの利得で、カンマの左側がプレイヤー1で右側がプレイヤー2の利得をそれ

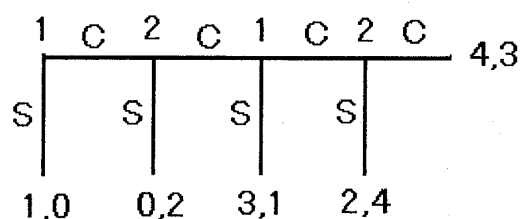


図1 むかでゲーム

ぞれ表しています。このゲームを後ろ向きの帰納法で解きますと、一番左端のノードでプレイヤー1がSを選び、プレイヤー1が1の利得、プレイヤー2が0の利得を得て終了することが最善になります。実際このゲームでは、各プレイヤーは自分の手番ではつねにSを選ぶのが最善です。

ところが、このゲームを実際に実験室で被験者にプレイさせると、必ずしもプレイヤー1が一番左端のノードでSを選ぶとは限らず、ゲームがさらに続くことがよくあります。例えば、プレイヤー1が一番左端のノードでCを選び、プレイヤー2が続く次のノードでCを選び、最後にプレイヤー1が次のノードでSを選んでゲームが終了したとします。この場合は、2人とも初めの手番では最適な行動から外れています。では、プレイヤー1が2回目の手番でCを選んでいたらどうなっていたのでしょうか。再びプレイヤー2がCを選んだのでしょうか。しかし、実際にはプレイヤー2の2回目の手番は回ってこなかったため、この実験を逐次選択法で行っていたならば結論が出せません。

ところが、実験をプラン選択法で行っていたならば、プレイヤー2が2番目の手番で何を選ぶつもりだったかそのプランを確認すればよいので、明確な結論を出すことができます。このように、到達されなかった意思決定ノードにおける行動をどのように考えるかという問題は、動学ゲームの実験を行っている者なら誰もが感じていると思います。プラン選択法を採用することで、とりあえずこの問題は回避できるわけです。

では、動学ゲームにおいてはプラン選択法の方が逐次選択法より優れているのでしょうか。これにはどうやら完全記憶ゲームならばという条件を付けなければいけないようです。それは、不完全記憶ゲームにおいては Piccione and Rubinstein[11]が健忘症のドライバー・パラドックス (absent-minded driver paradox) と呼んだ問題が生じるためです。

図2が健忘症のドライバーという1人ゲーム(意思決定問題)です。これは次のような状況をモデル化しています。いまバーでしこたま飲んだドライバーがどうやって家へ帰るか出発前にプランを立てている状況を考えます。家へ帰るには高速道路を走らねばならず、酒場から見て自分の家は1つ目と2つ目のインターチェンジのちょうど中間にあります。2つ目のインターチェンジで高速を降りた方が家に近いとします。1つ目のインターチェンジで高速を降りると人気のない

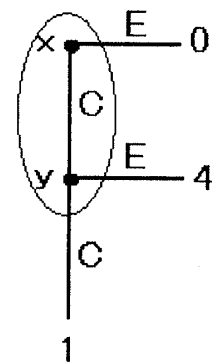


図2 健忘症のドライバー

複雑に入り組んだ田舎道を通ることになり、いまの自分の酔い具合ではとうてい家へたどり着く自信がありません。図2の意思決定ノード中、 x が1つ目のインターチェンジで、 C がインターチェンジを通りすぎる、 E が高速を降りる意思決定をそれぞれ表しています。1つ目のインターチェンジで高速を降りる場合の利得は0となっています。2つ目のインターチェンジで高速を降りると確実に家に帰り着くことができます。図2の意思決定ノード中、 y が2つ目のインターチェンジで、2つ目のインターチェンジで高速を降りる場合の利得は4となっています。2つ目のインターチェンジで降りずにさらに進んでいくと高速道路は終わり、モータールが立ち並ぶ郊外に行き着きます。ここで一夜を過ごす場合の利得は1となっています。

ところで、このドライバーは酔って意識が朦朧としているので、インターチェンジに着いたときには、それが1つ目のインターチェンジなのか、それともすでに1つ目を通り過ぎて2つ目に到着したのか記憶が不確かになっているとします。このことを、2つのノード x, y を囲う情報集合が表しています。すなわち、ドライバーは2つの意思決定ノードのどちらに今いるのか区別できないことをこの情報集合が表しています。情報の不確実性の中でも、このようにプレイヤーがすでに行われた一連の行動の一部または全部の記憶を失ってしまう場合を不完全記憶 (imperfect recall) といいます。健忘症のドライバーはこうした不完全記憶ゲームの一種になっているわけです。これがパラドックスと呼ばれるのは、酒場で帰りのプランを考えているとき、実際にインターチェンジに到達したときの最適な意思決定が異なってしまう、最適な選択が決定不能になるからです。

具体的に考えてみましょう。2つのインターチェンジを区別できない事態を想定して酒場でプランを立てている段階では、 C を2/3の確率で E を1/3の確率

で選ぶことが最適です。なぜなら、ドライバーはノード x とノード y を区別できないのですから、いずれにおいても同一の意思決定をしなければいけません。そこで、 C を確率 p で E を確率 $1-p$ で選ぶとすると、このときの期待効用は $(1-p) \times 0 + p(1-p) \times 4 + p^2 \times 1 = 4p - 3p^2$ であり、この期待効用を最大にするように p を選べば $p=2/3$ であることがわかります。

次に、ドライバーが酒場を出てインターチェンジに実際に到達した場合を考えてみます。ただし、ドライバーはそれが x なのか y なのかは区別できません。そこで、ドライバーは α の確率で今 x にいて、 $1-\alpha$ の確率で今 y にいると考えているとしましょう。このとき、 C を確率 p で E を確率 $1-p$ で選ぶとすると、このときの期待効用は $\alpha[p^2 + 4p(1-p)] + (1-\alpha)[p + 4(1-p)]$ であり、この期待効用を最大にするように p を選べば $p = \max\{0, (7\alpha - 3)/6\}$ であることがわかります。 $\alpha=1$ の場合にのみ $p=2/3$ となり、酒場でプランを立てているときと同一の選択になりますが、これ以外では一致しません。ところで、 $\alpha=1$ ということは、ドライバーは自分がいま確実に1つ目のインターチェンジにいると考えているということの意味ですが、これは完全記憶の場合にはほかなりません。すなわち、一般に不完全記憶がある場合には、このゲームをプレイする前にプランを立てたときの最適な行動と、実際にインターチェンジに到達した場合の最適な行動には食い違いが発生します。これを時間不整合性 (time inconsistency) の問題といいます。時間不整合性が発生したときには、プレイヤーは事前のプランに従うべきか、今現在において最適な選択に従うべきか決定不能になります。

ところで、展開形ゲームにおいて同一のプレイヤーが複数の異なる意思決定ノードで選択する場合、これを1人のプレイヤーが複数のノードに渡って選択すると解釈する場合と、1つのチームに属する複数のプレイヤーがそれぞれ別々の意思決定ノードにおける選択を担当すると解釈する場合があります (後者の解釈は Selten[12]によるものです)。後者の解釈に従うと、不完全記憶ゲームは、互いに連絡を取れない複数のエージェントがチームとして協調して問題解決を行う分散処理の問題となります。この場合にも、チームのメンバーである各エージェントは、チームで事前に決めたプランに従うべきか、自分に手番が回ってきた時点での最適な行動に従うべきか、やはり先ほどと同一のジレンマに直面することになります。いずれにせよ、

不完全記憶ゲームあるいはそれと同一視できるゲームでは、必ずしもプラン選択法が優れているとは言えないことに注意したいと思います。

4. 学習と進化

経済学における学習理論において古くから用いられている学習ルールは fictitious play 学習です。 t 回目における i 以外のプレイヤーの j 番目の純粋戦略の組 a_{-i}^j のプレイ頻度を $p_{-i}^j(t)$ とおくと、 i が $p_{-i}^j(t)$ を所与として効用 $u_i(a_i, a_{-i}^j(t))$ を最大にするような純粋戦略

$$a_i^* \in \arg \max_{a_i \in A_i} \sum_j p_{-i}^j(t) u_i(a_i, a_{-i}^j(t))$$

を $t+1$ 回目に用いる学習ルールを fictitious play 学習モデルといいます。ここで、 t 回目における i 以外のプレイヤーの純粋戦略の組 a_{-i}^j のプレイ頻度 $p_{-i}^j(t)$ は、 $t-1$ 回目までに a_{-i}^j が実際にプレイされた回数を $k_{-i}^j(t-1)$ とするとき、次の式によって更新されます。

$$k_{-i}^j(t) = k_{-i}^j(t-1) + \begin{cases} 1 & \text{if } a_{-i}(t) = a_{-i}^j \\ 0 & \text{それ以外} \end{cases}$$

$$p_{-i}^j(t) = \frac{k_{-i}^j(t)}{\sum_j k_{-i}^j(t)}$$

このように、他のプレイヤーがプレイした純粋戦略の組の相対頻度を所与として、その頻度のもとで期待効用が最大になるように純粋戦略を選ぶのが fictitious play 学習ルールです。ゼロ和ゲームにおいては、どんな初期値から出発しても fictitious play 学習によって必ず Nash 均衡 (ミニマックス解) に収束することが知られています。しかし、一般の非協力ゲームでは必ずしも Nash 均衡に収束しない場合があります²。なぜでしょうか。それは、この学習ルールに限らず、多くの学習ルールが環境の定常性を仮定しているからだと思います。ここで環境の定常性とは、自分が学習するとき、他のプレイヤーがあらかじめ決まったプランに従ってプレイしているとみなすことを意味します。言ってみれば、他のプレイヤーはあらかじめプログラムされた通りに混合戦略をプレイする機械であり、ゲームの間プログラムに変更はないという前提のもとに、各プレイヤーは毎回の手の系列から相手のプログラムの構造を推測して、その推測に基づき期待効用を最大にするように選択をするという学習が想定されている

² 学習理論全般については Fudenberg and Levine[5] を参照してください。

わけです。

しかし、一方では、各プレイヤーは他のプレイヤーが毎回選ぶ手の系列に応じて適応的に自分自身の行動パターンを変えているのです。実際には自分も含めて全員が毎回適応的に行動を変えているにも関わらず、他のプレイヤーについてはあらかじめプログラムされた行動を選んでいくに過ぎないとみなすことは、まったく無理な仮定ではないですが、問題を含んでいます。

この問題は、プレイヤーの間で学習ルールが共有知識となっていると仮定するならば、各プレイヤーが適応的学習を行う他のプレイヤーの将来の行動を予測して、その予測に基づき期待効用が最大になるように選択を行うと考えると良さそうに見えます。こうした考え方を極限まで推し進めると、将来にわたって他のプレイヤーの行動を正確に予測できるという合理的期待形成モデルになります。このような予測に基づく学習を実行するには学習ルールに関する連立方程式、あるいは差分方程式系を解くことになるでしょう。しかし、単純な有限ゲームでは、こうした方程式系を解くことより直接に混合戦略の Nash 均衡を求める方がはるかに簡単なはずで、こうなると、なぜプレイヤーが Nash 均衡を直接求めないで精緻な学習ルールを用いなければならないのか、という新しい問題が発生します。

ここで述べているのは、各プレイヤーが他のプレイヤーに関する適切なモデルを如何にして形成するかという問題です。各プレイヤーが形成する相手の行動パターンと、他のプレイヤーが実際に行っている行動との食い違いをどのように埋めていくか、という問題です。これはまた、プレイヤーは計算能力に限界があるので必ずしも合理的な選択ができず、そのために学習を通じた選択を行うにも関わらず、プレイヤーの行動を外から眺めている研究者の目から見て明らかに計算が容易な問題を解く代わりにより一層難しい問題にプレイヤーが取り組むのはなぜか、というプレイヤーのモデル化上の問題です。

動学ゲームを実験する際にこうした問題を避けるためには、やはりプラン選択法をもちいるのがよいでしょう。すべてのプレイヤーがあらかじめ作成したプラン通りにプレイすることが共有知識になっているかぎり、先に述べたような難しい問題に悩まされることはありません。しかし、一方でプレイヤーは事前に行動プランを作成しなければならないので、ゲームの構造や戦略的関係についてかなり十分に検討する機会を与

えないと、杜撰なプランしかできず実験は失敗するということにもなりかねません。また、なぜプレイヤーは自分自身が直接プレイしないで自分の戦略をプログラムに託さねばならないのか、その理由が明確でないといけません。ただ、プラン選択法を採用することで、人間の被験者が生み出すプランとコンピュータ・プログラムが生み出すプランとの比較研究ができる利点がありますので、動学ゲームの研究はプラン選択法をもちいることがこれから増加していくと考えられます。

続いて進化学習の問題に移ります。遺伝的アルゴリズムやクラシファイア・システムのような進化学習モデル (Goldberg[7]) は、マルチエージェント経済学でも多用される手法です。進化学習を経済モデルに適用する際の問題点は、進化ゲーム一般について言えることですが、それが不完備情報の展開形ゲームに適用できないということです。展開形ゲームは戦略形ゲームに比べて主体のもつ情報の質・量の違いを表現したり、意思決定のタイミングを適切に表現できる点で優れています。ところが、基本的に主体のランダム対戦を想定する進化ゲームにおいてはプレイヤーの意思決定のタイミングを問題にすることができません。また、進化ゲームにおいては完全・完備情報が仮定されています³。このことを考え合わせると、進化ゲームにおいては、(1)利得行列は対称でプレイヤー間で共有知識になっているので不完備情報ではなく、(2)各プレイヤーはどの時点においても各純粋戦略を使用する個体が集団にどれだけの割合で存在するかを知っているので、各プレイヤーが毎回選んだ手に関する情報が共有されており、ゲームの経過に関する完全情報をもっていることとなります。これらの前提があるからこそ、レプリケータ・ダイナミクス (Hofbauer and Sigmund[8]参照) で行われるように、各純粋戦略によって得られる利得と平均利得との差を計算できるのです。逆にいえば、ゲームの経過について不完全情報がある場合には、毎回どれだけのプレイヤーが各純粋戦略を選んだのかわからないので、集団内の各純粋戦略

³ ちなみに、不完備情報 (imcomplete information) とは利得に関する情報についてプレイヤー間に非対称性があることを意味し、不完備情報 (imperfect information) とはゲームの経過に関する情報についてプレイヤー間に非対称性があることを意味しています。

⁴ もちろん、意思決定において必要な戦略的情報はすべて戦略形ゲームに含まれているとして、戦略形ゲームを中心に考える Kohlberg and Mertens[9]などのゲーム理論家もいます。

の採用比率がわからず、平均利得を求めることができません。よって、プレイヤー間の情報の非対称性をモデル化する際には進化ゲームは適していないといえます。

もちろん、不完備情報の展開形ゲームを戦略形ゲームに変形することは容易にできますので、その上で進化学習を行えばよいと考えられるかもしれませんが、それでは展開形でしか表現できない意思決定のタイミングや情報の非対称性に関する情報が失われてしまいます⁴。

いずれにしても、進化モデルを不完備情報の展開形ゲームに適用するには注意が必要です。今後マルチエージェント実験経済学の研究が進んで、さらに不完備情報ゲームの研究を進めていく際には、進化ゲームのパラダイムを越えて、新しい学習方式を考えていく必要があると思います。

5. おわりに

ここまで実験経済学者の立場からマルチエージェント実験経済学的方法的基礎をなす3つのキーポイントについて検討してきました。経済現象をモデル化するにあたって不完備情報ゲームを取り扱う機会はこれからもますます増加していくと思われます。ここで述べた問題点をふまえて、不完備情報ゲームにおけるマルチエージェント実験経済学の方法論が、具体的なモデルを構成することでさらに深められていくことを願っています。

参考文献

- [1] Axelrod, R.: *The Evolution of Cooperation*, Basic Books, (1984).
- [2] Axelrod, R.: *The Complexity of Cooperation*, Princeton University Press, (1997).

- [3] Boyd, R. and Lorberbaum, J. P.: "No pure strategy is evolutionarily stable in the iterated prisoner's dilemma game", *Nature (London)*, 327, 58-59, (1987).
- [4] Friedman, D. and Sunder, S.: 『実験経済学の原理と方法』, 同文館, (1999).
- [5] Fudenberg, D. and Levine, D. K.: *The Theory of Learning in Games*, The MIT Press, (1998).
- [6] Gode, D. K. and Sunder, S.: "Allocative efficiency of markets with zero intelligence traders: market as a partial substitute for individual rationality", *Journal of Political Economy*, 101, 119-137, (1993).
- [7] Goldberg, D. E.: *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, (1989).
- [8] Hofbauer, J. and Sigmund, K.: *Evolutionary Games and Population Dynamics*, Cambridge University Press, (1998).
- [9] Kohlberg, E. and Mertens, J. F.: "On the strategic stability of equilibria", *Econometrica*, 54, 1003-1037, (1986).
- [10] Kreps, D., Milgrom, P., Roberts, J., and Wilson, R.: "Rational cooperation in the finitely repeated prisoner's dilemma", *Journal of Economic Theory*, 27, 245-252, (1982).
- [11] Piccione, M. and Rubinstein, A.: "On the interpretation of decision problems with imperfect recall", *Games and Economic Behavior*, 20, 3-24, (1997).
- [12] Selten, R.: "Reexamination of the perfectness concept for equilibrium points in extensive games", *International Journal of Game Theory*, 4, 25-55, (1975).
- [13] Selten, R., Mitkewitz, M., and Uhlich, G. R.: "Duopoly strategies programmed by experienced players", *Econometrica*, 65, 517-555, (1997).