

## 私の研究について

府内直樹 Naoki Funai  
滋賀大学 経済学部 / 教授

私はこれまで人々がゲーム（各人の選択が自身の利得だけではなく、相手の利得に影響を与える状況）や不確実性をともなう個人の意思決定問題に繰り返し直面し、過去の経験から学習した結果、人々の行動が長期的にある均衡に収束するかどうかを理論的に分析してきました。

例えば、歩行者2人が狭い通りですれ違う際に、右に避けるか左に避けるかの意思決定をしなければいけない状況を考えてみます（もちろん、もっと一般的なゲームを研究では対象にしています）。お互いが各人から見て同じ方向を選んで問題なくすれ違う方が、お互いが別の方向を選んでぶつかりそうになるよりも好ましいとします。

経済学では、このようなゲームにおける人々の行動をナッシュ均衡を用いて予測します。上記の歩行者のゲームにおいては、お互いに右に避けるのが一つのナッシュ均衡であり、お互いが左に避けるのがもう一つのナッシュ均衡となります（混合ナッシュ均衡もありますが、以後の議論では無視します）。

ナッシュ均衡が人々の行動予測として使用される根拠の一つは、ナッシュ均衡に関する以下の考え方です：人々がゲームを完全に理解し（自分や相手の選択肢や利得、相手が何を知っているかなど）、合理性を持ち、相手の行動に対し正しい推論を行えば、一発でナッシュ均衡が実現する。つまり、上記の例においては、実際にすれ違う前に、人々がゲームに関する知識、合理性をもとに正しく推論した結果、一発で皆が右に避ける（もしくは皆が左に避ける）ことが実現するというわけです。もちろん、重要な意思決定問題などでは上記のように推論し意思決定を行うと思いますが、ゲームは歩行者の問題も含め多々存在し、全てのゲームにおいて熟考するのは思考コストがかかります。また、講義で経済実験を行います。正しくゲームを理解し、正しく相手の行動を推論してナッシュ均衡を一発でプレイすることは難しそうです。

ここで、以下のナッシュ均衡に関する別の考え方を紹介します：意思決定者が繰り返し同じゲームに直面し、試行錯誤の結果、最終的にはナッシュ均衡に従って行動する。上記の歩行者のゲームの例においては、歩行者がぶつかりそうになったり、問題なくすれ違ったりということを何回も経験すれば、最終的に各人が右（もしくは各人が左）に避けるようになることに対応します。私はこちらの考え方に注目し、学習者がゲームを繰り返し経験し学習した結果、最終的にナッシュ均衡（や質的応答均衡, McKelvey and Palfrey, 1995）に従って行動するかどうか、収束するのであればその条件を理論的に考察しています。

特に、私が研究対象としている適応学習理論では、自分の選択肢に関する知識を持ち、プレイ後に利得を観測でき、自己の過去の経験から最も良い結果をもたらした選択肢を近視眼的に（最も高い確率で）選択するという比較的シンプルな条件のみを意思決定者に求めています。つまり、最初のナッシュ均衡の解釈のように、相手がどのような相手なのか、相手がどのような選択肢を持っているのか、相手が同じような学習者なのかなどに関する知識を要求していません。学習者は繰り返し同じゲームを経験した結果、最終的にあたかもゲームを完全に理解し、合理性を持ち、正しい推論ができるかのように振る舞うようになるのです。

さて、学習者の学習や行動基準をモデル化し、理論的に分析するのですが、どのようにモデル化すると良いでしょうか。人々のゲームや個人意思決定問題での学習に関しては、心理学や実験経済学で様々な研究が行われています。特に、実験経済学では、どのような学習モデルがもっとも実際の人々の行動を予測できるかを考察します。例えば、ErevとRoth（1995, 1998）は実験で観察されたデータを心理学のThorndike（1898）などの知見を生かした強化学習モデルで説明しようとし、CamererとHo（1999）は

ErevとRothの強化学習モデルやBrown (1951) や Fudenberg とKreps(1993) の仮想プレイ学習などを含むexperience-weighted attraction (EWA) 学習モデルでデータを説明しようとしています。一般的に、実験経済学では心理学などの知見を生かしたモデルが提唱され、観察されたデータをベースに、各モデルのパラメーターの推定などが行われます。

一方、私の研究では、実験経済学などで提唱された学習モデルの理論的考察を行います。特に、既存の均衡に収束するかどうか、また多々存在する学習モデルの共通点を模索しています。例えば、Funai(2019)では、EWA学習モデルなど、様々な学習過程を包括する適応学習過程が質的応答均衡に収束する条件を示し、Funai(2022)では、ErevとRothの強化学習において、既存の文献で考察されなかった各学習者が選ばれなかった他の選択肢の利得情報を得られる状況 (EWA学習では考察されているが、理論的分析は無し) を考察し、強化学習者の行動がナッシュ均衡や質的応答均衡に収束する条件を示しました。

私が研究対象としている学習モデルはEWA学習モデルなどと同じく、主に選択ルールと更新ルールから特徴づけられます。まず、各意思決定者は自分の選択肢を評価 (どれくらい利得が得られるか) し、その評価をベースに意思決定を行います。その際、選択ルールは各選択肢の評価をもとに各選択肢をどのような確率で選ぶかを示すもの、更新ルールはどのような利得情報を使って、選択肢の評価をどのように更新するかを示すものです。更新ルールに関しては、ErevとRothはThorndike (1989) の提唱した性質、「過去により良い結果をもたらした選択肢がまた選ばれる可能性が高くなる」をもとに、評価に対応する「過去の結果」を表すものとして、過去の利得 (選ばれた選択肢からの利得のみ) の総和を採用し、一方、Brown (1951) やFudenberg とKreps (1993) の仮想プレイモデルなどでは、過去の利得 (選ばれていない他の

選択肢の利得も含める) の加重和を採用しています。ErevとRothは正の利得を仮定しているので、利得の総和は経験すればするほど上昇しますが、利得の加重和は、ある科目の悪い成績がGPAを下げるようなことがあるように、経験した利得に応じて下がる場合があります。一方、選択ルールに関しては、ErevとRothなどではLuceの線形選択ルール、EWA学習モデルや仮想プレイ学習などでは、Logit選択ルールが採用されています。

この選択ルールと更新ルールの選択に関しては注意が必要です。私の2013年、2019年、2022年の論文は、選択ルール (やその正確性、どれくらいランダム性が強い) か、更新ルール (ルール自体、更新時に使用する利得情報の種類など) の各要素に関して、どれを選択するかで意思決定者の長期的な行動が異なることを示しています。例えば、(i) 評価が過去の利得の総和、(ii) 他の選択肢の利得を確認できず、(iii) 選択にランダム性がない場合は、最初に選ばれた (最適でない可能性のある) 選択肢が選ばれ続けます (評価がさらに上がるので)。ここで、評価を加重和に変えると、相対的に悪い利得を生み出す選択肢の評価は下がることがあるので、他の選択肢が選ばれる可能性があります。一方、(i) 評価が利得の加重和、(ii) 他の選択肢の利得を確認でき、(iii) 評価の差に応じて選択にランダム性が発生する場合は、長期的な評価の差に応じ、選択にランダム性が残ります (GPAなどの加重和の差は無限大にならないので、悪い選択肢もある程度選ばれる)。一方、評価を過去の利得の総和にする場合、一番良い選択肢とその他の選択肢の評価の差がどんどん大きくなるので、長期的にランダム性が小さくなります (ほぼ最適な選択肢が選ばれる)。つまり、最適な選択肢が最終的に (ほぼ確実に) 選ばれるかどうかは評価の形式だけに注目してもわからないということです。これをゲームの文脈で言い換えると、(ランダム性のない) ナッシュ均衡に収束するかどう

かは各要素のみに注目してもわからないことを示すことができます。例えば、データから意思決定者は選ばれた選択肢からの利得しか注目しないと判断されても、他の要素(どのように評価を更新するか、どれくらい選択に関してランダム性が強いかなど)についても注目しないと、最終的に意思決定者の行動がナッシュ均衡に対応するかどうかは判断できないということです。

次に、私が現在行っている研究の紹介をします。ここまで紹介した研究を含め、一般的な学習理論においては、同質な学習者同士がゲームをプレイし、同じように学習を行うケースを考察しています。しかし、同質な学習者同士のみがゲームをプレイして学習するというケースは稀ではないのかという疑問をもち、私の実体験と照らし合わせて考えてみました。私が海外のある地方に移住した際に、誰かに教育を受けたわけではありませんが、狭い通りで人とすれ違う時に、いつの間にか右に避けるようになりました。一方、帰国後に同じような状況に直面することがありましたが、うまくすれ違うことができないことが多々ありました(学生に質問してみると、状況に応じて右に避けたり、左に避けたりするようです)。なぜ私は意識せずに海外で右に避けるようになったのか、日本ではなぜそれが起きないのかに疑問を持ちました。一つの原因として、私のような新規移住者だけでなく、右に避けるという慣習に従い続ける既存住人の存在が大きいのではないかと感じました。歩行者の例で言えば、既存住人が右に避ける慣習に従い、学習をしていない人が多ければ、私のような新規移住者も最終的に右に避けるようになり、右に避けるという慣習が維持されるのではないかと考えました。

現在取り組んでいる研究(Funai, 2021b)では、ゲームにおける学習において、既存住人の慣習が新規移住者の学習にどのように影響を与えるかを理論的に考察しています。例えば、私のような、右に避ける

という慣習を知らない新規移住者が他の新規移住者だけでなく、慣習に従っている既存住人と交流する可能性がある場合、最終的に既存の慣習通りに行動する条件は何か、また、新規移住者間で異なる慣習を形成し、既存住人とのミスコーディネーションを引き起こす条件は何かを理論的に分析しています。

今後は、(i) 学習者が対戦相手を選択できるケースや(ii) 有限繰り返しゲームでの学習の理論的分析(Funai, 2021a)を行いたいと考えています。また留学や帰国後に知り合ったゲーム理論家、実験経済学者や数学者との共同研究なども視野に入れ、研究テーマを広げていきたいと思っています。

#### 参考文献

- ◎Brown, G. W., 1951. Iterative solutions of games by fictitious play, In *Activity Analysis of Production and Allocation*, T. C. Koopmans (Ed.), New York: Wiley.
- ◎Camerer, C., Ho, T. H., 1999. Experience-weighted attraction learning in normal form games. *Econometrica* 67, 827-874.
- ◎Erev, I., Roth, A. E., 1998. Predicting how people play games: reinforcement learning in experimental games with unique mixed strategy equilibria. *American Economic Review* 88, 848-881.
- ◎Fudenberg, D., Kreps, D. K., 1993. Learning mixed Nash equilibria, *Games and Economic Behavior* 5, 320-367
- ◎Funai, N., 2013. An adaptive learning model in coordination games. *Games* 4, 648-669.
- ◎Funai, N., 2014. An adaptive learning model with foregone payoff information. *B.E. Journal of Theoretical Economics* 14, 149-176.
- ◎Funai, N., 2019. Convergence results on stochastic adaptive learning. *Economic Theory* 68, 907-934.
- ◎Funai, N., 2021a. Adaptive learning in finitely repeated games. Available at SSRN: <https://ssrn.com/abstract=3830132> or <http://dx.doi.org/10.2139/ssrn.3830132>

- ◎Funai, N., 2021b. Stochastic adaptive learning with committed players in games with strict Nash equilibria. Available at SSRN: <https://ssrn.com/abstract=3944342> or <http://dx.doi.org/10.2139/ssrn.3944342>
- ◎Funai, N., 2022. Reinforcement learning with foregone payoff information in normal form games, *Journal of Economic Behavior and Organization* 200, 638–660.
- ◎McKelvey, R., Palfrey, T., 1995. Quantal response equilibria for normal form games, *Games and Economic Behavior* 10, 6–38.
- ◎Roth, A. E., Erev, I., 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior* 8, 164–212.
- ◎Thorndike, E. L., 1898. Animal intelligence: an experimental study of the associative processes in animals. *Psychological Monographs*, 2, 8.