

令和 6 年 9 月 30 日現在

機関番号：82401

研究種目：国際共同研究加速基金（国際共同研究強化(A））

研究期間：2022～2023

課題番号：21KK0247

研究課題名（和文）情報共有が限定された社会的ジレンマにおける間接互惠性の計算科学的研究

研究課題名（英文）Computational study on indirect reciprocity for social dilemma under private reputation

研究代表者

村瀬 洋介（Murase, Yohsuke）

国立研究開発法人理化学研究所・計算科学研究センター・研究員

研究者番号：30709770

交付決定額（研究期間全体）：（直接経費） 7,700,000円

渡航期間： 9ヶ月

研究成果の概要（和文）：本研究課題では、マックスプランク進化生物学研究所に九ヶ月滞在し、間接互惠性の進化に関して理論的に研究を行った。具体的な研究成果は主に二つある。一つ目は、間接互惠性における評判の双方向性と確率性を導入した新たなモデルを提案し、協利行動の維持メカニズムを理論的に解明した。二つ目は、社会規範の進化過程をスーパーコンピュータを用いて計算科学的に研究し、協利的な社会規範の進化にグループ構造が重要であることを明らかにした。

研究成果の学術的意義や社会的意義

どちらの研究も、今後の間接互惠性の理論、実験研究における今後の研究の参照点となる重要な知見である。また計算科学的な手法を初めて進化ゲーム理論に持ち込み、科学的意義の大きい知見を得た。新しい手法を開拓し、この手法の有効性を実証したという意義も大きい。長期的には、我々の現在の社会の成り立ちについての理解を深めるという文化的、基礎科学的な意義に加え、気候変動や紛争などグローバルスケールの課題において、協利関係を構築するメカニズムのデザインに寄与するものと期待している。

研究成果の概要（英文）：In this research project, I spent nine months at the Max Planck Institute for Evolutionary Biology conducting theoretical research on the evolution of indirect reciprocity. There are two main research outcomes. The first is the proposal of a new model that incorporates dual and stochastic reputation updates in indirect reciprocity, theoretically elucidating the mechanisms for maintaining cooperative behavior. The second is the computational study of the evolution of social norms using a supercomputer, revealing that group structure is crucial for the evolution of cooperative social norms.

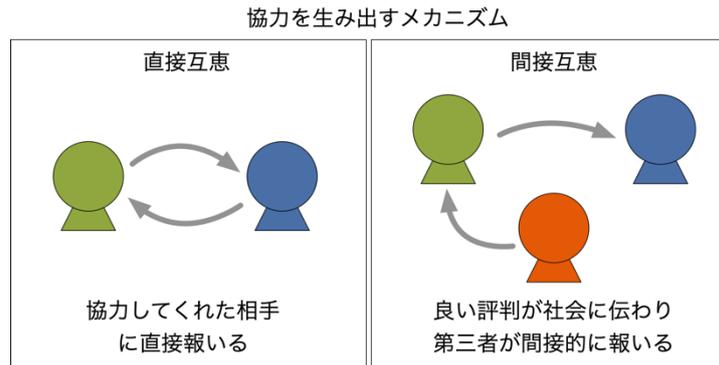
研究分野：数理科学

キーワード：進化ゲーム理論 協利 間接互惠性 大規模計算

## 1. 研究開始当初の背景

我々が日常的に出会う様々な問題の裏側には「個人が合理的にふるまった結果として社会的に望ましくない状態に至る」という社会的ジレンマが普遍的に存在している。社会的ジレンマを解決し、協力を生み出すメカニズムとしてはいくつか知られているが、「他社の評判」についての情報を共有することによって協力状態に至るメカニズムを間接互恵性という。本研究課題では、マックスプランク進化生物学研究所の Christian Hilbe 博士のグループに九ヶ月にわたり滞在し、間接互恵性についての理論研究を行った。

間接互恵性とは、「評判＝ある人についての情報」によって社会集団において協力を維持する仕組みのことである。例えば、A さんが B さんに対して協力的な行動をするかどうかを考えよう。B さんはそれによって利益を得るが、A さん自身は協力することに何かしらのコスト(労力、金銭、時間など)を伴う。今後 B さんと A さんは何度も出会うことはないとする、A さん視点では何も協力するインセンティブがなく、協力行動は Nash 均衡にならない。しかし、もしその行動を他の第三者が観察していて A さんについての噂が社会全体に共有されるとすると状況は全く異なる。A さんは、今後他の人々と交流があるので、彼らに悪い評判が伝わってしまうと協力が得られなくなってしまうかもしれない。つまり「評判」という情報が伝わる状況下においては、A さんが自分でコストを払ってまで利他的な行動をとることが合理的な判断の結果として現れるのである。このように協力を生み出す仕組みが間接互恵性である。



ちなみに、二人のプレイヤーが何度も繰り返し相互作用する状況において、協力してくれた相手に直接報いるメカニズムを「直接互恵」といい、基課題では「直接互恵」を主に研究している。

間接互恵のこれまでの理論研究における標準的なモデルは以下のようなものである。十分大きなサイズの集団を考える。各時刻において、ドナー役、レシピエント役の二人をランダムに選ぶ。ドナーはレシピエントに対して協力するか、協力しないかを行動を選択し、協力した場合は、ドナーはコスト  $c$  を支払い、レシピエントは利益  $b (>c)$  を得る。協力しない場合は、どちらにも利得の変化はない。ここでの行動は、集団の他の人にも観測され、ドナーの評判が更新される。以上をドナー、レシピエントを変えながら十分に長い期間繰り返す。

さて、このような状況において協力を安定に維持するためには「相手の評判におうじて適切に行動を変える」「ドナーの行動に応じて適切に評判を割り当てる」という二つのことが必要になる。前者の行動を決めるルールを「行動ルール」と呼び、後者の評判の割り当てを「評価ルール」と呼ぶ。この二つの組みを「社会規範」と呼んでいる。

従来の研究では、ある人についての評判が誤解なく集団全体で共有されているという仮定(公的評価: **public assessment**)をおく、研究が主流であった。このモデルの理論的な取り扱いには比較的容易であり、先行研究の蓄積も多かった。しかし、この仮定は必ずしも現実的ではなく、ある人が良い人か悪い人かについて意見の相違が起きることは日常的に起きている。その意見の相違がさらなる意見の相違を生み、社会の分断が起きることは我々も感覚的によく知っている。そのような意見の相違を取り入れたモデルを私的評価(**private assessment**)と呼ぶ。このモデルは、数理的な扱いが難しいだけでなく、公的評価でうまく機能していた社会規範も私的評価モデルではうまくいかず、協力を維持することが困難であることが知られている。(Hilbe et al. PNAS 2018 など)。本研究では、このような「情報共有が限定された状況」において協力を生み出すことは可能なのか、もし可能ならどのような条件が必要なのか、その背後にあるメカニズムは何かについて理解するための理論研究を行った。

## 2. 研究の目的

評判情報の共有が限定された私的評価モデルにおいて協力が促進されるのか、促進されるとしたらそのメカニズムは何か数理モデルを用いて明らかにする。

### 3. 研究の方法

研究期間中に主に以下の2つの研究(1, 2)を行った。

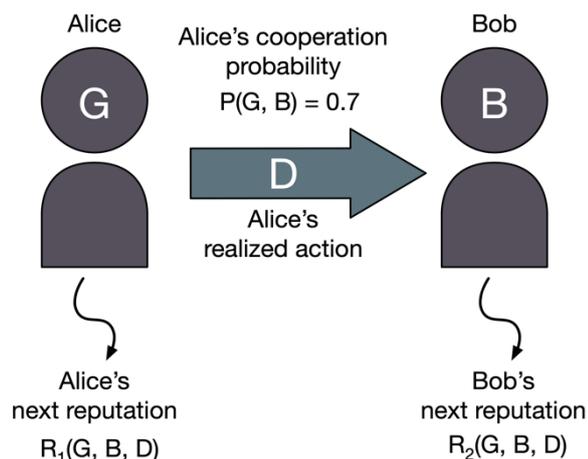
#### (1) 双方向のかつ確率的な評判更新における間接互惠性

この研究では、既存のモデルを二つの方向性に拡張したモデルを研究した。この研究は公的評価モデルであるが、私的評価モデルの研究の理論的基盤という位置付けになる。

一つ目の拡張は「評価の双方向性」である。これまでの理論モデルでは、評判の更新は行動を決定するドナーに対してのみ行われると仮定してきた。しかし、我々の日常の行動を考えてみると、行動の受け手側のレシピエントの評判も変わることが往々にして起きているように見える。

例えば、犯罪の被害者に対して同情的になり、そのような被害者をより道徳的な人であると評価するというバイアスが存在することが実験的に確かめられている (Jordan et al. Sci. Adv. (2021))。また、評判の悪い人から協力行動を受けた人が、その人の仲間だと思われて自分の評価を落とすというようなシナリオも考えられるだろう。これらのような行動の受け手の評判も変わるような要素を取り入れたモデル (ここでは「双方向的」モデルと呼ぶ) を数理的に分析した。

二つ目の拡張は「確率性」の導入である。これまでの研究では、行動の決定やその評価が決定論的に決まると仮定してきた。しかし、現実には社会の雰囲気や天候などさまざまなモデルの外部の要因にも依存するはずで、必ずしも行動や事前の評判のみに依存して評判が更新されるとは限らない。そのような要因をモデル化するために確率性を導入したモデルを考える。



この双方向確率モデル (上図) は、既存のモデルに比べて複雑度が増し、規範の数は無限大になるため、網羅的に列挙することもできない。我々はこのモデルに対して理論的な解析が可能であることを発見し、研究を進めた。

#### (2) 社会規範の進化

社会規範がどのように生まれ社会の中で定着するのか、その進化の過程についての研究を行った。既存の研究は、すでに社会規範が存在していると仮定し、その社会規範が Nash 均衡として協力解を持つかどうかを問うものが多かった。一方で、そのような社会規範は、そもそもどのように生まれ、どのような進化的な経路を辿ったのだろうか？その中で鍵となる社会規範はどのようなものなのか？

この問いは素朴で基礎的なものであるにもかかわらず、これまで十分に研究が行われていなかった。そのような進化プロセスを愚直に研究するためには、ありうる社会規範の数を  $K$ 、集団のサイズを  $N$  とすると  $O(NK^2)$  の計算量が必要になるため、シミュレーションをすることすら容易ではない。多くの既存研究では、ある特定の社会規範を恣意的に選び、それが ALLC や ALLD といった代表的な戦略に対して進化的に安定かどうかを調べることのみが行われてきた。また、社会規範が定着するまでは、プレイヤーによって利用している社会規範が異なり意見が必ずしも一致しないため、必然的に私的評価のモデルを採用する必要がある。このような理論的な困難さゆえに、この問題はこれまで限定的にしか研究されてこなかった。

我々はこのような制約を取り払い、社会規範を限定しない状況においてどのような社会規範が選択されていくかを初めて研究した。膨大な計算量が必要になるが、スーパーコンピューター富嶽を用いることでこの困難を克服した。

## 4. 研究成果

### (1) 双方向のかつ確率的な評判更新における間接互惠性

上で述べたモデルに対して、我々は理論的に解を求めることに成功した。より具体的には、協力 Nash 均衡を持つ社会規範の必要十分条件を厳密に求めることに成功した。

この厳密解からさまざまな興味深い知見が得られた。

- このモデルは単方向かつ決定論的なモデル（既存の理論モデルと同じ）を特殊な場合として含んでいる。この場合、すでによく知られている Leading eight という 8 種類の社会規範を再現する。しかし、それだけでなく、他にも 16 種類の解が存在することが明らかになった。（我々は “Secondary Sixteen” と名付けた。）また、この合計 24 種類以外に存在し得ないということも明らかにした。
- 双方向性を導入することにより、エラーが起きた際により早く評判を回復し、協力状態に戻りやすくなることを発見した。特に、「事前の評判の悪いレシピエントが、裏切り行動を受けた場合にレシピエントの評判を回復するかどうか」が重要な役割を果たす。しかし、一方で双方向性があることにより評判が早く回復するようになると、協力の利益  $b$  がより高い状況でないと協力が安定にならないという副作用もあることが明らかになった。
- 確率性を導入することにより、より低い  $b$  でも協力を安定かすることができるケースがあるということも発見した。さらに、「処罰の正当化」や「謝罪の受け入れ」といった行動が必ずしも起きなくても協力を維持できるということも発見した。「部分的な」処罰の正当化、「部分的な」謝罪の受け入れなども起きうるということが明らかになった。

この成果は論文として、Plos Computational Biology 誌に掲載され、また国内外の学会で発表された。

### (2) 社会規範の進化

社会規範の進化がどのように起きるかを計算的手法により研究した。その結果、素朴によく混合された単一の集団を考えた場合、ある程度複雑な社会規範を考えたとしても、協力は進化し得ないということが明らかになった。近年の先行研究で進化的安定な社会規範が報告されていたので、そのような社会規範が進化の結果として現れることを予想していたのだが、驚くべきことにそのようなことは起きないことが明らかになった。これは、既存研究がすでに確立した社会規範の安定性を議論しているのに対し、我々は、社会規範が初期にどのように発生し、定着していくかも含めて考えているためである。

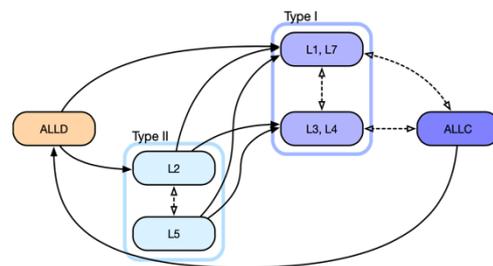
しかし他方で、集団の中に何か構造があると協力が進化しうることを発見した。グループ構造があると、グループ内での競争とグループ間での競争に勝ち抜く必要がある。そのような条件では、むしろ協力的な社会規範が進化しうるということが明らかになり、社会規範と集団の構造の複雑な関係についての洞察が得られた。また、社会規範がどのような経路で進化するのかについても明らかになった。特に、Leading Eight の中のある一つの社会規範「L1」が重要な役割を果たすことが明らかになった。（右図）

L1 はこれまでの研究ではあまり注目されてこなかった社会規範である。今後の実験的、あるいはフィールド研究において重要な指針となることが期待される。

この成果は論文として投稿し、本報告書執筆時点で査読中である。

これらの成果に加えて、滞在中に基課題についても進展が見られ、その成果は、Scientific Reports, Plos Computational Biology に出版された。また、これらの科学的成果に加えて、欧州滞在中に多くの研究者と知り合いになることができ、人脈の構築という点でも意義深い滞在となった。

Most important transitions among third-order strategies



## 5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 4件/うちオープンアクセス 4件）

1. 著者名 Murase Yohsuke, Hilbe Christian	4. 巻 19
2. 論文標題 Indirect reciprocity with stochastic and dual reputation updates	5. 発行年 2023年
3. 雑誌名 PLOS Computational Biology	6. 最初と最後の頁 e1011271
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pcbi.1011271	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する
1. 著者名 Murase Yohsuke, Baek Seung Ki	4. 巻 19
2. 論文標題 Grouping promotes both partnership and rivalry with long memory in direct reciprocity	5. 発行年 2023年
3. 雑誌名 PLOS Computational Biology	6. 最初と最後の頁 e1011228
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pcbi.1011228	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する
1. 著者名 Murase Yohsuke, Hilbe Christian, Baek Seung Ki	4. 巻 12
2. 論文標題 Evolution of direct reciprocity in group-structured populations	5. 発行年 2022年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 18645
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41598-022-23467-4	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する
1. 著者名 Oishi Koji, Ito Hiroto, Murase Yohsuke, Takikawa Hiroki, Sakamoto Takuto	4. 巻 17
2. 論文標題 Evolution of global development cooperation: An analysis of aid flows with hierarchical stochastic block models	5. 発行年 2022年
3. 雑誌名 PLOS ONE	6. 最初と最後の頁 e0272440
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pone.0272440	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Lee Sanghun, Murase Yohsuke, Baek Seung Ki	4. 巻 548
2. 論文標題 A second-order stability analysis for the continuous model of indirect reciprocity	5. 発行年 2022年
3. 雑誌名 Journal of Theoretical Biology	6. 最初と最後の頁 111202 ~ 111202
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jtbi.2022.111202	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕 計13件 (うち招待講演 3件 / うち国際学会 7件)

1. 発表者名 Y. Murase
2. 発表標題 Friendly-Rival Strategies for the Iterated Prisoner's Dilemma
3. 学会等名 CCP2023 - 34th IUPAP Conference on Computational Physics (国際学会)
4. 発表年 2023年

1. 発表者名 Y. Murase
2. 発表標題 Indirect reciprocity with stochastic and dual reputation updates
3. 学会等名 Collective Phenomena in complex systems: problems inspired by Ecology and Evolution (CoPhEE2023) (国際学会)
4. 発表年 2023年

1. 発表者名 村瀬洋介
2. 発表標題 確率のおよび双方向的な評判更新による間接互惠性
3. 学会等名 2023年度日本数理生物学会年会
4. 発表年 2023年

1. 発表者名 村瀬洋介
2. 発表標題 確率のおよび双方向的な評判更新による間接互惠性
3. 学会等名 日本物理学会第78回年次大会
4. 発表年 2023年

1. 発表者名 村瀬洋介
2. 発表標題 HPCを用いた進化ゲーム理論研究
3. 学会等名 HPCIC 計算科学フォーラム (招待講演)
4. 発表年 2023年

1. 発表者名 Y. Murase
2. 発表標題 Indirect reciprocity beyond dichotomic reputation updates
3. 学会等名 Complex Systems Research Exchange (国際学会)
4. 発表年 2023年

1. 発表者名 Y. Murase
2. 発表標題 Friendly Rival Strategies for iterated Prisoner's Dilemma
3. 学会等名 RHINOS 2023 (国際学会)
4. 発表年 2023年

1. 発表者名 村瀬洋介
2. 発表標題 単純二項対立ではない評価による間接互惠性
3. 学会等名 ヒトの協力行動の理解 -間接互惠性とその周辺- (招待講演)
4. 発表年 2024年

1. 発表者名 Yohsuke Murase
2. 発表標題 Meta-modeling of an Agent-Based Social Network Model
3. 学会等名 4th International Workshop on Agent-Based Modelling of Human Behaviour (ABMHuB'22) (国際学会)
4. 発表年 2022年

1. 発表者名 村瀬洋介, Minjae Kim, Seung Ki Baek
2. 発表標題 間接互惠における 3 種のラベルを用いた場合の進化的安定戦略
3. 学会等名 日本数理生物学会 2022
4. 発表年 2022年

1. 発表者名 Yohsuke Murase, Minjae Kim, Seung Ki Baek
2. 発表標題 Successful social norms in indirect reciprocity with ternary reputations
3. 学会等名 Roles of Heterogeneity in Nonequilibrium Collective Dynamics 2022 (RHINO2022) (国際学会)
4. 発表年 2022年

1. 発表者名 Yohsuke Murase, Minjae Kim, Seung Ki Baek
2. 発表標題 Successful social norms in indirect reciprocity with ternary reputations
3. 学会等名 Conference on Complex Systems 2022 (国際学会)
4. 発表年 2022年

1. 発表者名 村瀬洋介
2. 発表標題 間接互恵性における非バイナリ評価モデル:私的評価問題の解決をめざして
3. 学会等名 ゲーム理論ワークショップ 2023 (招待講演)
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
主たる渡航先の主たる海外共同研究者	ヒルベ クリスチャン  (Hilbe Christian)	マックスプランク研究所・Max Planck Research Group・ Group Leader	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
ドイツ	Max Planck Institute			
韓国	Pukyong National University			