なぜ人は協力し続けることができるのか？
－不確実性下の社会性を支える認知基盤の検討[1]－

（2024年１月11日　受理）

佐藤　友美 *[1]

How can People Continue to Cooperate?
-An Examination of the Cognitive Foundations of Sociality under Uncertainty-

（Received January 11, 2024）

Tomomi SATO

Humans are social animals and live in cooperative relationships with others. The purpose of this study was to elucidate the cognitive mechanisms underlying the maintenance of cooperative behavior, which is important for sociality. Previously, the cognitive basis of sociability was considered to be the ability to infer others' mental states correctly. However, since understanding the mental states of others is in principle uncertain, the basis of sociability is thought to be the recognition of uncertainty in one's understanding of the mental states of others. Therefore, in this study, we constructed a new social interaction task that incorporates uncertainty and continuity of others' mental states into the Prisoner's Dilemma and examined participants' cooperative behavior. 10 game experiments were conducted with 66 undergraduate students in which it was difficult to maintain cooperation. We analyzed the transition patterns of cooperation/noncooperation, the degree of liking for the partner, and the positive attitude toward cooperation/noncooperation, all of which were classified into two classes. The results suggest that two factors contribute to the maintenance of cooperative behavior: a positive prediction of cooperation that the partner will choose to cooperate, and the perception of uncertainty that the prediction may not always be true. The results also suggest that the perception of uncertainty about the partner's prediction of cooperative behavior may suppress the occurrence of conflict situations in group work, even when the partner is uncooperative.

問題と目的

ヒトは，他者と関わり合うことをその本質とする，社会的動物である（Aronson, 2018）。ヒトは，他者との協力関係の中に存在している。例えば，「見ず知らずのお年寄りに席を譲る」

[1] 本研究はJSPS科研費 JP20K20153の助成を受けたものです。
*[1] 九州工業大学教養教育院

「グループで団結して作業を行う」など，個体としての顕在的他者に協力する場合もあれば，「組織の不文律に従う」など，集団としての非顕在的他者に与する場合もある。これらは全て他者と協力する能力（社会性）の発現であり，これがあるために，ヒトは「メンバーの強みと弱みを互いに補って課題に取り組む」「先人の生み出したものに改良を加えて後世に伝える（文化の累積的進化）」など，個人の力では到達し得ない域にも到達してきた（Tomasello, 2014）。

では，ヒトの社会性を支える基盤とは如何なるものであろうか。他者と協力するためには，他者の行為や行動からその背後にある意図や感情（心的状態）を推測できなければならない。例えば，ふらつきながら荷物を運ぶ人を見て，「大変な思いをしているはずだ」と推測できなければ，代わりに荷物を持とうと思うことは難しい。同様の傾向は学術研究からも示唆されている。今，男性が自身の所有物を箱Aに入れた後，女性が密かに（男性がいない時に）それを箱Bに移したとする。では，戻ってきた男性はどちらの箱を見るだろうか。正答は，「男性は女性が移したことを知らないため，自身の所有物は箱Aに入っていると思い，箱Aを見る」となる。だが，社会性に障害を持つ自閉症児や社会的スキルの低い健常児などは「女性が移した先の箱Bを見る」と答えやすい（Astington & Jenkins, 1995; Lalonde & Chandler, 1995）。社会性に難のある者ほど，男性（他者）の心的状態を推測しにくいのである。このような知見から，これまで「社会性の認知基盤は他者の心的状態を正しく推測する能力にある」と広く考えられてきた（Bartsch & London, 2000）。

ところが，この想定には大きな問題がある。確かに，他者の心的状態を推測できなければ，その他者と協力することはできない。しかし，現実場面では行動を生み出した原因は数多く存在するため，他者の行動からその心的状態を正確に推測することは原理的に至難となる。先の「戻ってきた男性が箱Aを見た」という例であれば，「自分の所有物が移されたことを知らないため」かもしれないが，「自分の所有物が既に移されたことを確認するため」かもしれない。即ち，他者の行動からその原因となる心的状態を推測する行為は“逆問題”を解く行為に他ならず，「他者の心的状態の正しい推測」なるものは原理的に望めないのである。そうである以上，その能力が社会性の認知基盤になるとは考えにくい。事実，先の「戻ってきた男性がどちらの箱を見るか」という課題成績と，社会性との相関は弱い（Watson et al.,1999）。反対に社会的相互作用が不確実性を伴う以上，「他者の心的状態に関する自分の推測は間違っているかもしれない」と認識する能力こそが，他者の心的状態の柔軟な解釈と対応ひいては社会性に寄与すると考えられる。実際，「男性は自身の所有物が箱Aに無いことを確認すべく箱Aを見た」という場合，「所有物が移されたことを知らないために箱Aを見た」と判断して疑わない者は，その男性との意思疎通に齟齬を来すであろう。

そこで本研究では，他者の心的状態の理解は原理的に不確実であること，それゆえに，

他者の心的状態に関する自身の理解の不確実性を認識することが社会性の基盤となることを明らかにする。

ヒトの社会性の基盤について不確実性の観点から検討した研究は極めて少ないが，例外的研究として Vives & FeldmanHall (2018) がある。この研究は，相手が自分を裏切るかもしれないという不確実な状況下で，それでもなお相手を信じて協力するか否かを検討した。その結果，「相手は協力してくれるはずだ」と楽観的に捉える者ほど相手と協力することが明らかになった。他者への楽観視が社会性の基盤になるということである。

だが，ここで扱われた不確実性は「相手がこの後どう行動するのか（協力か非協力か）が分からない」であり，協力行動も一回限りであった。この場合は確かに，「相手は協力してくれるはず」という楽観視が協力行動につながるであろう。だが「協力してくれるはず」と信じていた相手が協力しなければ，その負の影響は大きい（Aronson & Linder, 1965）。つまり，他者への協力を継続するか否かが問われる場面（例えばグループ作業）では，「相手も協力してくれるはず」という楽観視は後続の協力行動を妨げかねない。寧ろ，「相手の行動の意図（協力か非協力か）は断定しにくい」という不確実性の認識こそ，その後も他者を信じて協力し続ける要因となり，重要になると予想される。言い換えると，社会性の認知基盤は他者との関係性構築の時間軸の中で変化しうるのである。

以上を踏まえて本研究では，不確実性下の継続的な社会的相互作用課題を新たに構築した。具体的には，囚人のジレンマに，新たに他者の心的状態の不確実性と継続性を含んだ状況を導入する。具体的には，３名の参加者がいて，各参加者が他の参加者を裏切って自分だけが利益を得る，またはほかの参加者に協力するために共同貯金に貢献する，という2つの選択肢を提示する。全ての参加者が他の参加者を裏切ると，最終的に得られる資金は最低金額になるが，１名だけが裏切った場合には，最終的に得られる資金が最高金額になるように設定する。また，３名がともに共同貯金に貢献し，最終的な共同貯金が200円以上になった場合には，共同貯金分が2倍になり，それを山分けすることができることとした。ゲームは10回繰り返し，共同貯金に貢献する（協力）か貢献しない（非協力）かを選択させる。選択後に相手の２人が共同貯金に貢献したか否かが知らされるが，それが確実な情報かどうかはわからないという状況を設定し，参加者はその情報を基に次の選択を行うことが求められた。つまり，他の参加者の出方を知ることはできるが，他の参加者の意図が協力なのか非協力なのかは確実性が低いという状況とした。その上で参加者には，他の参加者の意図（協力か非協力か）を答えてもらうことで，他者の心的状態を正しく推測する能力を測る。同時に，その推測がどれほど間違いないと思うかも答えてもらい，心的状態の推測における不確実性を認識する能力を測る。

ここでは，参加者が共同貯金に貢献すれば，参加者の意図は「協力」となる。本研究

では，10回のゲームを通じて協力し続けるためには，他者の心的状態を正しく推測するよりも，その不確実性を認識することが強い規定因子となることを検証する。

方法

実験参加者

合計66名の大学生（男性39名，女性27名）が参加した。参加者の平均年齢は19.333歳で，標準偏差は1.013であった。

材料

遠隔同期システムを使用した。教示は，実験者が表示しているMicrosoft PowerPointによるスライドを画面共有することで行った。

手続き

本研究では，実験者と3名の実験参加者が遠隔同期システムに同時に接続した。参加者間の匿名性を保持するため，システム上で参加者はそれぞれ「A」，「B」，「C」という識別子で表示されるよう設定された。参加者の音声および映像は無効化（ミュート）され，実験は実験者の音声のみで実施された。

実験中，実験者はPowerPointを用いて文字による指示および質問を提示し，これらを口頭で説明した。参加者は，実験者へのダイレクトメッセージを通じてのみ反応するように指示され，他の参加者との直接的なコミュニケーションは行わないよう求められた。

本実験は，３人１組の参加者によって10回の反復試行が行われた。各試行において，参加者は20円を持ち，その20円を共同貯金に貢献するか（"ペイ"）保持するか（"キープ"）を選択させた。この選択は参加者の利益に影響を与え，実際に支払われる謝金額が変動するようになっていた。ペイとキープの選択に応じて，支払い額が次のように定められていることが示された。

10回の試行全てに全員がペイした場合には1350円が支払われ，10回の試行全てに自身だけがキープした場合には最大の1417円が支払われ，10回の試行全てに自身だけがペイした場合には最小の1083円が支払われ，10回の試行全てに全員がキープした場合には1150円が支払われることが示された。さらに，共同貯金が200円以上になった場合，その金額は2倍され，3等分されたうちの1/3が支払われるが，共同貯金が200円未満の場合は基本の謝金のみが支払われることが示された。

ペイかキープかを参加者が選択してプライベートメッセージで実験者に報告した後，参加者には，2人ともがペイを選択したか，1人がキープを選択したか，または2人ともがキープを選択したかのいずれかをプライベートメッセージを通じて報告された。実際の参加者の選択に関わらず，各試行の結果は事前に定められた順序に基づいて報告された。この順序は次のように設定された。1回目：1人がキープ，2回目：1人がキープ，3回目：2人ともキープ，4回目：2人ともキープ，5回目：2人ともキープ，6回目：2人ともキープ，7回目：2人ともキープ，8回目：2人ともキープ，9回目：1人がキープ，10回目：1人がキープ，となっていた。ただし，この報告は「正しい時もあるが誤っている時もある」ことを明示した。この報告は正しいとは限らないとはいえ，自分以外の参加者が非協力的であることを示唆するものとすることで，協力行動を維持することが困難になる状況に設定した。

次に，参加者に他の2人の参加者に対する態度を測るため，他の2人の参加者に対する好意度を1から5のスケールで評価してもらった（1: 好ましくない，2: あまり好ましくない，3: どちらとも言えない，4: やや好ましい，5: 好ましい）。また，自身の選択（ペイ / キープ）が積極的なものであったか，あるいは消極的なものであったかを1から5のスケールで評価させた（1: 消極的，2: やや消極的，3: どちらとも言えない，4: やや積極的，5: 積極的）。

さらに，他の2人の実際の選択が「2人ともペイ」，「1人がキープ」，「2人ともキープ」のどれであったと思うかを尋ね，その推測の正確さ（確率）を50%以上で評価させた。最後に，次のゲームにおいて他の2人がどの選択をすると予測するかを尋ね，その予測の正確さ（確率）も50%以上で評価させた。

実験後，実際の参加者の選択に応じた謝金を参加者に支払った。

分析方法

本研究では，1回目から10回目に至るまでのゲームの推移を三つの側面から検討した。第一に，協力・非協力の行動パターンの推移を明らかにするため，二値のカテゴリカル潜在クラス成長分析（LCGA）を用いて，参加者の協力的行動と非協力的行動の推移パターンを同定した。第二に，相手への好意度のパターンの推移を明らかにするため，LCGAを用いて，各ゲームでの相手への好意度の変化を追跡し，好意度の変動パターンを同定した。第三に，協力・非協力への積極性のパターンの推移を明らかにするため，LCGAを用いて，参加者が示した協力または非協力行動の積極性の変化から，その推移パターンを同定した。

また，協力・非協力の推移パターンと，好意度および積極性の推移パターン間の関連

性をカイ二乗検定および相関係数を用いて検討した。これにより，行動パターンと好意度および積極性の間に統計的に有意な関連が存在するかを評価した。

最期に，協力・非協力の行動推移パターンによって予測の正確さが異なるかをt検定によって分析した。同様にt検定を用いて，協力・非協力の行動推移パターンによって予測と実際の曖昧性の認識が異なるかを分析した。

倫理的配慮

本実験への参加は自由意志であることや，参加による成績などへの影響はないこと，途中で実験を中断できること，取得した情報は研究以外に使用しないこと，疑問等はいつでも質問できることといった倫理的配慮についてPowerPointで提示しながら口頭で説明した。その上で同意する場合には同意する旨をプライベートメッセージにて送ってもらい，実験を開始した。なお，本実験は，九州工業大学教養教育院研究倫理委員会による承認を得た（受付番号28）。

結果

協力・非協力の推移パターン

2値の結果におけるLCGAを行った結果，BICを用いたモデル適合度の比較により，BIC値が868.350の2クラスが3クラス（BIC = 871.744）に比べて優れていることが示された。2クラスモデルのエントロピーは0.807であり，これは各クラスが適切に区分されていることを示唆している。LMRは，2クラスでは有意差は見られなかったが，3クラスでは有意差が見られた（p = .017)。BLRTは，2クラスでも3クラスでも有意差が見られた（ps < .001）。クラス１は１回目から10回目まで協力する選択を維持していたため，協力維持（16名，全サンプルの24.24%）とした。クラス2は1回目から非協力の選択をしており10回目にかけて非協力の選択が増えていることから，非協力増加50名（75.76%）とした。

相手への好意度の推移パターン

LCGAを行った結果，モデル適合度において二次の2クラスモデルはBIC値が1849.809，二次の3クラスモデルはBIC値が1783.907であった。LMRは，2クラスでは有意差が見られたが（p = .036)，3クラスでは有意差が見られなかった。BLRTは，2クラスでも3クラスでも有意差が見られた（ps < .001）。３クラスは１クラスが９名と全サンプルの10％未満のクラスが見られたことから，２クラスとした。クラス１は好意度

が一度下がってからまた上がっていることから，好意U字（36名，全サンプルの54.55%）とした。クラス2は1回目から10回目まで比較的好意度が高かったことから，好意度維持（30名，全サンプルの45.46%）とした。

Table 1
LCGA の適合度指標

		協力・非協力の推移		相手への好意度の推移		協力・非協力への積極性の推移	
		2 classes	3 classes	2 classes	3 classes	2 classes	3 classes
LL		-419.511	-414.924	-889.293	-847.962	-1057.323	-1043.1
No. of parameters		7	10	17	21	17	21
BIC		868.350	871.744	1849.809	1783.907	2185.869	2174.183
SSABIC		846.312	840.262	1796.29	1717.795	2132.35	2108.071
Entropy		0.807	0.682	0.911	0.938	0.908	0.802
Adj. LMR-LRT		24.262	234.761	156.361	78.006	102.889	26.843
p		.388	.017	.036	.216	.101	.508
BLRT		26.192	245.081	165.691	82.66	109.029	28.445
p		.000	.000	.000	.000	.000	.000
Group size	C1	16	15	36	6	16	30
%		24.242	22.727	54.545	9.091	24.242	45.454
	C2	50	9	30	33	50	29
%		75.758	13.636	45.454	50	75.758	43.939
	C3		42		27		7
%			63.636		40.909		10.606

協力・非協力への積極性の推移パターン

LCGAを行った結果，モデル適合度において二次の2クラスモデルはBIC値が2185.856，二次の3クラスモデルはBIC値が2174.183であった。LMRは，2クラスでも3クラスでは有意差が見られなかった。BLRTは，2クラスでも3クラスでも有意差が見られた（*ps* < .001）。3クラスは1クラスが7名と全サンプルの10%のクラスが見られたことから，2クラスとした。クラス1は積極性が一度下がってからまた上がっていることから，積極性U字（16名，全サンプルの24.24%）とした。クラス2は1回目から10回目まで比較的積極性が高かったことから，積極性維持（50名，全サンプルの75.76%）とした。

推移パターンのクラスごとの人数の偏り

Table 2に，協力･非協力の推移パターンのクラスごとの，相手への好意度の推移パターンのクラス，および協力･非協力への積極性の推移パターンのクラスでの人数を示した。協力･非協力の推移パターンのクラスごとの，相手への好意度の推移パターンのクラス，および協力・非協力への積極性の推移パターンのクラスでの人数の偏りは見られなかった（それぞれ$x^2(1) = 0.199$, *n.s.*, $x^2(1) = 2.441$, *n.s.*）

Table 2
協力・非協力の推移パターンのクラスごとの，相手への好意度の推移パターン，および協力・非協力への積極性の推移パターンの人数

		相手への好意度		協力・非協力への積極性	
		好意U字	好意維持	積極性U字	積極性維持
協力・非協力の推移	協力維持	10	6	7	9
	非協力増加	26	24	9	41
合計		36	30	16	50

推測の確信度，予測の確信度，推測の正確性および予測の正確性について相関係数を算出した（Table 3）。その結果，推測の確信度と予測の確信度は正の相関がみられた。推測の正確性と予測の革新性も正の相関がみられた。つまり，確信度が高い人は推測でも予測でも高く，正確性が高い人は推測でも予測でも高いことが明らかになった。

推測の確信度が高い程推測の正確性が高いことも明らかになったが，予測の確信度と予測の正確性の間には有意な相関がみられなかった。予測に関しては，確信度が高いからと言って正確性も高いとはいえないことが明らかになった。

Table 3
推測の確信度，予測の確信度，推測の正確性，予測の正確性についての相関係数

	予測の確信度	推測の正確性	予測の正確性
推測の確信度	.827**	.347**	—
予測の確信度	—	.325**	.204
推測の正確性	—	—	.547**

note: $p < .01$**

Table 4
協力・非協力の行動推移パターン，相手への好意度パターン，協力・非協力への積極性パターンの，クラス別平均値，標準偏差とクラスの比較（*t* 検定）

			mean	*sd*	*t* 値	自由度	*p* 値
協力・非協力の行動推移パターン	協力維持	推測の確信度	0.664	0.078	-1.646	64	.105
	非協力増加		0.703	0.082			
	協力維持	予測の確信度	0.647	0.082	-1.672	64	.099 †
	非協力増加		0.689	0.088			
	協力維持	推測の正確性	0.444	0.285	-2.087	64	.041 *
	非協力増加		0.592	0.235			
	協力維持	予測の正確性	0.285	0.229	-2.164	64	.034 *
	非協力増加		0.433	0.242			
相手への好意度のパターン	好意U字	推測の確信度	0.708	0.072	1.615	64	.111
	好意維持		0.676	0.091			
	好意U字	予測の確信度	0.696	0.076	1.817	64	.074 †
	好意維持		0.658	0.096			
	好意U字	推測の正確性	0.608	0.231	1.869	64	.066 †
	好意維持		0.493	0.269			
	好意U字	予測の正確性	0.410	0.246	0.475	64	.637
	好意維持		0.381	0.249			
協力・非協力への積極性のパターン	積極性U字	推測の確信度	0.694	0.092	0.023	64	.982
	積極性維持		0.693	0.080			
	積極性U字	予測の確信度	0.683	0.100	0.208	64	.836
	積極性維持		0.678	0.084			
	積極性U字	推測の正確性	0.613	0.239	1.023	64	.310
	積極性維持		0.538	0.258			
	積極性U字	予測の正確性	0.507	0.287	2.103	64	.039 *
	積極性維持		0.362	0.223			

note: $p < .10^{\dagger}$, $p < .05^{*}$

推測・予測の確信度と推測・予測の正確性の高さの違い

非協力が増加していく人たちは協力維持している人よりも，予測の確信度，推測の正確性，および予測の正確性が高かった。

相手への好意の高さが一度下がってからまた上がる人たちは相手への好意が一貫して高い人よりも，予測の確信度と推測の正確性が高い傾向があった。

さらに，協力・非協力への積極性の高さが一度下がってからまた上がる人たちは積極性の高さが一貫していた人よりも，予測の正確性が高かった。

考察

本研究の目的は，協力行動の維持をより強く規定する因子が，他者の心的状態を正しく推測する力よりも，その不確実性を認識する力であることを検証することであった。

協力維持が困難な10回のゲームを行った結果，ゲーム開始から一貫して協力的な人と，ゲーム開始から非協力的でその後より非協力傾向が高まる人に分かれた。非協力傾向が高まる人は実験者からの参加者の選択に関する報告をより信じたために応報戦略（e.g., Dawes, 1980）を取った可能性もある。しかし，推測の確信度，つまり相手が何を選択したかに対する自信は協力維持の人と差が見られなかったことから，この可能性は低いと考えられる。そのため，社会的相互作用場面において，特性的に協力維持傾向が高い人と非協力傾向が高まる人に分かれると考えられる。

非協力傾向が強まる人は協力維持の人に比べて，次に相手が何を出すかについての予測において自信があることが示された。そして実際にその予測は非協力傾向が強まる人は協力維持の人に比べて正確であった。つまり，協力維持傾向が高い人は，相手がどのような選択を行うかについてはよく分からないが協力行動をとってくれるのではないかという予測のもと，自分も協力行動をとっている。一方で非協力傾向が強まる人は，相手は非協力行動をとるだろうとある程度確信しているために，自分も非協力行動をとっていると考えられる。つまり，相手が協力を選択してくれるのではないかという協力への肯定的な予測と，しかしその予測は当たるとは限らないという不確実性の認識の二要因が，協力行動の維持に貢献していることが示唆された。

また，相手に対する好意を維持する人は，相手がどのような選択を行うかについてはよく分からないという予測に対する不確実性をより認識している傾向が高かった。相手に対する好意を維持している人は協力維持をするという関係性は見られなかったが，これは相手が協力しても協力しなくても，不確実性を認識していれば相手への好意は維持されると考えられる。したがって，相手の協力行動への予測に対する不確実性の認識は，グループワークなどで相手が非協力的であっても，葛藤状況の発生を抑制できる可能性が示唆された。

以上のことから本研究では，他者の心的状態に関する自身の予測の不確実性が，協力行動の維持や相手への肯定的態度を支える一部となっていることが新たに示唆された。今後，実際にどのような判断によって自身の協力・非協力行動を選択しているのかや，どのような理由で相手への好意を判断しているのかをインタビューデータをもとに質的に検討し，本研究の知見をよりサポートしていく必要があるだろう。

引用文献

Aronson, E. (2018). *The social animal* (12th ed.). New York: Worth.

Aronson, E., & Linder, D. (1965). Gain and loss of esteem as determinants of interpersonal attractiveness. *Journal of experimental social psychology, 1*(2), 156-171. http://dx.doi.org/10.1016/0022-1031(65)90043-0

Astington, J. W., & Jenkins, J. M. (1995). Theory of mind development and social understanding. *Cognition & Emotion, 9*(2-3), 151-165. http://dx.doi.org/10.1080/02699939508409006

Bartsch, K., & London, K. (2000). Children's use of mental state information in selecting persuasive arguments. *Developmental psychology, 36*(3), 352. http://dx.doi.org/10.1037/0012-1649.36.3.352

Dawes, R. M. (1980). Social dilemmas. *Annual review of psychology, 31* (1), 169-193. http://dx.doi.org/10.1146/annurev.ps.31.020180.001125

Lalonde, C. E., & Chandler, M. J. (1995). False belief understanding goes to school: On the social-emotional consequences of coming early or late to a first theory of mind. *Cognition & Emotion, 9*(2-3), 167-185. http://dx.doi.org/10.1080/02699939508409007

Sampaio, W. M. (2024). The uniqueness of human cooperation. *Nature Reviews Psychology,* 1-1. http://dx.doi.org/10.1038/s44159-023-00273-x

Tomasello, M. (2014). The ultra‐social animal. *European journal of social psychology, 44*(3), 187-194. http://dx.doi.org/10.1002/ejsp.2015

Vives, M. L., & FeldmanHall, O. (2018). Tolerance to ambiguous uncertainty predicts prosocial behavior. *Nature communications, 9*(1), 1-9. http://dx.doi.org/10.1038/s41467-018-04631-9

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of personality and social psychology, 76*(5), 820. http://dx.doi.org/10.1037/0022-3514.76.5.820