

# 非言語交渉における認知課題の分割が統合的解決能力の向上に与える影響

佐藤幹晃<sup>1\*</sup> 寺田和憲<sup>1</sup>

ジョナサングラッチ<sup>1</sup>

Motoaki Sato<sup>1</sup> Kazunori Terada<sup>1</sup> Jonathan Gratch<sup>1</sup>

<sup>1</sup> 岐阜大学

<sup>1</sup> Gifu University <sup>2</sup> 南カリフォルニア大学クリエイティブ・テクノロジー研究所

<sup>2</sup> Institute for Creative Technologies University of Southern California

**Abstract:** 交渉において、非言語情報から相手の選好とリミットを推定して統合的解決 (Win-Win) に至ることは、提案に対する満足度と選好の2つを同時に読み取ることに加え、Win-Win な分配を計算するという複雑な認知課題となっており困難である。そこで、非言語交渉を、選好推定のみ集中するフェーズと、提案に対する満足度とリミットの推定に専念するフェーズに分割したところ、Win-Win に至ることが可能と分かった。

## 1 はじめに

交渉において、相手の選好を正確に推定することは Win-Win な結果に至る上で重要である [Thompson 90, Kolodziej 16]。選好についての情報をお互いに交換しなければ、Win-Win な結果に至ることは困難である [Boven 03]。この困難さを説明する例として1つのオレンジを分け合う状況が知られている [Fisher 81]。二人の姉妹がオレンジを半分に分けてわけ、姉は果実を食べて皮を捨てたのに対し、他方で妹はパウンドケーキを作るために皮を用いて果実は捨てた。半分に分け合うことは公平な分配 (分配的交渉) ではあるが、オレンジ全てが利用されたわけではないため最適な分配とは言えない。しかし、彼女達はお互いに異なる選好についての情報を交換し合うことで、捨てることなく全てのオレンジを用いる最適な分配をすることが可能となる。このような分配は、分配的交渉と対比され、お互いにとって Win-Win に資源を分配することは統合的交渉と呼ばれる [Burchill 99]。相手の選好を知ることは重要であるが、交渉では選好を明らかにせず偽ることで相手よりも高い分配を得ることができるため [Gratch 16]、信頼性の低い虚偽の情報を無視しなければならない。明示的な情報交換に消極的であったとしても、交渉は強い感情を個人々に生じさせ、対人関係に影響を与える [Sharma 20, Kleef 18, Kleef 21, Hoegen 22]。交渉では、客観的な成果 (例: 金銭的な利得) だけでなく、個人の内的な経験から基づく主観的価値が感情として

現れるため [Curhan 06]、交渉の過程や結果にどのように反応しているかに注意を払うことで、相手の価値に関する信頼性の高い情報が得られる。そのため、もしそれが偽られた感情表出でないのであれば [Campagna 16]、相手の感情を相手の心的状態の推論に利用することは有用だと思われる [Kleef 04, Pietroni 08]。

典型的な交渉では、お互いに提案し合い、提案に対する反応を読み取り合うことでお互いが合意できる内容を模索する。提案内容は交渉相手の選好を洞察する上での貴重な情報であるが [Baarslag 15]、感情的な反応からその精神状態を理解することは依然として難しい。このようなフィードバックは、彼らの選好 [Pietroni 08, Sato 23] だけでなく、交渉で許容できる最低条件 (リミット) も反映し [Kleef 04, Dijk 08, Lelieveld 12]、精神状態の推測を複雑にする。例えば、「この提案にはちょっと怒りを感じる」といった否定的な感情的反応や、失望した表情は、交渉相手が提案を自分のリミット以下と見なして交渉が決裂する可能性が高いことを示しているのか、それとも提案された内容に不満を感じて別の提案を期待しているのか曖昧であり判断することは困難である。この曖昧さは、ある感情表現された理由が複数考えられることに起因しており、感情的表現から精神状態を推測することは本質的には不良設定問題である。

交渉における感情の重要性を示した研究は膨大にあるが、そのほとんどがリミットの推論に関するものであり、選好の推論とリミットの同時推論の認知計算課題に注目した研究は驚くほど少ない。本研究は、交渉における認知計算課題を定式化し、複雑な認知課題を

\*連絡先: 岐阜大学自然科学技術研究科知能理工学専攻  
〒501-1193 岐阜県岐阜市柳戸 1-1  
E-mail: motoaki.sato@ai.info.gifu-u.ac.jp

段階毎に分割することが、統合的解の実現に寄与することを検証した。

## 1.1 複数論点交渉における認知課題

複数論点交渉における認知課題は、選好の推定、リミットの推定、および Win-Win 計算である。以下では、感情表現を相互に行う非言語交渉の文脈で、選好の推定が、リミット推定と Win-Win 計算より先に行う必要がある重要な課題であることを示す。

### 1.1.1 選好の推定

典型的な複数論点交渉は、提案の内容とその提案に対する評価を相互に交換することで行われる。これらの評価は各提案への反応として現れる。各論点に対する分配の提案をベクトルとして、 $\mathbf{x}^o = [x_1^o, \dots, x_n^o]$  と表記する。ここで  $n$  は交渉の論点の数を示し、 $x_i^o$  は相手  $o$  に分配された  $i$  番目の論点の数を表す。自己 ( $s$ ) および相手 ( $o$ ) が獲得可能な分配の総量は、各論点に対する総量  $\mathbf{x}$  に等しく、 $\mathbf{x}^o + \mathbf{x}^s = \mathbf{x}$  と表現される。自己と相手が獲得可能な全体の効用は、選好の重みベクトル  $\mathbf{w}^o = [w_1^o, \dots, w_n^o]$  と分配ベクトル  $\mathbf{x}^o$  の内積によって決定され、次のように計算される：

$$u^o = \mathbf{w}^o \cdot \mathbf{x}^o \quad (1)$$

異なる論点に対する重みはエージェントによって異なり、これらの論点間に相互作用効果や依存関係がないという状況で、加算的に独立した効用関数を仮定する [Keeney 93]。また、限界効用逓減の原則は適用されない。例えば、オレンジの分配に関する交渉では、提案は果実と皮の量として  $\mathbf{x}^o = [x_{fruit}^o, x_{peel}^o]$  のように構成される。提案  $[x_{fruit}^o, x_{peel}^o] = [0, 1]$  で、0%の果実と100%の皮の場合、ジュースを作りたい人の場合、選好の重み  $[w_{fruit}^o, w_{peel}^o] = [1, 0]$  では効用  $u^o$  は0であるが、ケーキを作りたい人の場合、選好の重み  $[w_{fruit}^o, w_{peel}^o] = [0, 1]$  では効用  $u^o$  は1である。

相手の感情表現は  $u^o$  の評価から生じる。私たちは、効用が増加するにつれて、人々がよりポジティブな感情（例えば喜び）を表現するという比例関係を持つと仮定した： $joy_k \propto u$ （ここで  $k$  は喜びの程度を表す）。相手の課題は、観察された感情表現から推測されるスカラー  $u^o$  から重みベクトル  $\mathbf{w}^o$  を復号することである。これは逆評価として知られるプロセスを通じて行われる。理論的には、 $n$  個の異なる分配を提案し、相手による  $u^o$  の  $n$  個の評価を確認し、 $n$  個の線形方程式を解くことで  $\mathbf{w}$  内の  $n$  個の未知数を計算することができる。しかしこの方法は現実的ではない。より実証的な方法では、分配する他の論点の数を固定したまま単一の論

点の数を変化させることで相手の反応を観察することにより、選好の重み  $\mathbf{w}$  を復号することができる。これはローカルクエリとして知られている [Braziunas 08]。例えば、 $\mathbf{x}^o = [0, 1]$  と  $\mathbf{x}^o = [1, 1]$  の場合、相手の評価  $u^o$  が一定（感情表現の変化なし）であれば、最初の論点は彼らにとって価値がないことを示す。さらに、提案  $\mathbf{x}^o = [0, 0]$  と  $\mathbf{x}^o = [0, 1]$  に対する相手の感情表現が変化する場合、提案する側は相手の効用が第二の論点にのみ依存していることを知る。この方法では、各論点の数の変化ごとに効用の変化を観察する必要があるため、 $2n$  の提案をする必要がある。

### 1.1.2 リミットの推定

リミットは受け入れ可能な効用の下限であり、 $limit = u_l$  と表され、この下では提案が拒否される。基本的に、リミットは交渉中の論点の価値を反映している。したがって、少量の効用を得ることができる場合でも提案を受け入れることは経済的に合理的である。しかし、交渉者は様々な理由で提案を拒否することがある。一つの要因は、不公平または不平等が存在すると感じられる状況を避けようとする傾向である [Fehr 99]。例えば、単一の論点に関する一方向のゼロ和交渉である最後通牒ゲームでは、通常、総額の30%未満の不均等な提案は拒否される [Nowak 00]。この場合、リミットは交渉中の論点の効用だけでなく、不平等の要因も考慮に入れる。したがって、リミットには不平等を反映する  $\alpha$  という用語が含まれる [Fehr 99]、 $limit = u_l + \alpha$ 、または、効用が自己と相手間の利益の比率を反映するように修正される。さらに、回答者がより良い代替案 (Best Alternative to a Negotiated Agreement, BATNA) や、交渉において相手より優位であり、より多くの選択肢を持っている場合、それらを下回る提案は拒否される [Kleef 06, Kleef 07]。

上述のように、リミットは個人の要因や相手を持つ代替案によって影響を受ける個人的な価値である。したがって、交渉者は、低い提案に対する怒りの表現など、相手の行動からそれらを推測する必要がある [Sharma 20, Gratch 21]。しかし、感情表現だけでは正確な効用  $u_l^o$  は明らかにされない。式1で示されるように、効用は提案  $\mathbf{x}$  と内部の選好の重み  $\mathbf{w}$  の関数である。提案は可視化されるが、重みは選好の推定が完了するまで不明である。したがって、相手のリミットを理解するためには、まず相手の選好を特定することが重要である。

### 1.1.3 Win-Win の計算

提案者が相手のリミット  $u_l^o$  を知ると、自身の効用を最大化しつつ、それが相手のリミットを下回らないよ

うに自分の提案を最適化することができる：

$$\arg \max_{\mathbf{x}^s} (\mathbf{w}^s \cdot \mathbf{x}^s) \mid \mathbf{w}^o \cdot \mathbf{x}^o \geq u_i^o, \mathbf{x}^s + \mathbf{x}^o = \mathbf{x}. \quad (2)$$

## 1.2 関連研究

交渉において人が感情情報を用いて選好やリミットをどのように復号し、交渉に用いているか調査した研究はあるが、交渉における心的状態の推論という認知課題に注目した研究はとても少ない。Kleefらは、コンピュータを相手とした6回の3論点の携帯電話の売り手と買い手の交渉において、実験参加者の1回、3回、5回提案に対して、相手のエージェントが、怒り、喜びの感情を言語で伝達する場合に、実験参加者の要求水準がどのように変化するかを調査した [Kleef 04]。実験の結果、実験参加者は怒りを表出する相手には譲歩し、喜びを表出する相手には搾取することをが分かった。この実験では、相手のリミットや選好に関係なく感情が表出されていたために、実験参加者は具体的な選好やリミットを推論することは不可能であった。それに対して、Pietroniらは、2論点の4回の交渉（売り手と買い手および福利厚生について交渉）において、エージェントの選好を反映した感情（文章もしくは表情）を提示するタスクにおいて、実験参加者の固定バイ・バイアス（自身と相手と同じ選好を持っていると決めつけてしまう認知バイアス）の程度とエージェントに対する要求を計測した [Pietroni 08]。その結果、感情ありの場合には固定バイ・バイアスが減少し、より統合的解に至ることが確認された、このことは、実験参加者が感情から相手の選好を読み取りができることを示しているが、文章の感情的フィードバックは「モニターの提案には満足しているが、ハードディスクの提案には本当に腹が立つ」のように、各論点それぞれに対して独立に評価が与えられるものであり、利得の総量ではなく各選好に対して直接反応したものであった。また、表情は提案そのものに対する評価フィードバックではなく、各論点独立に与えるものであり、効用  $u$  を符号化した表情から選好ベクトル  $\mathbf{w}^o$  をローカルクエリや連立方程式の計算によって明らかにする認知課題ではなかった。

Satoらは、選好とリミットの両方を推測する必要がある非言語交渉タスクで、参加者がある程度の選好の推定を行う能力があることを実証した [Sato 23]。このゲームは2者間の交渉で、提案者である人と応答者であるエージェントが複数論点（アイテム）を分配する一回限りの交渉（複数論点最後通牒）であった。交渉中に相手の選好を明らかにするためには、参加者はスライダバーを操作して提案する内容を変更することで、ローカルクエリを行う必要があった。エージェントは参加者の提案に対する効用に応じて表情を提示し

た。また、エージェントの表情はリミット  $u_i$  も符号化していた。これは、Pietroniらの研究と異なり、効用の合計値  $u$  のみを観察し、 $w$  を復元するタスクであった。彼らの実験の参加者は、Win-Win 解に到達する代わりに、自身の分配を減らす傾向があった。この原因として、交渉中に、選好とリミットの推論と Win-Win の計算を同時に行うことが認知タスクとして困難であった可能性が示唆される。実際の実験参加者の選好推定の精度は 0.32（真の選好との順位相関係数）であり、実験参加者が拒否されない提案の探索を重視し、ローカルクエリによる選好推定を十分行わなかった可能性がある。

## 1.3 本研究

本研究では、交渉前に感情表現から AI エージェントの選好を学習することで、統合的な解決策を導くことができるという仮説を検証した。参加者は提案者として、AI エージェントと4論点の最後通牒ゲームを行った。参加者は、分配の提案を決定する前に、1回もしくは2回の異なる選好学習を行った。すべての参加者がエージェントの選好とリミットをスライダバーで同時に推論するタスクを行ったが、選好学習フェーズ有条件の参加者だけが、その前に、エージェントに交渉する各論点を見せて、それぞれの論点に対する感情表現を表情から観察する機会を与えられた。

## 2 実験方法

### 2.1 実験参加者

実験参加者は Yahoo! クラウドソーシングで募集した、19歳から70歳までの男性119人、女性28人であった ( $M_{age} = 46.31, SD_{age} = 11.092$ )。IDの不一致とアテンションチェック該当者を取り除き、147人（選好学習フェーズ有:75人、無:参加者72人）をデータとして集計した。

### 2.2 手順

参加者は Yahoo!クラウドソーシングから Qualtrics で作成したアンケートページに進み、年齢と性別を答えたのち、複数論点最後通牒ゲームの説明を受け、理解を確認する問題に答えた。次に参加者は16種類の食べ物に1から4(1:嫌い, 2:どちらでもない, 3:まあまあ好き, 4:とても好き)の水準で評価した。この4段階に少なくとも1つ以上の食べ物が割り当てられるよう、選好がばらけるような食べ物を事前に選定した。加えて、選好が極端に偏る参加者を考慮し、各水準に対し



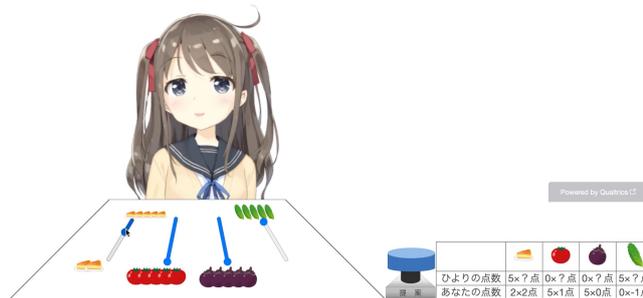
(a) アイテムごとに表出するエージェントの表情 ( $w^{\text{エージェント}} = -1$  (嫌い),  $w^{\text{エージェント}} = 0$  (どちらでもない),  $w^{\text{エージェント}} = 1$  (まあまあ好き),  $w^{\text{エージェント}} = 2$  (とても好き)).



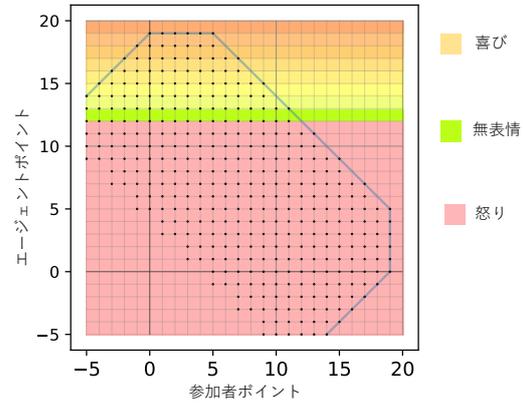
(b) 選好学習フェーズ

| アイテム (個数)           | A (7) | B (5) | C (5) | D (5) |
|---------------------|-------|-------|-------|-------|
| $w^{\text{参加者}}$    | 2     | 1     | 0     | -1    |
| $w^{\text{エージェント}}$ | 2     | 0     | -1    | 1     |

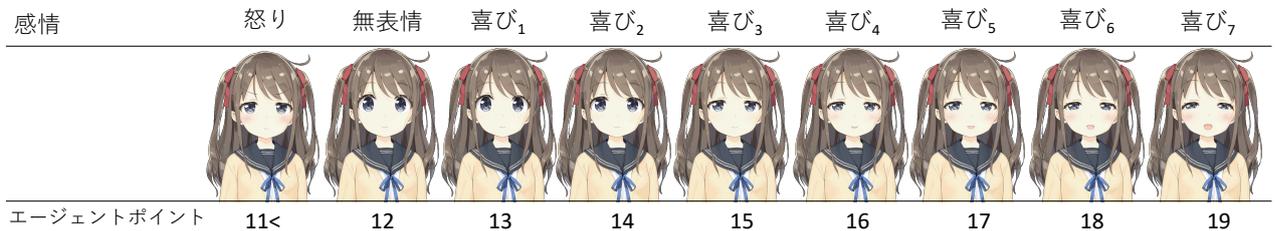
(d) アイテムの重み



(e) 複数論点最後通牒ゲーム



(c) エージェントが分配をどのように評価するかと複数論点最後通牒ゲームにおけるパレート最適を示した評価理論のヒートマップ



(f) エージェントの利得に応じてエージェントの表情が変化する

図 1: 方法: 選好学習のインターフェース

て、少なくとも1つ以上の食べ物を割り当てるように強制した。各水準からランダムに1つの食べ物を選択し、4つのアイテムを後の課題に用いた。選好学習を行う前にエージェントの選好をどのように推測したかを評価するために、参加者はエージェントの選好を推定し、1から4まで順位をつけた。

次に、全ての参加者は複数論点最後通牒ゲームを行った。参加者はスライダーバーを用いて、アイテムの分配をエージェントに提示することができた。表1dのような重みの設定であり、式(1)で示されるように分配されたアイテムの重みの合計がエージェントのポイントとなった。エージェントはポイントが高いほどより喜んだ表情を表出した。エージェントの喜びの表情は、

喜びの段階を適切に表しているか妥当性を検証した8段階を用いた(図1f)。エージェントはリミットを下回ると怒った表情を表出した。参加者がリミット以上で分配を決定すると、参加者とエージェントともに分配に応じたポイントを獲得した。参加者がリミット未満で分配を決定すると、エージェントに分配を拒否されて両者ともにポイントを獲得できなかった。図1cのヒートマップはエージェントのポイントと表情の対応関係を示している。リミットはSatoらに倣って、パレート最適でエージェント参加者間で均等な分配となる12ポイントとした[Sato 22]。参加者の課題は、エージェントに拒否されることなく参加者のポイントを増やすことであった。そのために、参加者はエージェントの選

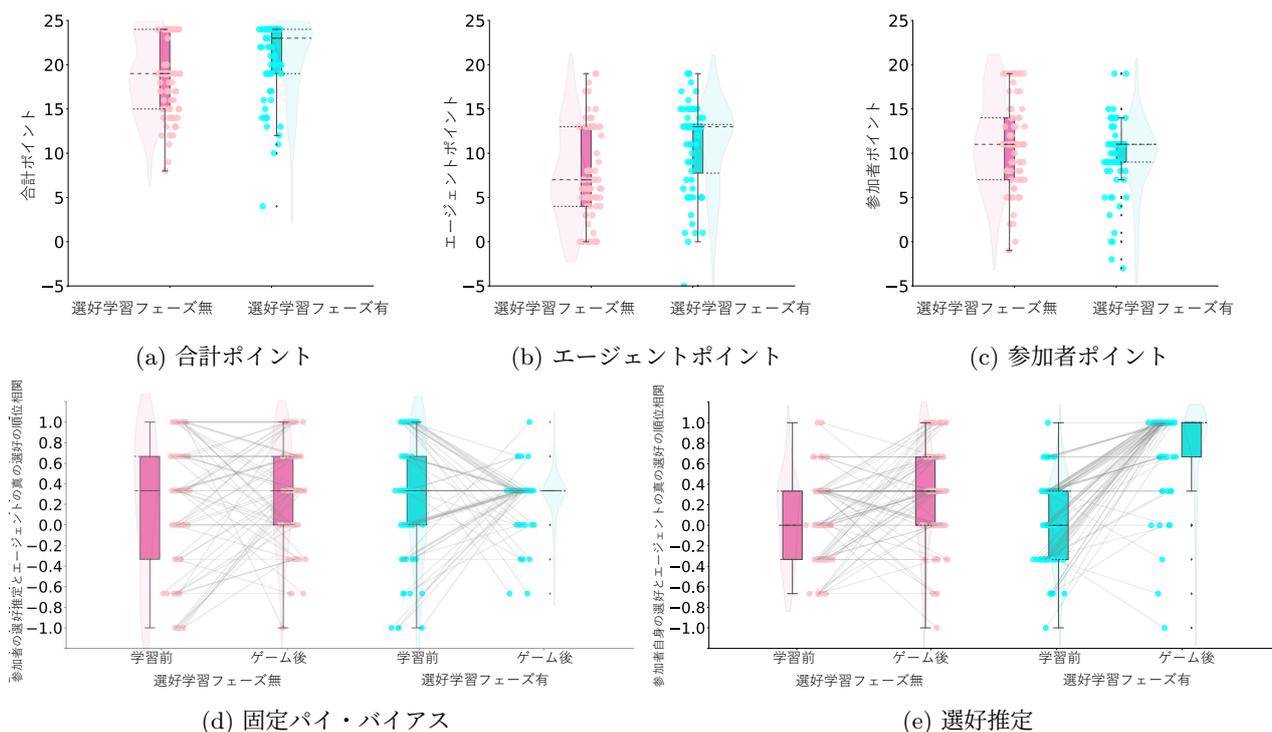


図 2: 実験結果

好とエージェントのリミットを推論する必要があった。図 1e のインタフェースを用いることで、エージェントの選好とリミットの同時推論が理論的に可能である。

しかし、前述の通り選好とリミットを同時に推論することは難しいため、選好学習フェーズ有条件の参加者のみに、交渉前に選好学習を行う機会を与えた。選好学習フェーズでは、選好を符号化しているエージェントの表情を観察することでエージェントの選好を学習した (図 1b)。エージェントの選好は  $w^{\text{エージェント}} = [-1, 0, 1, 2]$ 。エージェントが嫌いな問題には -1 が、エージェントがとても好きな問題には 2 が割り当てられている。  $w^{\text{エージェント}}$  と表情の関係を図 1a に示す。最後に全ての参加者は再度エージェントの選好推定を行った。

交渉結果に影響を与える可能性のある重要な特徴の 1 つは、エージェントの見た目のリアルさや単純さである。寺田らは、単純な線画からなるソフトウェア・エージェントを用いて、最後通牒ゲームにおいて口の角度が上下するにつれて人々の向社会的行動が変化することを示した [Terada 17]。deMelo らは、実際の人に似たエージェントを用い、囚人のジレンマにおいて、エージェントが競争的な感情表現と比較して協力的な感情表現を示すと、人々はより協力的になることを確認した [Melo 14]。そこで本研究では、単純な見た目と実際の見目の中間であるアニメ風のキャラクターをエージェントとして用いた。

## 2.3 報酬

参加者には、参加費として PayPay ポイントを 200 ポイント支払った。さらに、複数論点最後通牒ゲームで獲得した参加者自身のポイントに応じて、より多くのお金を獲得する機会を与えた。今回の実験では 150 人中上位 5%、つまり 7 人の参加者に Amazon ギフト券 3000 円分を与えた。

本研究は、岐阜大学大学院医学系研究科の医学審査委員会 (IRB) の承認を得ている。参加者には、データセットに参加者の情報が使用されることについて、インフォームド Consent を行った。

## 2.4 測定と分析

複数論点最後通牒ゲームにおいて、エージェントが獲得したポイント、参加者が獲得したポイント、両者の合計ポイントを検証した。また、参加者が推定したエージェントの選好順位を測定し、[Johnson 20] に倣い、参加者が推定したエージェントの選好と参加者の選好 (固定バイ・バイアス)、参加者が推定したエージェントの選好とエージェントの実際の選好 (選好推定) の間の Kendall の順位相関係数をそれぞれ算出した。

## 2.5 分析

実験計画は2(選好学習フェーズ: 有/無)の参加者間要因配置で行った。エージェントが獲得したポイント(エージェントポイント), 参加者が獲得したポイント(参加者ポイント), 両者が獲得したポイント(合計ポイント)について, 2水準(選好学習フェーズ: 有/無)の一元配置分散分析を行なった。参加者は, 複数論点最後通牒ゲームの前と後の2回, エージェントの選好を推定したので, 固定パイ・バイアスと選好推定について, 最初・最後の4アイテムの複数論点最後通牒ゲームのインタラクション要因(学習前/ゲーム後水準: 参加者内)×選好学習フェーズの有無の外見要因(選好学習フェーズ: 有/無)の2要因混合計画で行った。

## 3 実験結果

合計ポイントについて, 一元配置分散分析の結果, 選好学習フェーズ有の条件 ( $M = 20.6, SD = 4.3$ )の方が, 選好学習フェーズ無の条件 ( $M = 18.5, SD = 4.3$ )よりも有意に高いことがわかった ( $F(1, 146) = 9.8, p = .002, \eta_p^2 = .063$  (図 2a))。この結果は, 選好学習フェーズが, エージェントと参加者にとって Win-Win な分配をすることに寄与することを示す。エージェントポイントについて, 一元配置分散分析の結果, 選好学習フェーズ有の条件 ( $M = 11.0, SD = 5.0$ )の方が, 選好学習フェーズ無の条件 ( $M = 7.5, SD = 5.2$ )よりも有意に高いことがわかった ( $F(1, 145) = 17.3, p = .000, \eta_p^2 = .105$  (図 2b))。この結果は, 選好学習フェーズにより, 参加者がエージェントに譲歩するようになったことを示している。参加者ポイントについて, 一元配置分散分析の結果, 選好学習フェーズ有の条件 ( $M = 9.6, SD = 4.1$ )の方が, 選好学習フェーズ無の条件 ( $M = 10.9, SD = 5.1$ )に有意な差はなかった ( $F(1, 145) = 2.6, p = .107, \eta_p^2 = .018$  (図 2c))。エージェントポイントの結果も踏まえるとこの結果は, エージェントの選好を正しく推定することで, 参加者は自身のポイントを下げることなく, エージェントのポイントを上げることを示している。

固定パイ・バイアスについて, 繰り返しありの二要因分散分析を行った結果, 学習前/ゲーム後水準に主効果はなく, 選好学習フェーズの有/無の主効果もなく ( $F(1, 145) = 1.64, p = .202, \eta_p^2 = .011$ ), 交互作用も確認されなかった ( $F(1, 145) = 0.922, p = .339, \eta_p^2 = .006$  (図 2d))。このことから, 固定パイ・バイアスは選好学習や学習前後に影響されないことがわかった。選好推定について, 繰り返しありの二要因分散分析を行った結果, 学習前/ゲーム後水準に主効果が認められた ( $F(1, 145) = 105.05, p < .001, \eta_p^2 = .420$ )。Bonferroniの方法による多重比較を行った結果, 参加者が推定したエージェントの選好は, 学習前 ( $M = .54, SD = .51$ )

と比較して, ゲーム後 ( $M = .05, SD = .42$ )の方が有意に高かった ( $p < .001$ )。加えて, 学習前/ゲーム後と有/無の交互作用が確認された ( $F(1, 145) = 105.05, p < .001, \eta_p^2 = .420$ )。Bonferroniの方法による多重比較を行った結果, 選好学習フェーズ有の参加者は ( $M = .79, SD = .42$ ), 無の参加者 ( $M = .31, SD = .47$ )と比較して, ゲーム後にエージェントの選好をより良く推定した ( $F(1, 145) = 41.56, p < .001, \eta_p^2 = .223$  (図 2e))。このことは, 選好学習フェーズにより, 参加者がエージェントの選好をより良く推定できるようになったことを示している。

## 4 議論

本研究では, 選好学習により, 交渉において感情表現から相手の心理状態を推測する能力が向上し, Win-Win な解決策を導くことができるかを検討した。その結果, 選好学習フェーズ有の参加者は, 無の参加者に比べて, 相手の選好をより良く推定し, エージェントと参加者の合計ポイントが増加することが示された。これは, 選好推定が正確になることで, より Win-Win な分配をすることにつながったことを示している。

我々の研究は, 感情表現が固定パイ・バイアスを克服するという既存の研究と一致しなかった。しかし, タスクの種類, 感情表現の仕方, 固定パイ・バイアスの尺度に違いがあったことに留意すべきである。以前の研究は, 2 論点の売り手と買い手の交渉タスクにおいて, 文章と画像を介して限られた回数感情を伝えたのに対し [Pietroni 08], 我々は AI エージェントを用いた 4 論点の最後通牒ゲームにおいて, リアルタイムで表情を伝えた。固定パイ・バイアスは軽減されなかったが, 相手の選好を推定する精度は改善された。エージェントの選好のより良い推定が固定パイ・バイアスの軽減を必ずしも意味するわけではない。参加者の選好と相手の選好のランク相関係数は完全に負の相関ではなく, 部分的に独立している。Kleef らは怒りを出す相手に譲歩し, 喜びを出す相手から搾取することを示したが [Kleef 04], 交渉の前に相手の選好を感情表現から推論することで, 喜びと怒りの境界である, 統合的解決かつ均等な分配を探索することができるようになった。Sato らは AI エージェントと 4 論点の複数論点最後通牒ゲームにおいて参加者自身の点数を減らしてエージェントのポイントを高くする譲歩を引き出すことを示したが [Sato 23], 本研究では参加者自身の点数を減らすことなく, エージェントのポイントを高くすることができた。4 論点の分配をする前に, 文章や単純化したタスクを複数回行いトレーニングするより, 選好を表情から学習することが人と AI エージェントが Win-Win に至ることに寄与すると考えられる。

Win-Win な分配の内訳について、エージェントのポイントは選好推定により上がったが、参加者のポイントは変わらなかった。協力主義者は個人主義者よりも Win-Win な分配に容易に到達可能である [Carnevale 86]。そのため、現在の自分のポイントにしか興味がない個人主義から協力主義なエージェントにデザインすることで、参加者自身のポイントも上げることができ、より Win-Win な解決策を導き出すことが容易になる可能性がある。

## 謝辞

本研究は JSPS 科研費 21H03782 および JST, CREST (JPMJCR21D4), 未来社会創造事業 (JPMJMI22J3) の支援を受けたものである。

## 参考文献

- [Baarslag 15] Baarslag, T., Hendriks, M. J. C., Hindriks, K. V., and Jonker, C. M.: Learning about the opponent in automated bilateral negotiation: a comprehensive survey of opponent modeling techniques, *Autonomous Agents and Multi-Agent Systems*, Vol. 30, No. 5, pp. 849–898 (2015)
- [Boven 03] Boven, L. V. and Thompson, L.: A Look into the Mind of the Negotiator: Mental Models in Negotiation, *Group Processes & Intergroup Relations*, Vol. 6, No. 4, pp. 387–404 (2003)
- [Braziunas 08] Braziunas, D. and Boutillier, C.: Elicitation of Factored Utilities, *AI Magazine*, Vol. 29, No. 4, pp. 79–92 (2008)
- [Burchill 99] Burchill, F.: Walton and McKersie, *A Behavioral Theory of Labor Negotiations* (1965), *Historical Studies in Industrial Relations*, No. 8, pp. 137–168 (1999)
- [Campagna 16] Campagna, R. L., Mislin, A. A., Kong, D. T., and Bottom, W. P.: Strategic consequences of emotional misrepresentation in negotiation: The blowback effect., *Journal of Applied Psychology*, Vol. 101, No. 5, pp. 605–624 (2016)
- [Carnevale 86] Carnevale, P. J. and Isen, A. M.: The influence of positive affect and visual access on the discovery of integrative solutions in bilateral negotiation, *Organizational Behavior and Human Decision Processes*, Vol. 37, No. 1, pp. 1–13 (1986)
- [Curhan 06] Curhan, J. R., Elfenbein, H. A., and Xu, H.: What Do People Value When They Negotiate? Mapping the Domain of Subjective Value in Negotiation, *Conflict & Dispute Resolution* (2006)
- [Dijk 08] Dijk, van E., Kleef, van G. A., Steinel, W., and Beest, van I.: A Social Functional Approach to Emotions in Bargaining: When Communicating Anger Pays and When It Backfires, *Journal of Personality and Social Psychology*, Vol. 94, No. 4, pp. 600–614 (2008)
- [Fehr 99] Fehr, E. and Schmidt, K. M.: A Theory of Fairness, Competition, and Cooperation, *The Quarterly Journal of Economics*, Vol. 114, No. 3, pp. 817–868 (1999)
- [Fisher 81] Fisher, R., Ury, W., and Patton, B.: Getting to Yes: Negotiating Agreement Without Giving in (1981)
- [Gratch 16] Gratch, J., Nazari, Z., and Johnson, E.: The Misrepresentation Game: How to Win at Negotiation While Seeming Like a Nice Guy, in *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems, AAMAS '16*, pp. 728–737, Richland, SC (2016), International Foundation for Autonomous Agents and Multiagent Systems
- [Gratch 21] Gratch, J.: The Promise and Peril of Automated Negotiators, *Negotiation Journal*, Vol. 37, No. 1, pp. 13–34 (2021)
- [Hoegen 22] Hoegen, J., DeVault, D., and Gratch, J.: Exploring the Function of Expressions in Negotiation: the DyNego-WOZ Corpus, *IEEE Transactions on Affective Computing*, pp. 1–12 (2022)
- [Johnson 20] Johnson, E. and Gratch, J.: The Impact of Implicit Information Exchange in Human-agent Negotiations, in *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, ACM (2020)
- [Keeney 93] Keeney, R. L. and Raiffa, H.: *Decisions with multiple objectives: preferences and value trade-offs*, Cambridge University Press (1993)
- [Kleef 04] Kleef, van G. A., Dreu, C. K. W. D., and Manstead, A. S. R.: The interpersonal effects of anger and happiness in negotiations, *Journal of Personality and Social Psychology*, Vol. 86, No. 1, pp. 57–76 (2004)

- [Kleef 06] Kleef, G. A. V., Dreu, C. K. D., Pietroni, D., and Manstead, A. S. R.: Power and emotion in negotiation: power moderates the interpersonal effects of anger and happiness on concession making, *European Journal of Social Psychology*, Vol. 36, No. 4, pp. 557–581 (2006)
- [Kleef 07] Kleef, van G. A. and Côté, S.: Expressing anger in conflict: When it helps and when it hurts, *Journal of Applied Psychology*, Vol. 92, No. 6, pp. 1557–1569 (2007)
- [Kleef 18] Kleef, van G. A. and Côté, S.: Emotional Dynamics in Conflict and Negotiation: Individual, Dyadic, and Group Processes, *Annual Review of Organizational Psychology and Organizational Behavior*, Vol. 5, No. 1, pp. 437–464 (2018)
- [Kleef 21] Kleef, van G. A. and Côté, S.: The Social Effects of Emotions, *Annual Review of Psychology*, Vol. 73, No. 1 (2021)
- [Kolodziej 16] Kolodziej, R., Hesse, F. W., and Engelmann, T.: Improving negotiations with bar charts: The advantages of priority awareness, *Computers in Human Behavior*, Vol. 60, pp. 351–360 (2016)
- [Lelieveld 12] Lelieveld, G.-J., Dijk, E. V., Beest, I. V., and Kleef, G. A. V.: Why Anger and Disappointment Affect Other’s Bargaining Behavior Differently, *Personality and Social Psychology Bulletin*, Vol. 38, No. 9, pp. 1209–1221 (2012)
- [Melo 14] Melo, de C. M., Carnevale, P. J., Read, S. J., and Gratch, J.: Reading people’s minds from emotion expressions in interdependent decision making, *Journal of Personality and Social Psychology*, Vol. 106, No. 1, pp. 73–88 (2014)
- [Nowak 00] Nowak, M. A., Page, K. M., and Sigmund, K.: Fairness Versus Reason in the Ultimatum Game, *Science*, Vol. 289, No. 5485, pp. 1773–1775 (2000)
- [Pietroni 08] Pietroni, D., Kleef, G. A. V., Dreu, C. K. D., and Pagliaro, S.: Emotions as strategic information: Effects of other’s emotional expressions on fixed-pie perception, demands, and integrative behavior in negotiation, *Journal of Experimental Social Psychology*, Vol. 44, No. 6, pp. 1444–1454 (2008)
- [Sato 22] Sato, M. and Terada, K.: The effect of visualizing other’s appraisal process on the cognitive ability to reach win-win outcomes in a multi-issue ultimatum game, in *Conference of the International Society for Research on Emotion (ISRE2022)* (2022)
- [Sato 23] Sato, M., Terada, K., and Gratch, J.: Teaching Reverse Appraisal to Improve Negotiation Skills, pp. 1–14 (2023)
- [Sharma 20] Sharma, S., Elfenbein, H. A., Sinha, R., and Bottom, W. P.: The Effects of Emotional Expressions in Negotiation: A Meta-Analysis and Future Directions for Research, *Human Performance*, Vol. 33, No. 4, pp. 331–353 (2020)
- [Terada 17] Terada, K. and Takeuchi, C.: Emotional Expression in Simple Line Drawings of a Robot’s Face Leads to Higher Offers in the Ultimatum Game, *Frontiers in Psychology*, Vol. 8, pp. 1–9 (2017)
- [Thompson 90] Thompson, L. and Hastie, R.: Social perception in negotiation, *Organizational Behavior and Human Decision Processes*, Vol. 47, No. 1, pp. 98–123 (1990)