

多サイト経済実験はどのようにすれば可能か： 被験者の属性情報と認知能力スコア

小川一仁・尾崎祐介・川村哲也・高橋広雅・田口聡志

藤井陽一郎・渡邊直樹



文部科学大臣認定 共同利用・共同研究拠点

関西大学ソシオネットワーク戦略研究機構

Research Institute for Socionetwork Strategies,
Kansai University

Joint Usage / Research Center, MEXT, Japan

Suita, Osaka, 564-8680, Japan

URL: <http://www.kansai-u.ac.jp/riss/index.html>

e-mail: riss@ml.kandai.jp

tel. 06-6368-1228

fax. 06-6330-3304

多サイト経済実験はどのようにすれば可能か：

被験者の属性情報と認知能力スコア*

Exploring the Possibility of Multi-Sited Economic Experiments:

Subjects' Cognitive Ability and Attribute Information

小川一仁[†] 尾崎祐介[‡] 川村哲也[§] 高橋広雅[¶] 田口聡志^{||}

藤井陽一郎^{**} 渡邊直樹^{††}

2020年3月28日

概要

本稿の著者たちは Guerci et al. (2017) の重み付き投票を背景に持つ二腕バンディット (二者択一問題) 実験の追試を実施し、各実験サイトにおける被験者の正答率を比較することで、多サイト経済実験を可能にするには被験者の属性情報と認知能力スコアによる共変量調整が有効となる場合があることを明確にした。主要な結論は次のとおりである。同じ地域にある実験サイトでは、通常の被験者管理において収集する被験者の一般的属性情報 (年齢、性別、所属学部など) に加えて、彼らの認知能力スコアを収集し、それらを使って共変量調整すると、被験者行動が均質化されることを確認した。一方、実験サイトが別々の地域にある場合、被験者の一般的属性情報と認知能力スコアだけでは彼らの行動の均質性を担保できなかった。また、被験者による意味のある学習 (Rick and Weber, 2010) が観察されるのは被験者の認知能力スコアが高い実験サイトであることも判った。

キーワード: 多サイト経済実験、認知能力、属性情報、重み付き投票、バンディット実験

JEL 分類番号: C91, D72, D83

*本研究は日本学術振興会科学研究費挑戦的研究 (萌芽) (科研費番号 17K18573) と 2017 年度関西大学若手研究者育成経費 (研究課題「多地点実験環境の構築とそれをを用いた経済実験の実施」) より研究費を受け、関西大学 RISS の支援を得てなされた。秋山英三氏、石川竜一郎氏、竹内幹氏、西村直子氏、花木伸行氏、濱口泰代氏、肥前洋一氏には本研究についてコメントをいただいた。ここに記して、感謝する。

[†] 関西大学社会学部 Email: kz-ogawa@kansai-u.ac.jp

[‡] 早稲田大学商学部 E-mail: osakiy@waseda.jp

[§] 日本経済大学経営学部 E-mail: t-kawamura@tk.jue.ac.jp

[¶] 広島市立大学国際学部 E-mail: htaka@hiroshima-cu.ac.jp

^{||} 同志社大学商学部 E-mail: staguchi@mail.doshisha.ac.jp

^{**} 明治大学商学部 E-mail: fujii@meiji.ac.jp

^{††} 慶應義塾大学大学院経営管理研究科、223-8526 神奈川県横浜市港北区日吉 4-1-1、Phone: +81-45-564-2039、E-mail: naoki50@keio.jp

1 はじめに

1.1 本研究の背景

実験結果の頑健性と一般性を保証するには、複数の実験サイトで同じ結果が得られるかを確かめること（外的妥当性の検証）が重要である。しかし、異なる実験サイトでは被験者の属性や特徴は異なるので、サイト間での実験結果の比較には慎重にならねばならないだろう。一方、最近では、論文1本あたりの被験者数は顕著な増加傾向にあり、1つのサイトでは十分な数の被験者を集めることが困難になってきているため、複数の実験サイトで被験者の属性や特徴を共通の基準で制御することで十分な数の被験者を確保しようという動きもある。本稿はこのように様々な動機を背景とする多サイト経済実験を実施するための方法論的基盤整備を目標とする研究の一環として書かれた¹。

この20年ほどの間に、経済実験の成果に対する研究者の認知は顕著に進んだ。それと共に、実験実施環境と実験方法論の整備も進められた²。これらのことを背景として、1つの実験研究に必要な被験者の数は、近年、顕著な増加傾向にある。実験室実験への参加者に限れば、評価の定まった国際専門誌に掲載された実験研究に関する論文1本あたりの被験者数は、2000年ごろには100人足らずであることも少なくなかったが、2015年には300人以上の被験者を集めた実験の報告が多くなり、最近では500人を越える被験者から収集したデータをまとめた論文も少なくない³。

上述のように、複数のサイトで同じ実験を実施し、同様の結果を確認することで、実験結果の頑健性と一般性を保証することに対する必要の高まりが被験者数増加の背後にあるのだろう。同じサイトで何度か同じ実験を実施しても、被験者の属性を含む実験実施状況を完全に制御することは難しく、セッションごとにやや異なった結果が出ることもある。被験者数が多い場合には実験結果に関わる様々な要因が平準化されることもあり、このことも論文1本あたりの被験者数の増加に繋がっているのかもしれない。実験内容も多様化し、その目的によっては被験者の意思決定の背景にあるモデルの関数形を特定できないが、その場合にはノ

¹この方面での先駆的な研究に Roth et al. (1991) や Cason, et al. (2002) などがある。ここでは実験結果の国際比較を行うために留意すべき注意事項がノートされている。

²この時期には、ちょうど、経済実験用ソフトウェアである zTree (Fischbacher, 2007) などの普及により、プログラミングの知識があまりなくても、多くの研究者が経済実験をより容易に実施できるようになり、また、ORSEE (Greiner, 2003) 等の被験者管理支援ツールの開発により、被験者募集をより効率的に行えるようになったため、多くの被験者に実験に参加してもらえる体制が整えられてきた。それと同時に、経済実験方法論の整備も進められ、Journal of Economic Behavior and Organization ではそれに関する特集号 (Rosser and Eckel, eds., 2010) が編集された。この特集号では、たとえば、被験者の実験サイトにおけるランダムな抽出を妨げるような情報を実験関係者（実験者、被験者、実験補助者など）が漏らさないようする必要が説かれている。最近では、理論研究者も、自らの理論の実験による検証を厳密に行うため、実験方法論に関する論文を書いている。たとえば、Azrieli et al. (2018) は被験者への謝金ルールと実験目的の整合性を論じている。

³この点については、2013年に開催された京都大学経済実験室開室記念カンファレンスでの竹内幹氏の講演においても言及された。本研究の出発点は竹内氏によるこの講演にある。

ンパラメトリック検定が用いられる。しかし、ノンパラメトリック検定で明確な実験結果を得るには十分な数のサンプルが必要である。このことも論文1本あたりの被験者数の増加をもたらした一因なのだろう⁴。

論文1本あたりの被験者数増加の結果として、研究者は実験内容に応じて被験者に課された制約のために1つの実験サイトでは研究課題の遂行に必要な数の被験者を集めることが難しくなっている。実験室実験の被験者は実験者とその共同研究者が所属する大学の学生であることが多いが、日本国内での状況を例として取り上げると、大規模私立大学に比べて学生数の少ない国公立大学に所属する実験者にとっては、研究のみならず、学生の指導においても十分な数の被験者を集められなくなっている。このような場合、複数のサイトで同じ仕様の実験を実施し、そこで収集したデータを統合して、データ解析の結果を報告することになる。実際、そのような論文の数は増えてきており、そこには、何らかの指標を参照し、それに関する基準を設けるなど、複数のサイトで収集した実験データを統合 (merge) することを正当化するための説明が書かれてはいるが、その統計学的吟味は十分ではない⁵。このように、実験結果頑健性を保証することを目的としない場合でも、複数のサイトで実験を実施することが増えている。しかし、実施した実験で収集したデータの分析方法について、実験経済学における統一的基準は定まっていない。

データ融合 (data fusion) の方法として、カーネルマッチング (Heckman et al., 1998) は最も簡便なものの一つであるが、それは人工的にデータの重み付けを行うことによりデータを融合する。複数の実験サイトにおける被験者の属性や特徴が異なれば、実験データに欠測値があることになり、そのような方法の利用も有用だろう。重要なことは、どのような指標を参照すれば、複数のサイトで収集した実験データを分析し、より偏りのない推論結果を導出できるかという点にある。カーネルマッチングを用いるにしても、その重みの算出には被験者の意思決定に関わる重要な指標を探し出すことが必要となる。

1.2 概要と結果

本研究では、被験者の認知能力をレーヴンテストを用いて計測し、被験者の一般的属性情報を含めて、複数の大学における実験結果に違いをもたらす要因を考察する。大学生の被験

⁴並べ替え (permutation) の技法を用いた各種検定は、サンプル数が少ない場合でも、正確な p 値を算出することができるといわれているが、20年前ごろまでは大型計算機でしか利用できなかった。最近では、高性能計算機の使用が格段に容易になったこともあり、実験データの解析に並べ替え技法を用いた検定が標準的に用いられるようになった。たとえば、permutation test や permuted Brunner-Munzel test (Brunner and Munzel, 2000; Ernst, 2004; Neubert and Brunner, 2007) が挙げられる。これらは総じて、並べ替えの考え方自体は Fisher (1935) によるものであるが、当時はその膨大な計算量ゆえに、実用的ではなかったようである。

⁵たとえば、Guerci et al. (2014) ではモンペリエ大学と筑波大学で実施した実験のデータを、Guerci et al. (2017) では大阪大学と筑波大学で実験を実施し、個別の実験サイトで同じ結果が得られたことにより、統合されたデータを用いて、結果を記述した。

者と一般社会人との比較も重要な研究課題ではあるが、本稿では大学生の被験者に絞って、大学間での外的妥当性を担保するための被験者の特徴を検討する。具体的には、複数の実験サイトにおいて、各サイトの被験者に同一の意思決定問題に対して繰り返し回答してもらい、その問題の背後にある利得生成パターンを学習できるか否かを問う。レーヴンテストでは被験者の視覚的パターン認識能力が計測される⁶。取り扱う意思決定問題として、Guerci et al. (2017) の実験の仕様を踏襲し、重み付き投票に関する二者択一問題を選択した。この問題を選んだ理由は、実験サイトは異なるが、彼らの実験が本研究と同じ日本で実施され、被験者の認知能力テストにおけるスコアも試験的にではあるが既に収集していたことにある⁷。

重み付き投票では、各投票者が異なる票数を持ち、予め定められた票数を集めた提案が採択される。この集団意思決定システムは、議会や株主総会など、様々な実務において用いられている。しかし、そこで起こりうることを予測することは時に困難であり、それゆえに、現在でも理論、実証、実験による研究対象となっている⁸。Montero et al. (2008) や Guerci et al. (2014) の実験では、被験者に複数回の投票を行わせたところ、各回における投票結果にはばらつきが見られた。これは、被験者である投票者間の戦略的相互作用の下で、彼らがとるべき行動に関する学習が行われることによる混乱に起因するのかもしれない。そこで、Guerci et al. (2017) は実験の仕様を大きく簡素化し、被験者が実際に投票を行うことはなく、実験者によって提示された2つの投票状況のうち、彼ら自身にとって有利と思われる一方を選択する。このような実験を重み付き投票を背景にもつ（二腕）バンディット実験といい、この実験の二者択一問題において期待利得がより高い選択肢を正解であるとする。

実験は関西大学、大阪産業大学、同志社大学、広島市立大学で実施した。各被験者は、二者択一問題に回答した後、認知能力テストを受けた。これらのサイトは認知能力テストの平均的正答数に対応する謝金額が予め参加謝金に加算されたグループと政党数に応じて謝金を支払ったグループに分けられている。前者のグループには関西大学と大阪産業大学の被験者、後者のグループには前者とは独立にリクルートした関西大学の被験者、同志社大学、広島市立大学の被験者が属する。前者のグループでは、被験者の認知能力の平均だけでなく、彼らの性別比にも顕著な差がある。関西大学の被験者は男女ほぼ半分ずつであるが、大阪産業大学の被験者は男子学生に極端に偏っている。後者のグループでは、同志社大学の被験者には

⁶被験者の認知能力の差異に起因する実験結果の違いについては、近年、立て続けに報告がなされている。たとえば、Hanaki et al. (2016a)、Hanaki et al. (2016b)、Gill and Prowse (2016)、Basteck and Mantovani (2018)、Proto et al. (2019)、Kawamura and Ogawa (2019)、Hanaki (forthcoming) がレーヴンテストを用いた。Hanaki et al. (2017) は現在改訂中の原稿であるが、ここでもレーヴンテストが用いられている。Watanabe et al. (2020) では、学校選択実験において、レーヴンテストを用いて計測された認知能力スコアが異なる2グループ間では被験者の行動とその帰結であるマッチングの効率性指標に有意な差が生じうることを示した。

⁷Watanabe (2018) は Guerci et al. (2017) の実験を補完する実験を実施し、さらに彼らのデータも用いて、その結果を再検証した。その際にも認知能力テストを試験的に用いている。

⁸欧州議会における事例は Felsenthal and Machover (1998) を見よ。

理工系学生がおらず、広島市立大学の被験者には経済学部生がいないが、関西大学では理工系学生と経済学部生も満遍なく被験者として実験に参加してもらった。広島市立大学の被験者では女子学生が顕著に多い、関西大学の被験者は男女半々である。また、関西大学、同志社大学、広島市立大学でのセッションでは、実験者の声や表情などが被験者の行動に影響を与えないように、音声読み上げソフトウェアを使ってインストラクションを行った。

多サイトで実施した重み付き投票を背景にもつバンディット実験に関する本稿の主要な結果は次のとおりである。(1) 同一地域にあるが、被験者の認知能力と男女比に差がある関西大学と大阪産業大学について、年齢、性別、経済学部生か否か、理工系学生か否かという4つの属性情報だけでは二者択一問題における正答率に有意な差が検出されるが、それらに加えて、認知能力も制御すると、正答率に有意な差はなくなる。(2) 同一地域にあり、被験者の認知能力と男女比に差がない関西大学と同志社大学について、年齢、性別、経済学部生か否かという3つの属性情報だけで、正答率に有意な差が検出されることはなかった。(1)と(2)より、同じ地域にある実験サイトでは、通常の被験者管理において収集する被験者の属性情報(年齢、性別、所属学部)に加えて、彼らの認知能力スコアを使い、それらの分布に平均的には有意な差がないならば、被験者の行動の均質性を保証できる可能性があることが判った。一方、次の結果は、実験サイトが別々の地域にあると、被験者の認知能力と属性情報だけでは彼らの行動の均質性を担保することが難しく、実験サイトごとにデータを分析する必要があることを示唆している。(3) 別地域にあるだけでなく、被験者の認知能力と男女比に差がある関西大学と広島市立大学について、認知能力と年齢、性別、理工系学生か否かという3つの属性情報を用いても、正答率に有意な差がなくなるわけではない。

以上の結果は関西大学を軸とする2校間比較によって得られた。多サイト経済実験の実施可能性を検討する上で最も重要なことは分析対象となる変数(本研究では、二者択一問題における正答率)に有意な差があるかどうかを偏りなく推論することである。したがって、我々は被験者の認知能力スコアと属性(年齢、性別、学部)を変数とする2校への割当確率を求めることで傾向スコアマッチングを行い、傾向スコアを用いて共変量調整された2群間で一定期間における被験者の二者択一問題に対する正答率に有意な差があるかを検討した。

また、本研究の後半では、被験者に二者択一問題に繰り返し回答してもらうことで、彼らが投票状況の背後にある票数と期待利得の関係(正しい選択肢)を学習することができるかを問うている。さらに、関西大学では、意思決定の結果を被験者にフィードバックしない場合においてのみ、被験者による「意味のある学習」が観察された。同じ状況を繰り返し経験したあと、似てはいるが異なる二者択一問題に被験者が直面したとき、それまでに学習したことを応用して、彼らがより大きな期待利得をもたらす選択肢を正しく選ぶならば、それまでの学習は意味のあるものであると考えられる。正確な定義は4.3節で述べるが、これらの4校の

被験者よりもレーヴンテストのスコアが幾分高い大阪大学と筑波大学の被験者では、フィードバックなしの場合においてのみ、この高次の学習が繰り返し観察された⁹。このことも経済実験において被験者の認知能力を制御する必要を示唆している。また、関西大学経済実験室が収集した一般社会人の認知能力スコアと比べると、平均的には、意味のある学習が観察された被験者の認知能力スコアはかなり高かった。つまり、本研究で用いた二者択一問題における推論にはかなり高い認知能力が必要とされることが判った。

学習と意味のある学習の検出においては、Guerci et al. (2017) が提示した判断基準に加えて、被験者が計測期間において Win-Stay-Lose-Shift (WSLS) 行動 (Nowak and Sigmund, 1993) をとっていないという基準も加えた。被験者が高いポイントを得たとき、彼または彼女は直前に選んだ選択肢を次の期も継続して選び、低いポイントを得たとき、次の期には直前に選んだ選択肢とは別の選択肢を選ぶ行動を WSLS 行動という。被験者が WSLS 行動をとるということは、彼または学習した内容に確信が持てていないことを意味する。学習と意味のある学習について、本稿では Guerci et al. (2017) の判定基準に加えて、WSLS 行動に関する基準が適用された。Guerci et al. (2017) の基準であれば、関西大学のみならず、同志社大学でも被験者による意味のある学習が観察されていた。

本稿の構成は次のとおりである。第2章では実験のデザインを説明する。まず、実験の基本仕様について述べ、認知能力テストの概要はその後で説明する。第3章では各実験サイトにおける被験者の認知能力テストの結果と彼らの属性情報について説明する。第4章では、実験結果について述べる。第5章では多サイト経済実験の実施に関する今後の課題について簡潔に整理する。Appendix には、各実験サイトで実施されたセッションに関する詳細な情報、データの時系列プロット、WSLS 行動の検出に用いた数値の表、インストラクションが記述されている。なお、本稿の目的は理論から導かれた仮説の検証ではない。よって、2章以降の記述において、仮説の提示とその検証という形をとっていない。

2 実験のデザイン

2.1 実験の基本仕様

Guerci et al. (2017) の実験は3つのトリートメントを有している。本節ではまず、各トリートメントに共通する事項について述べ、トリートメントの違いについては後述する。すべてのトリートメントにおいて、各セッションは60期からなり、各期において、被験者は4人からなる委員会を2つ提示され、そのどちらかを選ぶよう求められる。委員会のメンバーは合

⁹Guerci et al. (2017) と Watanabe (2018) を参照せよ。

計 120 ポイントをメンバー間で分ける配分案の可否を投票によって決めており、それぞれ数票を与えられている。その票数はメンバー間で異なりうる。1つの委員会は $[q; v_1, v_2, v_3, v_4]$ で表され、 q は配分案の可決に必要な票数であり、 v_i ($i = 1, \dots, 4$) はメンバー i に割り振られた票数を表す。この委員会は、ゲーム理論の用語では、重み付き投票という。本実験では、被験者は常にメンバー 1 の役割を担うが、彼らに求められているのは提示された両委員会において決定される配分を予想しつつ、そのうちの 1 つを選択することであり、その選択後に実際に投票を行うわけではない。このような二者択一問題から被験者が得る利得は、Deegan and Packel (1978) による投票力指数 DPI に基づいて、外生的に決められる。

DPI について説明しよう。投票に参加するメンバーの集まりを提携という。ある委員会が選択されたとする。このとき、メンバー内で $\sum_{i \in S} v_i \geq q$ を満たす票数を集められる提携 S を勝利提携、そうでない提携を敗北提携という。最小勝利提携 (minimal winning coalition, MWC) とは、提携内のどのメンバーが欠けても、敗北提携になってしまう勝利提携のことである。表 1 にはこの実験で用いる二者択一問題がリストされている。そこに表示されている問題 A の選択肢 1 に対応する委員会 $[14; 5, 3, 7, 7]$ を例として、そこでのメンバー 1 の DPI を計算してみよう。この委員会には 3 つの MWC が存在し、すべて 3 人のメンバーからなる。票数の組合せで表すと、それらはともに 7 票を持つメンバーからなる提携、7 票と 5 票と 3 票を集める (7 票を持つメンバーが異なる) 2 つの提携である。被験者がその役割を担うメンバー 1 は 5 票を持つので、それらの MWC が等確率で生起し、MWC 内では均等配分がなされるとすると、メンバー 1 の DPI は $2/3 \times 1/3 = 2/9$ と計算される。120 ポイントがこの DPI に基づいてメンバー 1 に配分されるので、メンバー 1 の期待利得は $120 \times 2/3 \times 1/3$ である。1/3 の確率でメンバー 1 が属する MWC は選ばれず、メンバー 1 の利得はゼロだが、メンバー 1 が属する MWC が選ばれれば、どの MWC でも 40 ポイントを得る。

被験者はこの利得生成の仕組みを知らされず、ある委員会内の意思決定に関するある理論に基づいて、二者択一の意思決定の結果としての配分額が機械的に決まるとのみ知らされる。この実験では、被験者が自らの選択の結果に関するフィードバック情報をどの程度与えるとより多くの期待利得をもたらす選択肢を学習できるかを考察対象とする。この点において、我々は、DPI だけでなく、投票力指数に関する既存の理論であるシャプリー・シュビク指数 (Shapley and Shubik, 1954) やバンザフ指数 (Banzhaf, 1965) でも DPI と同じ選択肢が被験者により大きな利得をもたらす委員会を設定した。したがって、この実験では、DPI を被験者が学習するかどうかを検証する実験にはなっていない。

被験者がより多くの期待利得をもたらす選択肢を学習するかを観察するため、最初の 40 期では同一の二者択一問題が、それに続く 20 期では (最初の 40 期とは異なる) 同一の二者択一問題が被験者に提示される。具体的には、表 1 にリストされている二者択一問題について、

表 1: 実験で用いられた二者択一問題

問題	選択肢 1	メンバー 1 の期待利得	選択肢 2	メンバー 1 の期待利得
A	[14; 5 , 3, 7, 7]	$120 \times 2/3 \times 1/3$	[14; 5 , 4, 6, 7]	$120 \times 3/4 \times 1/3$
B	[6; 1 , 2, 3, 4]	$120 \times 1/3 \times 1/3$	[6; 1 , 1, 4, 4]	$120 \times 2/3 \times 1/3$
C	[14; 3 , 5, 6, 8]	$120 \times 2/3 \times 1/3$	[14; 3 , 6, 6, 7]	$120 \times 3/4 \times 1/3$
D	[9; 1 , 3, 5, 6]	$120 \times 1/3 \times 1/3$	[9; 1 , 2, 6, 6]	$120 \times 2/3 \times 1/3$

Note: 被験者はメンバー 1 の役割を担う。この表では、メンバー 1 の票数は太字で示してある。

$A \rightarrow B$ 、 $B \rightarrow A$ 、 $C \rightarrow D$ 、または、 $D \rightarrow C$ のいずれかの順序が適用される。ここで、矢印の元の問題は最初の 40 期に、矢印の先の問題はそれに続く 20 期に被験者が直面する問題である。ただし、同じ二者択一問題での選択を一定期間求められること自体は被験者には通知されないが、4 人からなるメンバーのうち、メンバー 1 以外のメンバーを他の被験者が担当することはないと通知される。すべての二者択一問題において、表 1 から判るように、より大きな期待利得をもたらすのは選択肢 2 である¹⁰。

本章の初めに言及したトリートメントについて説明しよう。Guerci et al. (2017) にならい、この実験でも、各期における被験者の選択の結果について、3 つのトリートメントを用意し、以下では、それぞれ、(1) フィードバックなし、(2) 部分フィードバック、(3) 完全フィードバックということにする。フィードバックなしトリートメントでは、被験者の選択の結果に関するあらゆる情報が彼らには伏せられたままであり、実験の最後に彼ら自身の累積得点のみが知らされる。部分フィードバックトリートメントではメンバー 1 が獲得した利得のみが選択直後の被験者に知らされ、完全フィードバックトリートメントでは 4 人のメンバー全員の利得 (120 ポイントの配分) が選択直後の被験者に知らされる。

被験者は 30 秒以内に二者択一問題に回答するよう求められ、回答後の 10 秒間、選択の結果を閲覧できる。フィードバックなしトリートメントでは、その 10 秒間、次の二者択一問題が提示されるまで待つように指示される。もし、30 秒以内に被験者が回答できなかった場合には、その期の当該被験者の得点はゼロであり、部分フィードバックトリートメントと完全フィードバックトリートメントでは、その旨が被験者に通知される。

本実験では zTree (Fischbacher, 2007) を用いており、実験中、実験者からの二者択一問題の提示と指示、被験者の回答はすべて計算機ネットワークを通じて行われる。よって、30 秒間の選択時間が経過する前に選択を終えた被験者には、彼らの計算機モニター上に待機を指示する文言が掲示される。同一セッションに参加したすべての被験者の選択が完了するか、30 秒の選択時間が経過すると、一斉に 10 秒間の情報閲覧時間に進む。フィードバックなしトリート

¹⁰Watanabe (2018) は最初の 40 期と 20 期で選択肢 1 と 2 を入れ替えた場合を取り扱っている。

トメントでは、上述の通り、次の二者択一問題が提示されるまで待つようにとの指示が被験者の計算機モニターに掲示される。被験者には、全 60 回を通じて彼または彼女が獲得したポイントが、1 ポイント 1 円で換算され、参加謝金に追加される。参加謝金については、次節を参照してほしい。実験インストラクションは本稿の Appendix D に添付されている。

2.2 レーヴン APM テスト

レーヴンテストのスコアは被験者の認知能力を測る尺度としてよく知られており、言語を通じて被験者の認識を問うものではなく、いくつかの選択肢の中から一つを選択することを彼らに求める問題からなっている。各問題では 8 つの模様が描かれており、被験者はそれらの視覚的パターンに合致する模様を選択肢の中から選択する。レーヴンテストには 3 種類あり、問題の難易度が低いものから順に、Coloured Progressive Matrices (CPM) 版、Standard Progressive Matrices (SPM) 版、Advanced Progressive Matrices (APM) 版と呼ばれている。この実験では APM 版を用いる¹¹。APM 版には全部で 48 問で構成されているが、すべてに回答するには 30 分を要するので、被験者プールの構築のためには 10 分ほどで回答を終えられるようにしなければならない。そこで、本実験では 48 問から 16 問をランダムに選択し、二者択一問題への回答を全て終えた後で、被験者に回答してもらった¹²。本稿では、この 16 問でのレーヴンテストの正解数をレーヴンスコアと呼ぶ¹³。

インストラクションの時間は 5 分とし、被験者には 10 分で 16 問に回答してもらった。回答時間中はどのような順番で回答しても良く、一度回答した問題に再度回答することも可能とした。実験終了後、被験者から求めがあった場合には、その被験者のレーヴンスコアを彼または彼女に対してのみ開示した。我々は、レーヴンテストの各問題について、正解か否かに関わらず一律に参加謝金 500 円を支払った被験者のグループ（関西大学 A と大阪産業大学）と、正解ごとに 50 円の金銭的報酬を支払った被験者のグループ（関西大学 B、同志社大学、広島市立大学）と設定した。表 2 には各実験サイトにおけるレーヴンテストの結果の基礎統計が記されている。表中では、関西大学の被験者と比較対象校の被験者の間に平均的には差があるかを Brunner-Munzel 検定で検証するための p 値（両側検定）も併記されている¹⁴。

¹¹最近、レーヴンテストを用いる経済実験の成果がいくつか知られるようになってきているが Gill and Prowse (2016) と Basteck and Mantovani (2018) は SPM 版を用いており、Proto et al. (2019) は APM 版を用いた。

¹²APM 版 48 問のうち、我々は Set 1 に収録されている問題 1、4、7、10 と Set 2 に収録されている問題 1、4、7、10、13、16、19、22、25、28、31、34 を用いた。Proto et al. (2019) は 48 問中 30 問を用いている。

¹³これらの 16 問は Hanaki et al. (2016a)、Hanaki et al. (2016b)、Guerci et al. (2017)、Watanabe (2018)、Kawamura and Ogawa (2019)、Hanaki (forthcoming)、Watanabe et al. (2020) でも共有されている。

¹⁴レーヴンスコアの平均値は、次節でノートする被験者数からして、実験サイトの代表値とはいえず、これらの平均スコアは実験参加者のものである。被験者からは認知能力テストを行うことについて同意を得ており、実験中いつでも参加を取り止めてよいと事前に説明した。

表 2: 学生被験者のレーヴンスコア：基礎統計量

実験サイト	被験者数	平均	標準偏差	p 値	最低	最高
関西大学 A	240	11.208	2.170		3	16
大阪産業大学	120	10.625	3.041	0.027	3	16
関西大学 B	197	11.518	2.398		2	15
同志社大学	135	11.578	2.300	0.890	5	16
広島市立大学	124	10.976	2.441	0.038	3	15

表 3 は関西大学が位置する大阪府北部に居住する一般社会人の方々のレーヴンスコアに関する基礎統計量であり、図 1 はそのヒストグラムである。関西大学経済実験室が 2015 年から 2018 年にかけてこれらのデータを収集し、その一部は Kawamura and Ogawa (2019) の実験に参加した被験者のものである。これらの図表により、本実験の被験者である大学生の認知能力スコア（平均で 10.625 から 11.578）は一般社会人のそれ（平均 7.986）よりも高いことが判る。一般社会人に向けて様々な制度の導入を検討する際には、彼らの認知能力とその影響を受けるであろう行動を意識して、それらの制度の性能評価実験を実施する必要があるだろう。本研究の目的は大学生を被験者とするデータの統合においてどのような因子に関するデータが必要であるかを考察することであり、制度の設計ではないので、一般社会人が被験者となる実験結果の詳細な分析は他稿に譲ることとする。

表 3: 一般社会人のレーヴンスコア：基礎統計量

	被験者数	平均	標準偏差	最低	最高
関西大学経済実験室	1,057	7.986	3.326	0	16

Note: 計測期間は 2015 年 4 月から 2018 年 3 月まで。年齢は個人情報であるため、すべての被験者が回答したわけではないが、ほぼ全員から回答を得た。回答者の平均年齢は 55 歳前後である。

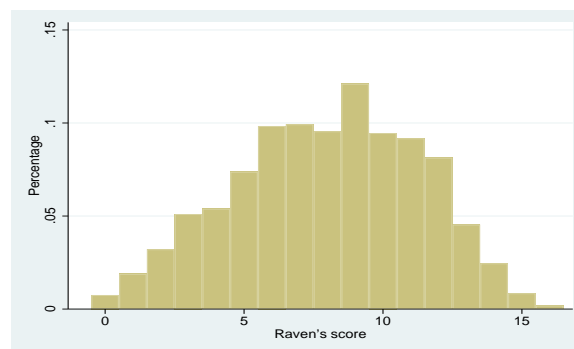


図 1: 一般社会人のレーヴンスコアのヒストグラム：関西大学経済実験室

3 実験の実施

実験は2018年3月2日から2019年10月17日にかけて、各実験サイトにて実施された。全セッションでは816人の被験者がこの実験に参加し、謝金として支払った金額は平均で2534円であった。セッションの詳細は実験サイトごとにAppendix Aにまとめられている。(表5から表9にも、トリートメントごとの平均謝金額がリストされている。小数点以下の数字は四捨五入してある。)表2より、レーヴンスコアについて、関西大学Aの被験者の方が大阪産業大学の被験者よりも平均的には有意に高く、関西大学Bと同志社大学の被験者の間には有意な差はなく、関西大学Bの被験者の方が広島市立大学の被験者よりも平均的には高いことが判る。また、レーヴンテストの正答数によらず一律に参加謝金500円を支払った被験者のグループ(関西大学Aと大阪産業大学)では、同じ実験者がインストラクションを行い、レーヴンテストの各問題について正解ごとに50円の金銭的報酬を支払った被験者のグループ(関西大学B、同志社大学、広島市立大学)では、実験インストラクションを読み上げソフトウェアを用いて行った。関西大学、大阪産業大学、同志社大学は同一地域(関西地方)にあり、広島市立大学はそれらの3校とは別地域(中国地方)にある。

大阪産業大学と同志社大学は関西大学から鉄道を使って1時間以内の距離に立地しており、関西大学から広島市立大学へは新幹線を使っても3時間以上かかる。被験者の意思決定に関する文化的背景を考慮するならば、実験サイトの立地だけではなく、被験者の出身地についてもデータを収集すればよかったのかもしれないが、本研究の目的は文化的背景が意思決定に与える影響を考察することではないので、ここでは被験者の文化的背景は問わない。

表4: 被験者の性別と所属学部

実験サイト	被験者数	女性	男性	p 値	経済学部	理工系	その他
関西大学 A	240	112	128		21	53	166
大阪産業大学	120	21	99	<0.001	47	38	35
関西大学 B	197	99	98		24	41	132
同志社大学	135	59	76	0.828	21	0	114
広島市立大学	124	75	49	0.047	0	24	100

男女比に関するフィッシャー検定のための p 値は関西大学との比較において計測された。

表4には、各実験サイトの被験者の属性に関する内訳が記されている。関西大学の被験者では男女比がほぼ均等であるが、大阪産業大学では男子学生が顕著に多く、広島市立大学では女子学生が顕著に多い¹⁵。関西大学の被験者の男女比との差はフィッシャー検定の p 値(片

¹⁵これらの情報は、各大学が公開している在学者の内訳より、予想されたことであり、本研究の実験サイト選定においても参考にした。

側検定)で示されている。関西大学と同志社大学の被験者の男女比には、両側検定の結果、有意な差はない。また、同志社大学では、理工学部が別のキャンパスにあるため、理工系の被験者がおらず、広島市立大学には経済学部が存在しない。表 5 から表 9 では、各実験サイトにおいて、トリートメントと適用された問題の順序ごとに、セッションへの参加者数と平均謝金額をリストしている。関西大学 A と大阪産業大学のグループでは、問題 A と B だけでなく、問題 C と D でも被験者の行動を観察した。関西大学 B と同志社大学、広島市立大学からなるグループでは、問題 A と B のみにおいて被験者の行動を観察した。

表 5: 関西大学 A

トリートメント	A→B	B→A	C→D	D→C	被験者数	平均謝金額
フィードバックなし	20	20	20	20	80	2480
部分フィードバック	20	20	20	20	80	2537
完全フィードバック	20	20	20	20	80	2483
被験者数	60	60	60	60	240	2495

表 6: 大阪産業大学

トリートメント	A→B	B→A	C→D	D→A	被験者数	平均謝金額
フィードバックなし	10	10	10	10	40	2420
部分フィードバック	10	10	10	10	40	2506
完全フィードバック	10	10	10	10	40	2517
被験者数	30	30	30	30	120	2481

表 7: 関西大学 B

トリートメント	A→B	B→A	被験者数	平均謝金額
フィードバックなし	45	27	72	2401
部分フィードバック	27	26	53	2571
完全フィードバック	29	43	72	2643
被験者数	101	96	197	2586

表 8: 同志社大学

トリートメント	A→B	B→A	被験者数	平均謝金額
フィードバックなし	27	20	47	2567
部分フィードバック	28	20	48	2632
完全フィードバック	20	20	40	2543
被験者数	75	60	135	2581

表 9: 広島市立大学

トリートメント	A→B	B→A	被験者数	平均謝金額
フィードバックなし	20	20	40	2424
部分フィードバック	23	20	43	2661
完全フィードバック	21	20	41	2593
被験者数	64	60	124	2519

4 分析

各実験サイトでの二者択一問題の各期における被験者の平均正答率をプロットした図については、Appendix Bを見てほしい。本章ではまず10期ごとの正答率を関西大学の被験者とその他の実験サイトの被験者の間で比較する。次に、学習と意味のある学習について、各実験サイトの結果を提示する。その際、Win-Stay-Lose-Shift 行動 (Nowak and Sigmund, 1993) を用いて、Guerci et al. (2017) による考察を強化する。同じ状況における意思決定を繰り返すことで正しい選択肢を選べるようになることを「学習」といい、その内容を似てはいるが異なる状況においても適用し、被験者が期待利得のより高い選択肢を正しく選択できたとき、学習は「意味のある」ものであるという (Rick and Weber, 2010)。

2.1 節で述べたように、被験者 (プレイヤー 1) は、どの問題においてどちらの選択肢を選んでも、確率的に 0 ポイントか 40 ポイントを得る。以下では、被験者が 40 ポイントを得たとき、彼または彼女は直前に選んだ選択肢を次の期も継続して選び、0 ポイントを得たとき、次の期には直前に選んだ選択肢とは別の選択肢を選ぶ行動を Win-stay-Lose-Shift (WSLS) 行動ということにする。被験者が WSLS 行動をとるということは、彼または彼女がまだ委員会構成の背後にある利得構造について確信を持つに至っていないことを意味する。その利用の仕方は 4.2 節と 4.3 節で説明する。

次節では、実験サイト間での正答率の差をできるだけ偏りなく推定し、その差が統計的に有意なものであるかを検定する。本章で用いる各種検定では、有意水準を 5% に設定する。

4.1 傾向スコアマッチング

複数のサイトで実験を行う場合、各実験サイトへの被験者の割当は無作為に行われているわけではない。経済実験の被験者の多くは大学生であり、入学前、彼らは受験校を選び、受験を通して大学に選抜される。よって、何らかの補正が施されなければ、サイト間での二者選択問題の正答率を偏りなく比較することはできない。そこで、2つのサイトへの被験者の割当が、仮に、被験者の属性に関わるいくつかの要因によって決まったと考えてみよう。被験者の属性として、たとえば、認知能力スコアを取り上げると、そのスコアは二者択一問題の正答率にも実験サイトの選択にも影響を与える共変量であろう。被験者の所属学部や性別、年齢も同様に共変量の代理変数となりうる。

ここで、被験者の認知能力や年齢、性別、所属学部などの共変量とその代理変数が同じであれば、被験者の各実験サイトへの割り当てと二者択一問題の正答率の分布は独立であると仮定すると、サイト間での二者択一問題の正答率の（平均値の）差を偏りなく推定することができる。2つの実験サイトのどちらかに被験者が割り当てられる確率を e とすると、共変量とその代理変数を要素とするベクトルを一つの変数 e で表すことができ、この確率の近さで被験者をマッチし、二者選択問題の正答率に差があるかどうかを考えることができる。この確率 e を傾向スコアといい、傾向スコアを用いて、2群の被験者をマッチすることを傾向スコアマッチングという。以下では、傾向スコアを2項ロジスティック回帰を使って求める。つまり、被験者の実験サイト1と2の割当について、被験者 i が実験サイト1へ割り当てられる確率を

$$e_i = \frac{1}{1 + \exp(-(b_0 + b_1x_{i1} + \dots + b_px_{ip}))}$$

と推定する。 x_{i1}, \dots, x_{ip} は被験者 i に関する共変量とその代理変数である。

この分析では、2群において傾向スコアが最も近い被験者のペアはマッチされ、彼らの認知能力、年齢、性別、所属学部などの変数の組が大きく異なることはない。そのため、彼らは2つの実験サイトに（疑似的にではあるが）「ランダムに」割り当てられたものとみなすことができる。本研究では、単純な正答率の比較ではなく、このような補正を行った上での正答率の差の比較を行い、 z 検定によりその差が統計的に有意なものかを判断する。

関西大学と大阪産業大学

本節では、関西大学と他校の被験者からなる2群の被験者について傾向スコアマッチングを行い、2群間で10期ごとの平均正答率に有意な差（実験サイトが異なるという処置に対する平均処置効果）があるかを z 検定で検証した。ここでは、フィードバックに関する3つのト

リートメントと問題の順序を無視し、これらのデータをマージした正答率で比較を行う。これにより、トリートメントと問題の順序間での異なる結果は丸められるので、このような比較においても平均正答率に有意な差が検出される場合には、認知能力と通常の被験者管理における属性情報では被験者の行動の均質性を保証することが困難であるといえるだろう。

まず、被験者の男女比に顕著な偏りがある大阪産業大学とそれがほとんどない関西大学の比較を行う。2.2章で言及したように、関西大学の被験者の方が大阪産業大学の被験者よりもレーヴンスコアが有意に高い。表 10 では、トリートメントと利得順のすべてのケースで傾向スコア算出に利用した。Raven's score は Raven、経済学部ダミーは econ、理工系ダミーは sci-eng とした。この表より、被験者の年齢、性別、認知能力について、それぞれ単独の制御では 41-50 期の正答率に有意な差が生じていることが判る。年齢、性別、経済学部生であるか否か、理工系であるか否かといった通常の被験者管理における属性情報だけでは、多くの場合、正答率に有意な差が検出されるが、それらに加えて、認知能力も制御すると、正答率に有意な差はなくなる。以上より、次の結果を得た。

表 10: 平均処置効果の z 検定のための p 値：関西大学 A vs. 大阪産業大学

Score 算出に用いた変数	1 - 10	11 - 20	21 - 30	31 - 40	41 - 50	51 - 60
age	0.095	0.302	0.728	0.871	0.015	0.048
age, gender	0.640	0.975	0.492	0.564	0.069	0.228
age, gender, econ, sci-eng	0.948	0.635	0.662	0.965	0.005	0.043
gender	0.734	0.680	0.534	0.148	0.05	0.157
gender, econ, sci-eng	0.929	0.719	0.267	0.366	0.087	0.423
Raven	0.372	0.252	0.529	0.759	0.014	0.013
Raven, age	0.062	0.977	0.952	0.535	0.016	0.180
Raven, gender	0.483	0.914	0.695	0.789	0.025	0.145
Raven, age, gender	0.822	0.532	0.776	0.275	0.079	0.471
Raven, age, gender, econ, sci-eng	0.937	0.648	0.532	0.323	0.295	0.387

結果 1 同一地域にあるが、被験者の認知能力と男女比に差がある関西大学と大阪産業大学について、年齢、性別、経済学部生か否か、理工系学生か否かといった通常の被験者管理における 4 つの属性情報だけでは正答率に有意な差が検出されるが、それらに加えて、彼らの認知能力も制御すると、正答率に有意な差はなくなる。

関西大学と同志社大学

次は、関西大学とは同一地域にあり、被験者の男女比と認知能力には差がないが、理工系の被験者がいない同志社大学と理工系の被験者が多い関西大学のデータを比較する。ここでも、トリートメントと利得順のすべてのケースで傾向スコア算出に利用した。同志社大学の被験者に理工系学生はいないため、傾向スコア算出のための変数に sci-eng は用いていない。

表 reftab:BvsDoshisha より、被験者の年齢、性別、経済学部生であるか否か、といった通常の被験者管理における属性情報だけで、正答率に有意な差がなくなることが判る。以上より、次の結果を得た。

表 11: 平均処置効果の z 検定のための p 値：関西大学 B vs. 同志社大学

Score 算出に用いた変数	1 - 10	11 - 20	21 - 30	31 - 40	41 - 50	51 - 60
age	0.648	0.364	0.024	0.339	0.834	0.432
age, gender	0.570	0.745	0.888	0.649	0.517	0.741
age, gender, econ	0.972	0.906	0.430	0.622	0.606	0.368
gender	0.793	0.866	0.157	0.467	0.691	0.100
gender, econ	0.758	0.891	0.138	0.597	0.727	0.109
Raven	0.878	0.939	0.135	0.472	0.513	0.050
Raven, age	0.539	0.996	0.093	0.661	0.656	0.133
Raven, gender	0.509	0.659	0.134	0.770	0.753	0.132
Raven, age, gender	0.243	0.320	0.355	0.377	0.957	0.087
Raven, age, econ	0.745	0.887	0.341	0.731	0.959	0.122
Raven, age, gender, econ	0.769	0.924	0.371	0.697	0.537	0.051

結果 2 同一地域にあり、被験者の認知能力と男女比に差がない関西大学と同志社大学について、年齢、性別、経済学部生か否かといった通常の被験者管理における 3 つの属性情報だけで、正答率に有意な差が検出されることはなかった。

関西大学と広島市立大学

次に、関西大学とは別地域にあり、被験者の男女比と認知能力には差があるだけでなく、経済学部を持たない広島市立大学と経済学部を有する関西大学のデータを比較する。ここでも、トリートメントと利得順のすべてのケースで傾向スコア算出に利用した。広島市立大学の被験者に経済学部生はいないため、傾向スコア算出のための変数に econ は用いていない。表 12 より、被験者の認知能力、年齢、性別、理工系学生であるか否かという属性情報だけでは正答率に有意な差がなくなることにはなかった。以上より、次の結果を得た。

結果 3 別地域にあるだけでなく、被験者の認知能力と男女比に差がある関西大学と広島市立大学について、認知能力と年齢、性別、理工系学生か否かという 3 つの属性情報を用いても、正答率に有意な差がなくなるわけではない。

表 12: 平均処置効果の z 検定のための p 値 : 広島市立大学 vs. 関西大学 B

Score 算出に用いた変数	1 - 10	11 - 20	21 - 30	31 - 40	41 - 50	51 - 60
age	0.035	0.747	0.013	0.123	0.145	0.201
age, gender	0.062	0.838	0.024	0.105	0.320	0.811
age, gender, sci-eng	0.086	0.953	0.025	0.378	0.146	0.970
gender	0.116	0.414	0.018	0.067	0.062	0.131
gender, sci-eng	0.208	0.440	0.056	0.133	0.158	0.344
Raven	0.160	0.905	0.154	0.252	0.162	0.095
Raven, age	0.392	0.999	0.179	0.092	0.114	0.163
Raven, gender	0.100	0.947	0.142	0.200	0.113	0.339
Raven, age, gender	0.091	0.642	0.102	0.227	0.550	0.744
Raven, age, gender, sci-eng	0.367	0.903	0.104	0.279	0.030	0.339

表 12 では、被験者の年齢、性別、理工系学生であるか否かという属性情報だけでは正答率に有意な差がなくなることはなかったが、認知能力と年齢または性別を制御すると、正答率に有意な差はなくなる。しかし、被験者の認知能力、年齢、性別、理工系学生であるか否かを制御すると正答率に有意な差が現れる。これらの変数のうち、どれを制御すれば良いかを実験前に推測することは困難であろう。この結果は、恐らく、被験者の認知能力、年齢、性別、所属学部だけでなく、地域性を制御する必要があることを示唆していると思われる。

データ処理を問題 A→B について行った結果でも、表 13 に示されているように、結果 3 は保持されている。実験サイトの立地という点では、同志社大学は関西大学と同一地域にあるので、広島市立大学とは異なる地域にある。さらに被験者の認知能力という点では、同志社大学の被験者と関西大学の被験者の間に平均的には有意な差はない。そのため、広島市立大学と同志社大学でも類似した結果が観察されるはずである。広島市立大学には経済学部生の被験者がおらず、同志社大学では理工系学生の被験者はいない。よって、傾向スコアの計算に所属学部に関する情報を反映させることはできない。それゆえに平均処置効果に偏りが潜んでいる可能性は否定できない。そのため、強くは主張できないが、両実験サイトでの被験者行動の均質性を通常の被験者管理における被験者属性である年齢、性別だけで保証できるかということ、表 14 に示されているように、それは難しいようである¹⁶。広島市立大学と同志社大学の間でも被験者の認知能力を制御する必要はあるだろう。

表 13: 平均処置効果の z 検定のための p 値 : 広島市立大学 vs. 関西大学 B, 問題 A→B

Score 算出に用いた変数	1 - 10	11 - 20	21 - 30	31 - 40	41 - 50	51 - 60
age	0.032	0.371	0.025	0.278	0.081	0.274
age, gender	0.025	0.220	0.014	0.138	0.070	0.348
age, gender, sci-eng	0.025	0.059	0.067	0.286	0.065	0.404
gender	0.082	0.404	0.024	0.120	0.009	0.062
gender, sci-eng	0.144	0.316	0.092	0.176	0.066	0.218
Raven	0.132	0.735	0.270	0.196	0.058	0.114
Raven, age	0.201	0.859	0.237	0.311	0.052	0.146
Raven, gender	0.110	0.967	0.175	0.201	0.092	0.126
Raven, age, gender	0.082	0.648	0.149	0.350	0.079	0.323
Raven, age, gender, sci-eng	0.014	0.247	0.668	0.914	0.093	0.429

4.2 学習と WSLs 行動

1.2 節と 2.1 節で述べたように、本稿は Guerci et al. (2017) の実験の仕様を踏襲している。彼らの研究目的は意味のある学習がどのような情報のフィードバックによってもたらされるかを考察することであった。多サイト経済実験を実施するための基盤整備のためには、同じ

¹⁶この 2 校の被験者では男女比にも有意な差がある。フィッシャー検定の p 値 (両側検定) は 0.0089 である。

表 14: 平均処置効果の z 検定のための p 値 : 広島市立大学 vs. 同志社大学, 問題 A→B

Score 算出に用いた変数	1 - 10	11 - 20	21 - 30	31 - 40	41 - 50	51 - 60
age	0.004	0.050	0.948	0.332	0.199	0.608
age, gender	0.028	0.161	0.712	0.670	0.083	0.690
gender	0.020	0.044	0.544	0.358	0.054	0.902
Raven	0.014	0.109	0.750	0.339	0.052	0.544
Raven, age	0.131	0.823	0.763	0.788	0.102	0.406
Raven, gender	0.019	0.084	0.923	0.471	0.035	0.580
Raven, age, gender	0.078	0.431	0.735	0.687	0.123	0.741

仕様で実施された先行論文の結果が我々の実験でも観察されたかどうかを検証し、結果 1 から結果 3 に対して付加すべき何らかの示唆を得るべきだろう。本節では「学習」に考察の焦点を当てる。学習が検出されないならば、意味のある学習を考察することに意味はない。

表 15 は各トリートメントにおける第 1 期と第 41 期に期待利得の高い選択肢を正しく選択した被験者の割合を実験サイトごとにリストしている。ここでは、すべての実験サイトにおけるほぼすべてのトリートメントにおいて、問題 A よりも問題 B の方が正当者数の比率が低く、関西大学 A と大阪産業大学におけるほぼすべてのトリートメントにおいて、問題 C よりも問題 D の方が正当者数の比率が低い。問題 A と C では、選択肢の一方ではより多くの票を持つ 2 人の投票者のみで勝利提携を形成することが可能であるが、他方はそうではない。問題 B と D では、両選択肢でより多くの票を持つ 2 人の投票者のみで勝利提携を形成することが可能である。このことが被験者の選択を困難にし、問題 B と C において正答者数の比率が低いことの背後にあると思われる。このように、2.1 節で設定した問題 A と B のペアと問題 C と D のペアには、見かけ上の票数などが異なるが、背後にある理論上の構造は同一である。被験者もこの点を認識していたのであろう。

個人 i が隣接する 5 期ごとの k 番目のブロックにおいて正しい選択肢を選択する相対頻度を FR_k^i と表そう。たとえば、 FR_2^i は被験者 i が 6 期から 10 期までに正しい選択肢を選択した回数を 5 で除したものである。被験者 i が l 番目と m 番目のブロックで正しい選択肢を選ぶ相対頻度の変化分を

$$\Delta FR_{l,m}^i = FR_l^i - FR_m^i.$$

と定義する。以下では、特に断らない限り、上付き文字での記号 i を省略し、 $\Delta FR_{l,m}^i$ を $\Delta FR_{l,m}$ と書くことにする。被験者は、フィードバックなしでも、心の中で思考を反芻する「内省」によってより高い期待利得をもたらす選択肢を選ぶことを学習することがあると考える。

各二者択一問題について、片側符号順位 (SR) 検定における p 値の結果を表 16 から表 19 にリストしてある。そこでは、帰無仮説は $\Delta FR_{2,1} \leq 0$ ($\Delta FR_{8,1} \leq 0$) であり、対立仮説は $\Delta FR_{2,1} > 0$ ($\Delta FR_{8,1} > 0$) である。Guerci et al. (2017) は第 1 ブロックと第 8 ブロックの間

	第 1 期	No-fb	Part-fb	Full-fb	第 41 期	No-fb	Part-fb	Full-fb
関西大学 A	問題 A	0.70	0.70	0.60	問題 A	0.70	0.55	0.55
	問題 B	0.20	0.25	0.20	問題 B	0.40	0.50	0.40
	問題 C	0.70	0.70	0.65	問題 C	0.55	0.70	0.55
	問題 D	0.45	0.55	0.40	問題 D	0.80	0.45	0.45
大阪産業大学	問題 A	0.60	0.70	0.30	問題 A	0.40	0.60	0.40
	問題 B	0.40	0.30	0.50	問題 B	0.30	0.30	0.30
	問題 C	0.60	0.60	0.60	問題 C	0.80	0.70	0.70
	問題 D	0.60	0.10	0.20	問題 D	0.20	0.30	0.50
関西大学 B	問題 A	0.689	0.741	0.621	問題 A	0.778	0.538	0.535
	問題 B	0.333	0.308	0.349	問題 B	0.489	0.370	0.345
同志社大学	問題 A	0.593	0.679	0.800	問題 A	0.700	0.550	0.500
	問題 B	0.400	0.450	0.350	問題 B	0.519	0.464	0.350
広島市立大学	問題 A	0.500	0.609	0.714	問題 A	0.600	0.250	0.550
	問題 B	0.350	0.300	0.350	問題 B	0.500	0.478	0.381

表 15: 第 1 期と第 41 期において期待利得の高い選択肢を正しく選択肢した被験者の割合。No-fb はフィードバックなし、Part-fb は部分フィードバック、Full-fb は完全フィードバックに対応するトリートメントを表す。

で正しい選択肢が選ばれる相対頻度の上昇が確認されたとき、被験者は学習したとみなした。たとえ $\Delta FR_{2,1} \leq 0$ が棄却されても、 $\Delta FR_{8,1} \leq 0$ が棄却されなければ、被験者による学習がなされたとは判断しない。この基準を適用すると、フィードバックなしのトリートメントにおいて、広島市立大学における問題 B 以外では被験者による学習が観察されなかった。

一方、Guerci et al. (2017) では問題 B と D で、Watanabe (2018) ではすべての問題で、フィードバックなしでも被験者による学習が観察された。Watanabe (2018) の実験は関西大学と同一地域にある大阪大学で実施されたので、実験サイトの所在地の違いがこの結果の違いを生み出す源泉ではないだろう。経済学部と理工系学部に所属する被験者もその実験に参加した。よって、フィードバックなしのトリートメントにおける被験者の学習に関する結果の違いはの認知能力スコアの違いが関わっているのかもしれない。Guerci et al. (2017) と Watanabe (2018) の実験サイトである大阪大学と筑波大学の被験者のレーヴンスコアはともに平均で 12.3 であり、本稿のための実験に参加した被験者のそれよりも有意に高い。

フィードバックなしのトリートメントでは、被験者は意思決定の結果を実験終了後まで知ることができない。そのため、WSLS 行動をとりようがない。しかし、上述のように被験者による学習がほぼなされていないならば、その他のトリートメントに対して、WSLS 行動が観察された場合、実際には学習は行われていなかったと判断する基準を適用しても支障はないだろう。Appendix C には、実験サイトとトリートメント (Part-fb は部分フィードバック、Full-fb は完全フィードバックのトリートメントを意味する)、問題ごとに、第 1 ブロック (B1) と第 8 ブロック (B8) において観察された得点の頻度 (freq)、得点を観察した直後の選択肢の変更頻度 (switch)、それらの比 (ratio)、両側フィッシャー検定のための p 値が記されて

いる。被験者が0ポイントを得たときに選択肢の変更を行った頻度が40ポイントを得た時のそれよりも相対的に高いことがれば、彼らはWSLS行動をとったと判断する。5期からなるブロックでは、個人のこのような行動を統計的に検出することは困難なので、本稿では5期からなるブロックにおいて観察された被験者の行動すべてを集計して検定を行う。

本稿における被験者による学習とは次の2条件を満たすことをいう。

- $\Delta FR_{8,1} > 0$
- 第8ブロックにおいてWSLS行動が観察されない。

これらの基準を適用し、我々は次の結果を得た。検定結果の詳細は表16から表19を参照せよ。多サイト経済実験に関する結果の総括は、最終章において、意味のある学習に関する結果（次節）と併せて行う。

結果4 関西大学Aと大阪産業大学では、部分フィードバックでの問題Bにおいて、ともに被験者による学習が観察された。関西大学Aではフィードバックなしでの問題Bと問題Dにおいて被験者による学習が観察された。

結果5 関西大学Bと同志社大学では被験者による学習が観察されなかったが、広島市立大学では問題Bにおいてフィードバックなしのトリートメントと完全フィードバックのトリートメントでそれが観察された。

表 16: 片側符号順位検定のための p 値 ($H_0: \Delta FR_{2,1} \leq 0$)、2校
関西大学 A 大阪産業大学

	関西大学 A			大阪産業大学		
	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
問題 A	0.029	0.593	0.395	0.813	0.938	0.500
問題 B	0.212	0.783	0.402	0.637	0.344	0.910
問題 C	0.363	0.910	0.598	0.813	0.891	0.109
問題 D	0.105	0.090	0.710	0.227	0.145	0.500

太字で p 値が表示されているものはほとんどなく、まだ学習は進んでいないことが伺える。

表 17: 片側符号順位検定のための p 値 ($H_0: \Delta FR_{8,1} \leq 0$)、2校
関西大学 A 大阪産業大学

	関西大学 A			大阪産業大学		
	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
問題 A	0.227	0.387	0.212	0.688	0.125	0.363
問題 B	0.048	0.001	0.001	0.855	0.008	0.500
問題 C	0.500	0.212	0.227	0.813	0.500	0.109
問題 D	0.038	<0.001	0.059	0.227	0.500	0.172

太字で p 値が表示されているもののうち、関西大学 A の Part-fb における問題 D と Full-fb の問題 B では、第8ブロックでのWSLS行動が観察された。WSLS行動の検出については、Appendix C の表27から表30を参照せよ。

表 18: 片側符合順位検定のための p 値 ($H_0 : \Delta FR_{2,1} \leq 0$)、3 校
 関西大学 B 同志社大学 広島市立大学

	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
問題 A	0.819	0.252	0.661	0.227	0.500	0.315	0.090	0.985	0.324
問題 B	0.132	0.001	0.500	0.227	0.605	0.910	0.500	0.500	0.151

太字で p 値が表示されているものはほとんどなく、まだ学習は進んでいないことが伺える。

表 19: 片側符合順位検定のための p 値 ($H_0 : \Delta FR_{8,1} \leq 0$)、3 校

関西大学 B 同志社大学 広島市立大学

	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
問題 A	0.066	0.668	0.115	0.808	0.808	0.151	0.395	0.315	0.105
問題 B	0.402	<0.001	0.007	0.090	<0.001	0.240	0.011	0.227	0.006

太字で p 値が表示されているもののうち、関西大学 B と同志社大学の Part-fb と関西大学 B の Full-fb における問題 B では、第 8 ブロックでの WSLs 行動が観察された。WSLS 行動の検出については、Appendix C の表 35 から表 40 を参照せよ。

4.3 意味のある学習と WSLs 行動

Guerci et al. (2017) では、フィードバックなしでも被験者による学習（内省）が観察され、さらに、フィードバックなしのトリートメントでのみ「意味のある学習」が観察されたことが報告されている¹⁷。Guerci et al. (2017) と Watanabe (2018) は次の 2 条件を満たすとき、意味のある学習が観察されるとして、問題 D ではフィードバックなしでも意味のある学習を観察した。FR₁ と FR₉ は、当該ブロックにおいて被験者 i が期待利得のより高い選択肢を選んだ回数を 5 で除した値を第 i 成分とするベクトルである。

- 1 期と 41 期でより期待利得の大きな選択肢を選んだ被験者の数が有意に増えている。
- FR₁ の（相加）平均よりも FR₉ のそれの方が有意に大きい。

1 期と 41 期、第 1 ブロックと第 9 ブロックでは問題が異なるが、上述（Table 15 参照）のように、問題が異なれば難易度も異なる。そのため、第 1 期と第 1 ブロックにおける回答者を「経験のない」被験者、第 41 と第 9 ブロックにおける回答者を「経験のある」被験者と呼び、同じ問題で未経験な被験者と経験のある被験者の行動を比較する。上述の 2 論文ではすべてのトリートメントと問題で被験者数が同じだが、本稿の関西大学 B、同志社大学、広島市立大学ではそれが統一されていないので、1 期と 41 期に関する上記の基準は適用することができない。したがって、それに代えて、ここでは WSLs 行動を付加して意味のある学習が観察されたと判断する。より正確には次の基準を用いる。

- FR₁ の（相加）平均よりも FR₉ のそれの方が有意に大きい。
- 第 9 ブロックと第 12 ブロックにおいて WSLs 行動が観察されない。

被験者が 40 期までの学習内容に確信を持っているならば、第 9 ブロックで WSLs 行動はとらないだろう。しかし、第 12 ブロックになって WSLs 行動をとるならば、その確信は揺らいでいることになるので、意味のある学習が観察されなかったと判断する。多サイト経済実験に関する結果の総括は、最終章において、学習に関する結果（前節）と併せて行う。

表 20 から表 21 は、経験のない被験者と経験のある被験者のデータ（FR₁ と FR₉）について、平均正答率と Brunner-Munzel 検定のための p 値をリストしてある。平均正答率とは、各実験サイトでのトリートメントと問題ごとの FR _{k} ($k = 1, 9$) の相加平均である。Appendix C には、実験サイトとトリートメント、問題ごとに、第 9 ブロック (B9) と第 12 ブロック (B12)

¹⁷Weber (2003) と Rick and Weber (2010) は意思決定に至るまでに被験者が先読みを繰り返す必要がある p -美人投票ゲームにおいてフィードバックを控えることが意味のある学習を促進することを観察しているが、Neugebauer et al. (2009) は公共財供給ゲームではフィードバックなしだと被験者は支配戦略を選択することさえ学習しないことを報告している。

において観察された得点の頻度 (freq)、選択肢の変更頻度 (switch)、それらの比 (ratio)、両側フィッシャー検定のための p 値が記されている。

結果 4 より、関西大学 A でも部分フィードバックでの問題 B において被験者による学習が観察されたが、表 20 には FR_1 と FR_9 に関する上記基準を満たさないことが示されている。結果 5 より、広島市立大学のみで被験者による学習が観察された（フィードバックなしと完全フィードバックのトリートメント）が、表 21 にはそれらのトリートメントにおいて、 FR_1 と FR_9 に関する上記基準を満たさないことが示されている。ここで注意したいのは、学習の検出に WSLs に関する基準が適用されたことである。Guerci et al. (2017) の基準であれば、関西大学 A における部分フィードバックでの問題 D、および、同志社大学における部分フィードバックでの問題 B においても、被験者による意味のある学習が観察されていた。

結果 6 関西大学 A ではフィードバックなしでの問題 D において意味のある学習が観察されたが、大阪産業大学ではどのトリートメントにおけるどの問題でも意味のある学習は観察されなかった。

結果 7 関西大学 B、同志社大学、広島市立大学では、問題 A と問題 B のみが試されたが、どのトリートメントにおけるどの問題でも意味のある学習は観察されなかった。

Guerci et al. (2017) では問題 B と D において、Watanabe (2018) では問題 D において、フィードバックなしの場合に被験者による学習が観察された。結果 6 は関西大学でも同じ結果が問題 D で観察されたことを示している。なお、フィードバックなしのトリートメントでの行動には、WSLS 行動に関する基準は適用されていない。関西大学 B、同志社大学、広島市立大学では問題 A と問題 B のみが試されたが、結果 2 より関西大学と同志社大学の被験者は同質性が高いので、問題 C と問題 D も試されていた場合には、関西大学 A と同様に、関西大学 B と同志社大学でもフィードバックなしでの問題 D において、意味のある学習が観察されていたかもしれない。

学生被験者のレーヴンスコア（表 2）、結果 6、結果 7、一般社会人のレーヴンスコア（表 3）、Guerci et al. (2017) と Watanabe (2018) の結果より、意味のある学習はレーヴンスコアの高い被験者によってなされると推測できる。では、レーヴンスコアの高い被験者でもフィードバックがある場合において、なぜ意味のある学習が観察されなかったのだろうか。本節の最後にそれを確認しておこう。

Watanabe (2018) は Guerci et al. (2017) の実験データを再検証し、部分フィードバックと完全フィードバックのトリートメントにおいて、被験者は WSLs 行動をとっていることを報告した。この結果はフィードバックが意味のある学習を妨げる混乱要因となっていることを

表 20: B1（経験なし）と B9（経験あり）での平均正答率：関西大学 A、大阪産業大学
 関西大学 A 大阪産業大学

問題 A	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
経験なし	0.530	0.630	0.680	0.700	0.720	0.580
経験あり	0.670	0.500	0.560	0.620	0.480	0.380
p-value	0.089	0.316	0.322	0.682	0.166	0.313

問題 B	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
経験なし	0.380	0.450	0.440	0.460	0.440	0.620
経験あり	0.500	0.540	0.490	0.360	0.440	0.500
p-value	0.243	0.260	0.466	0.328	0.971	0.323

問題 C	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
経験なし	0.690	0.530	0.560	0.660	0.700	0.600
経験あり	0.620	0.640	0.730	0.740	0.620	0.540
p-value	0.777	0.206	0.034	0.498	0.449	0.625

問題 D	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
経験なし	0.430	0.510	0.550	0.480	0.380	0.400
経験あり	0.620	0.610	0.510	0.200	0.440	0.600
p-value	0.031	0.182	0.560	0.024	0.456	0.103

太字で P 値が表示されているもののうち、関西大学 A における完全フィードバックでの問題 C、大阪産業大学におけるフィードバックなしでの問題 D では、被験者による学習が B1 と B8 の間で観察されていない（結果 4）参照。

表 21: B1（経験なし）と B9（経験あり）での平均正答率：関西大学 B、同志社大学、広島市立大学
 関西大学 B 同志社大学 広島市立大学

問題 A	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
経験なし	0.680	0.593	0.641	0.622	0.614	0.670	0.560	0.626	0.571
経験あり	0.689	0.685	0.586	0.690	0.610	0.560	0.590	0.410	0.560
p-value	0.840	0.136	0.632	0.305	0.885	0.258	0.569	< 0.001	0.891

問題 B	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb	No-fb	Part-fb	Full-fb
経験なし	0.452	0.492	0.493	0.490	0.500	0.490	0.420	0.510	0.430
経験あり	0.538	0.563	0.469	0.541	0.636	0.460	0.520	0.548	0.571
p-value	0.231	0.270	0.761	0.581	0.039	0.517	0.235	0.475	0.069

太字で P 値が表示されているもののうち、同志社大学における部分フィードバックでの問題 B、広島市立大学における部分フィードバックでの問題 A では、被験者による学習が B1 と B8 の間で観察されていない（結果 5）参照。

示唆している¹⁸。実際、Watanabe (2019) は筑波大学で Guerci et al. (2017) の実験とほぼ同じ内容でマウストラッキング実験を実施し、意味のある学習を行った被験者は意思決定直後の利得よりも自分自身の累積利得をより長く閲覧していることを報告している¹⁹。これらの結果が、本稿において、学習と意味のある学習の判定基準に被験者の WSLS 行動を加えた根拠である。

5 今後の課題

本稿では、Guerci et al. (2017) の実験について、被験者の一般的属性情報（年齢、性別、所属学部）に加えて、彼らの認知能力スコアをとり、それらを用いて共変量調整すると、同一地域内の実験サイト間では被験者行動に一定の均質性があることが示された。このことは、被験者の認知能力スコアを予め収集してあれば、複数のサイトで実施した実験のデータを用いた分析をより容易に実施することが可能かもしれないことを示している。ただし、異なる実験サイト間で予め収集した被験者の認知能力スコアと属性情報の分布がほぼ同一になるような被験者募集を実施することは、各サイト内でのランダムな被験者の募集を阻害しかねないことに注意すべきではあるだろう。

本稿では、また、実験サイト間で被験者による学習と意味のある学習が観察されるかどうかという面でも考察を行った。結果 4 と結果 6 は関西大学と大阪産業大学で観察されたことに言及している。これらの実験サイトでは被験者の認知能力と男女比に差があったので、学習と意味のある学習に関する結果は異なった。ただし、これら 2 つの実験サイトは同一地域内にあり、被験者の属性情報や認知能力スコアで共変量調整すれば、被験者の行動に均質性もたらされる（結果 1）。結果 5 と結果 7 は関西大学、同志社大学、広島市立大学で観察されたことに言及している。被験者の認知能力と男女比に差がなく、同一地域にある関西大学と同志社大学では、学習においても、意味のある学習においても、ほぼ同じ結果を得たといえるだろう。これは結果 2 が示唆することと矛盾しない。関西大学と広島市立大学の被験者の間では様々な点が異なっており、学習に関する結果に差が生じた。このことも結果 3 の意味することと矛盾はない。

1.1 節冒頭でも触れたように、実験結果の外的妥当性を確保することは重要である。オン

¹⁸Watanabe (2018) はフィードバックなしの場合のみ Guerci et al. (2017) の追試を実施し、Guerci et al. (2017) のデータとの照合によって、次のことを報告している。(1) 41 期目に初めて問題 D での選択に直面したとき、フィードバックなしのセッションにおいて最初の 40 期を経験した被験者は部分フィードバックや完全フィードバックでそれを経験した被験者と平均的には同じ長さの時間をかけて選択している。(2) 1 期から 5 期に問題 D での選択に直面した被験者よりも、41 期から 45 期に直面した被験者の方が選択までにかかる時間が長い。

¹⁹そこでは、被験者のモニターにおいて二者択一問題における委員会メンバーの票数や意思決定後に実現する利得を隠しておき、被験者がそれらの情報を閲覧する際にはマウスをモニター上のその場所に持って行き、クリックする必要がある。

ライン実験では、その実施環境が特定の物理的サイトに限定されないため、外的妥当性の確保がより容易になることがある。よって、最近ではオンラインでの実験がしばしば実施されるようになってきた。オンライン実験において最も多く使用されているプラットフォームは Amazon Mechanical Turk (MTurk) であり、たとえば、Arechar et al. (2018) はそれを用いて繰り返し公共財供給ゲームの実験を実施し、オンライン実験における注意点をいくつか提示している²⁰。本稿において提示された主要な結果は、実験サイトが同一地域にない場合、被験者の一般的な属性情報や認知能力スコアだけでは彼らの行動を説明できないことを示唆している。このことは、地域ごとに被験者の行動は文化的背景を持ち、実験結果を吟味する際にはその違いを考慮する必要があるかもしれないことを示唆する。しかし、実験サイトが同一地域にあるならば、文化的背景などをあまり考慮しなくても、被験者の一般的属性や認知能力スコアで彼ら行動の均質性を確保できる可能性があることも、本稿の結果が示唆するところである。

本研究の被験者は大学生であり、彼らの行動に限定して、本稿では多サイト経済実験を実施する上での注意点を明確にした。そのため、実験結果のより一般的な外的妥当性を検証したわけではないが、本稿では、一般社会人の認知能力スコアも併記した。理論的考察を経て設計された制度が実用化されたとき、その下で実際に意思決定を行うのは一般社会人であることが多い。しかし、被験者への謝金などの実験にかかる費用を工面できない場合、設計された制度が実際に適用される前にその性能を評価するには、まず、学生を被験者として実験を実施し、その結果を参照することになる。一般社会人と学生では被験者としての属性が異なるが、被験者となる学生の認知能力が一般社会人のそれに対してどの程度異なるものであるかを認識しておくことは重要であろう。

最後に、次の点にも言及しておこう。本研究で観察された実験結果は被験者による個別の意思決定に関するものであり、他の個人との相互依存的関係の下でなされる意思決定（メカニズムやゲーム理論）に関するものではない。本研究に続いて、メカニズムやゲーム理論の実験でも、被験者の認知能力と属性情報の相違によって実験結果に相違が観察されるか否かに関する研究がなされ、さらに国内のみならず、国外の実験室を含めて、複数のサイトで実験データを収集する際の方法論に関する研究がさらに進むことを期待したい。レーヴンテストを実施するためのキットおよび回答用紙には著作権が発生しているため、それらを有料で購入する必要がある。しかし、同種の認知能力テストに Matrix Reasoning Quiz (MRQ) が開発されており、経済実験だけではなく、他の社会科学諸分野との連携を取ろうとする動き (<https://icar-project.com/>) が出てきている。

²⁰たとえば、実験室実験では実験への参加を実験中に取りやめる被験者はほとんどいないが、オンライン実験ではそれが頻繁に発生することを Arechar et al. (2018) は報告している。

Appendix A: Session Details

各サイトにおける実験実施日などの詳細は次のとおりである。

表 22: 関西大学でのセッション A

実験実施日	フィードバック	問題の順序	被験者数	平均	最高	最低	標準偏差
2019/02/16	partial	A→B	20	2654	2840	2400	132
2019/02/16	partial	B→A	20	2470	2840	2120	213
2019/02/26	fullinfo	A→B	20	2572	3040	2160	211
2019/02/26	fullinfo	B→A	20	2452	2800	2200	158
2019/03/08	no info	A→B	20	2574	2840	2240	151
2019/03/08	no info	B→A	20	2454	3080	2080	259
2019/06/13	no info	C→D	20	2532	2960	2320	176
2019/06/13	no info	D→C	20	2358	2920	2000	257
2019/06/15	partial	C→D	20	2520	2840	2240	138
2019/06/15	partial	D→C	20	2438	2680	2160	157
2019/07/11	fullinfo	C→D	20	2526	2840	2160	193
2019/07/11	fullinfo	D→C	20	2384	2720	2120	164
Total			240	2495	3080	2000	205

表 23: 大阪産業大学でのセッション

実験実施日	フィードバック	問題の順序	被験者数	平均	最高	最低	標準偏差
2018/11/19	no info	A→B	10	2456	2800	2080	167
2018/11/19	partial		10	2464	2560	2360	57
2018/11/19	fullinfo		10	2596	2840	2280	187
2018/11/20	no info	B→A	10	2336	2600	2000	174
2018/11/20	partial		10	2516	2840	2320	185
2018/11/20	fullinfo		10	2444	2720	2120	164
2018/12/03	no info	C→D	10	2516	2840	2120	203
2018/12/03	partial		10	2596	2760	2360	117
2018/12/03	fullinfo		10	2540	2920	2120	224
2018/12/10	no info	D→C	10	2372	2760	2000	230
2018/12/10	partial		10	2448	2640	2280	130
2018/12/10	fullinfo		10	2488	2680	2200	160
Total			120	2481	2920	2000	188

表 24: 関西大学でのセッション B

実験実施日	フィードバック	問題の順序	被験者数	平均	最高	最低	標準偏差
2018/03/02	fullinfo	A→B	17	2,755	3,050	2,390	156
2018/03/02	fullinfo		12	2,664	2,950	2,370	211
2018/03/05	no info	A→B	9	2,599	3,000	2,250	242
2018/03/05	no info		13	2,718	3,010	2,260	203
2018/05/16	partial	A→B	14	2,592	2,860	2,250	171
2018/05/16	parital		13	2,635	3,040	2,180	237
2018/05/21	fullinfo	B→A	12	2,558	2,750	2,250	155
2018/05/21	fullinfo		11	2,541	2,850	2,200	207
2018/06/20	no info	B→A	8	2,450	2,600	2,230	139
2018/10/03	partial	B→A	13	2,517	2,780	2,230	144
2018/10/03	parital		13	2,539	3,030	2,310	198
2019/05/27	fullinfo	B→A	19	2419	2770	1920	243
2019/11/22	no info	A→B	23	2128	2420	1820	158
Total			197	2,586	3,050	1,900	222

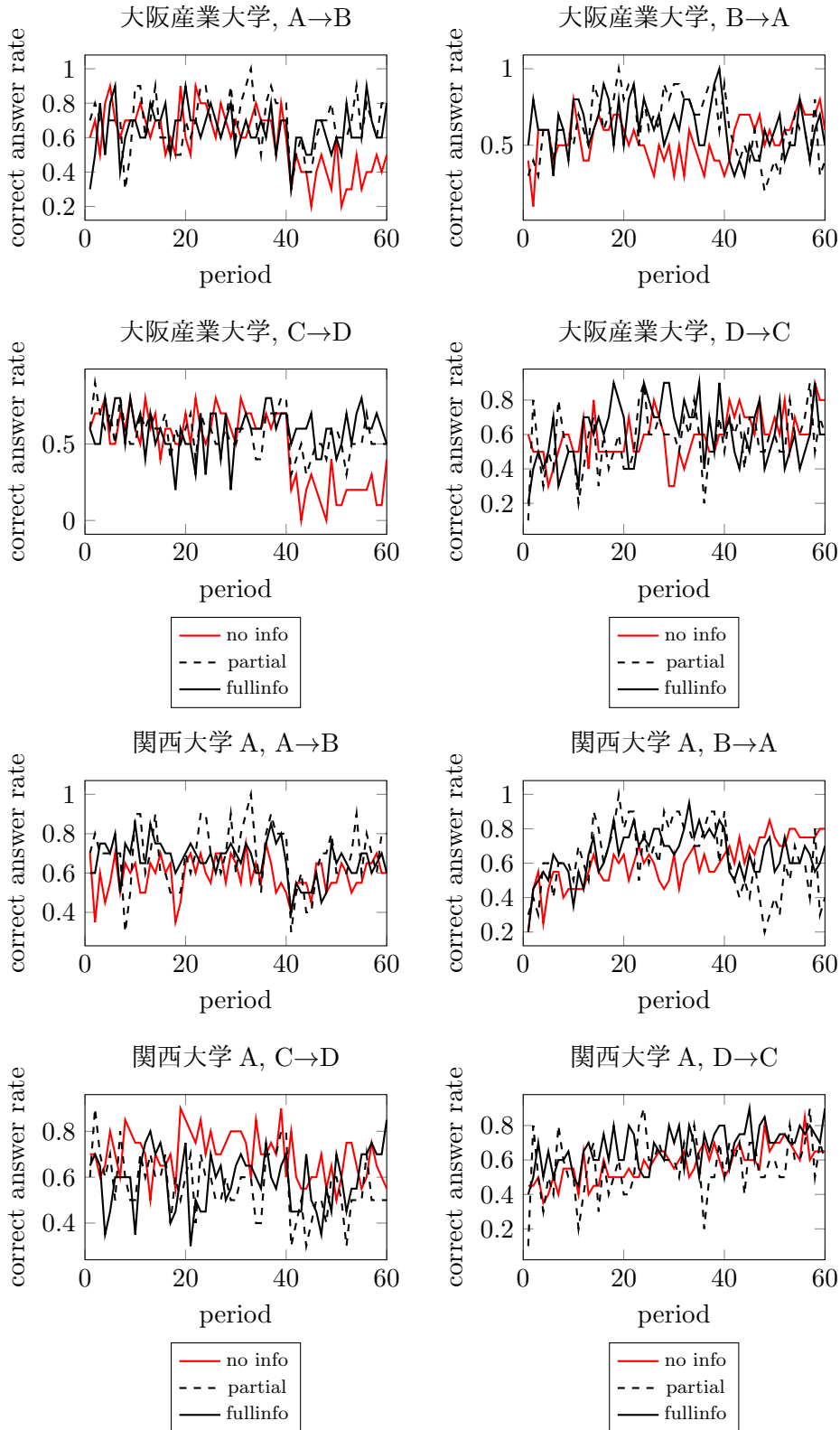
表 25: 同志社大学でのセッション

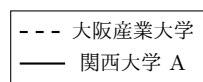
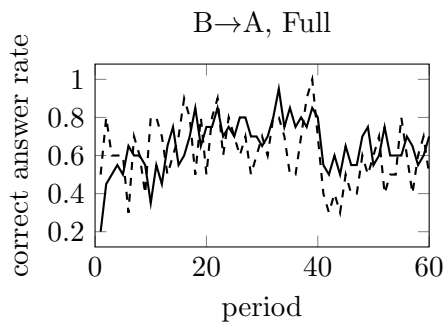
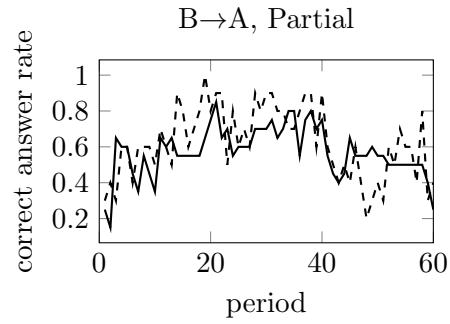
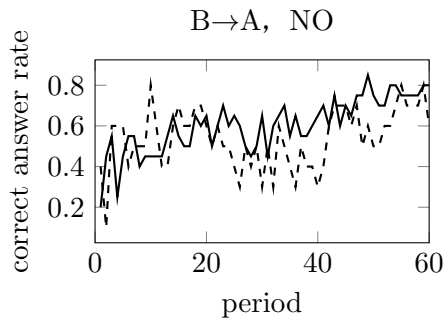
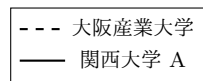
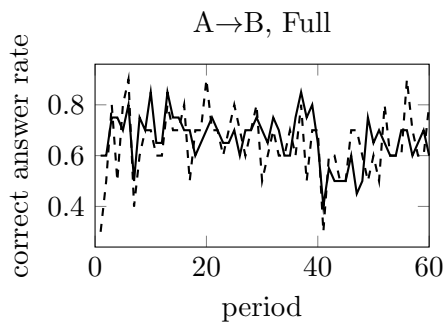
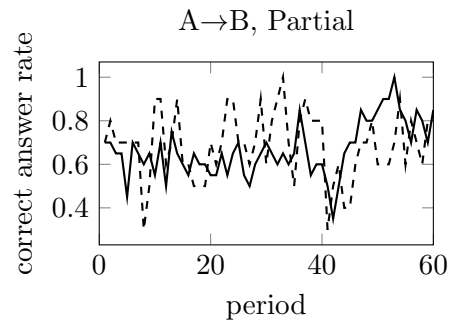
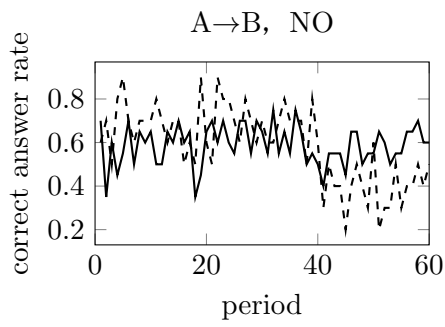
実験実施日	フィードバック	問題の順序	被験者数	平均	最高	最低	標準偏差
2018/07/13	fullinfo	A→B	9	2606	2960	2060	283
2018/07/13	fullinfo		11	2734	2980	2440	159
2018/11/02	no info		12	2721	2960	2380	151
2018/11/02	no info		15	2532	2840	2110	230
2018/11/01	partial		15	2714	2960	2450	134
2018/11/01	parital		13	2687	3080	2460	198
2019/07/12	fullinfo	B→A	10	2372	2790	1750	278
2019/07/12	fullinfo		10	2460	2670	2250	122
2019/07/12	no info		10	2406	2680	2110	171
2019/07/18	no info		10	2596	2960	2260	232
2019/10/17	partial		10	2589	2730	2170	164
2019/10/17	partial		10	2479	2710	2180	169
Total			135	2584	3080	1750	229

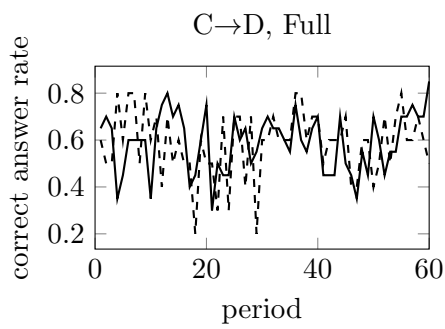
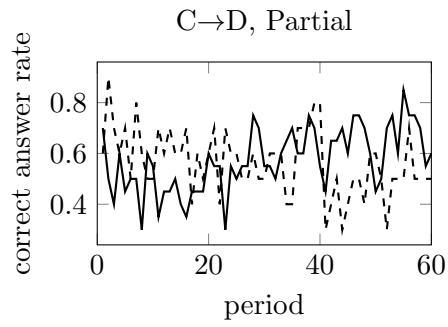
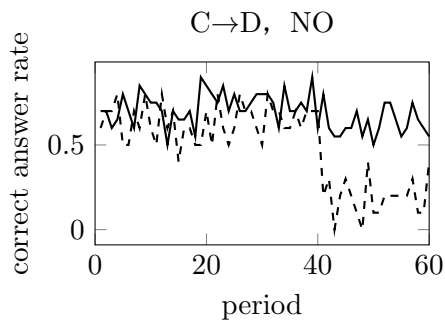
表 26: 広島市立大学でのセッション

実験実施日	フィードバック	問題の順序	被験者数	平均	最高	最低	標準偏差
2018/05/29	fullinfo	A→B	10	2709	2930	2460	172
2018/05/30	fullinfo		11	2628	3030	2070	241
2018/07/04	no info	A→B	10	2602	3080	2230	246
2018/07/25	no info		10	2484	2760	2310	140
2018/10/31	partial	A→B	14	2688	2970	2220	194
2018/10/31	parital		9	2508	3060	2200	231
2018/11/28	partial	B→A	10	2364	2630	2080	181
2018/11/28	partial		10	2554	2770	2240	178
2019/05/29	fullinfo	B→A	10	2448	2700	1930	216
2019/05/29	fullinfo		10	2584	3010	2140	232
2019/06/12	no info	B→A	10	2337	3040	1810	350
2019/06/12	no info		10	2272	2520	2040	179
Total			124	2521	3080	1810	257

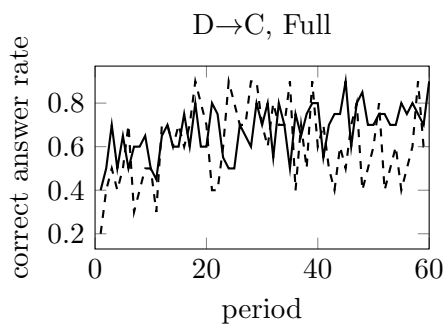
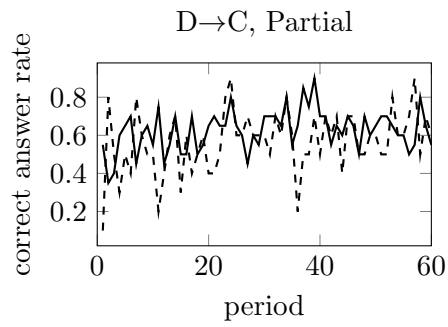
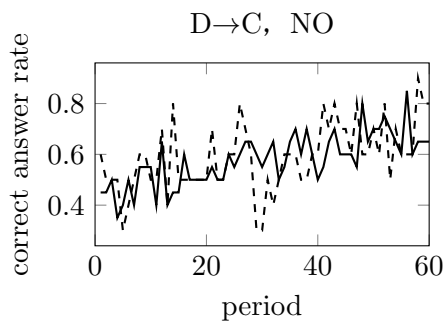
Appendix B: Time Series Plots



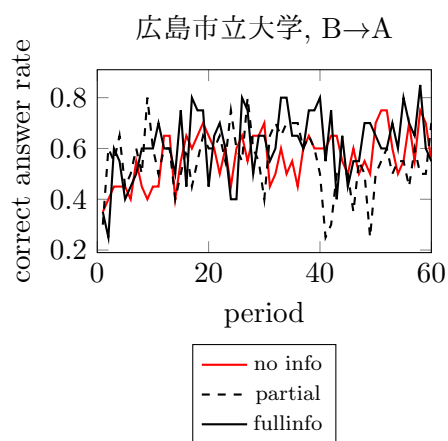
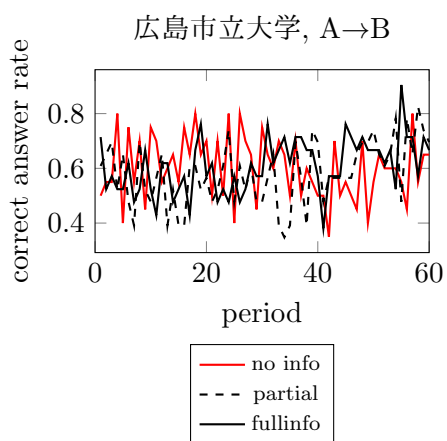
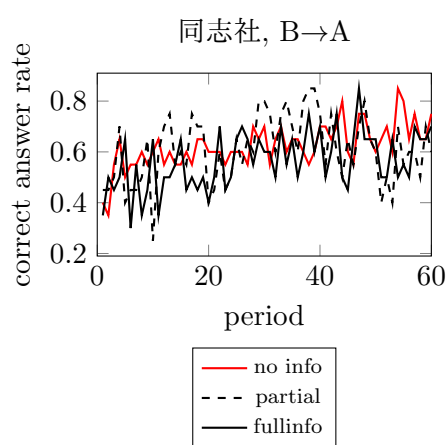
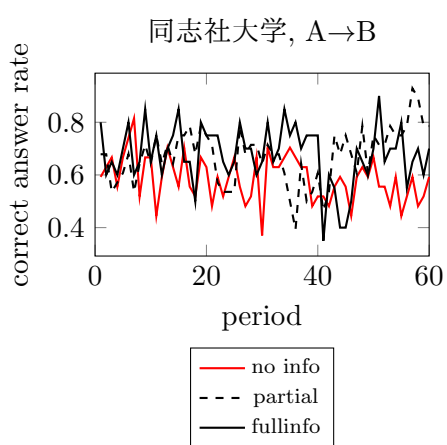
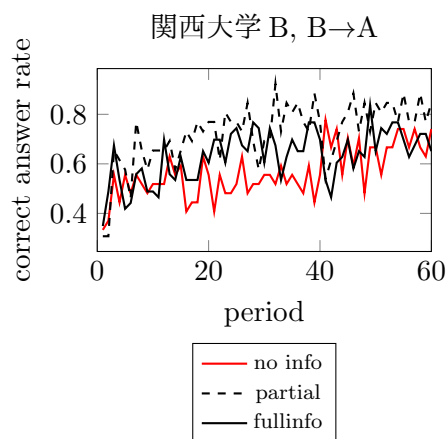
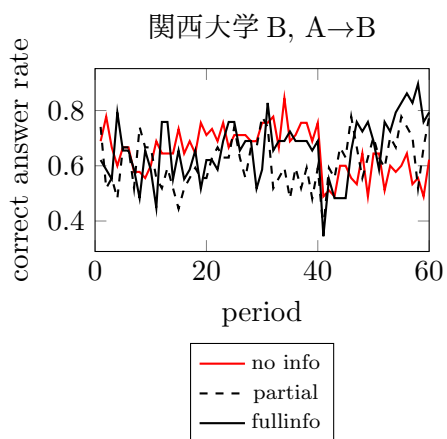


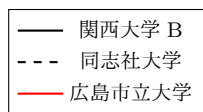
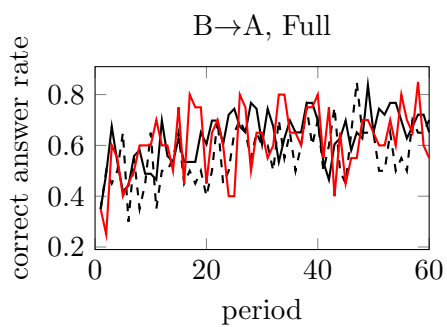
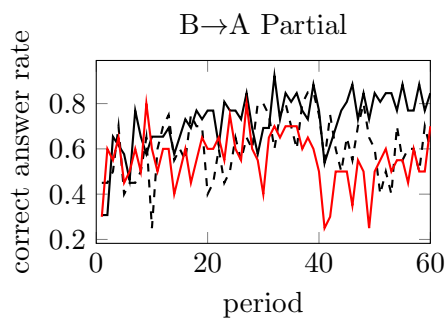
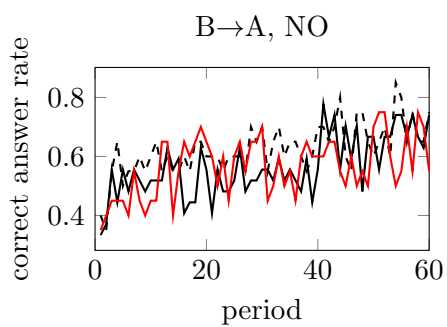
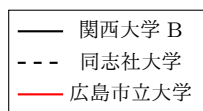
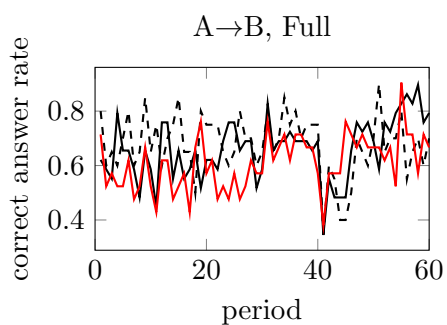
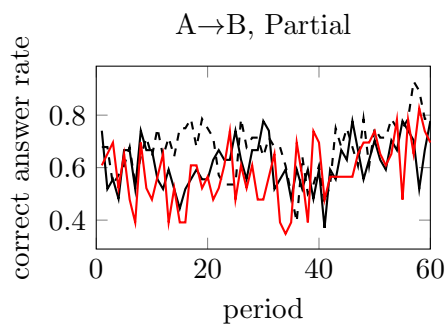
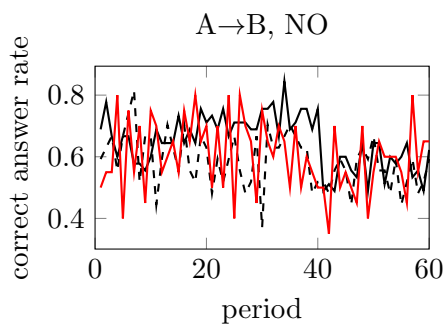


--- 大阪産業大学
— 関西大学 A



--- 大阪産業大学
— 関西大学 A





Appendix C: Win-Stay-Lose-Shift Behavior

表 27: 選択肢の変更頻度 B1 : 関西大学 A (両側フィッシャー検定)

Part-fb (20)				Full-fb (20)			
		0 point	40 points			0 point	40 points
問題 A	freq	23	57	問題 A	freq	26	53
	switch	11	22		switch	10	9
	ratio	0.478	0.386		ratio	0.385	0.170
	p-value		0.464		p-value		0.050
問題 B	freq	42	36	問題 B	freq	46	34
	switch	24	8		switch	24	8
	ratio	0.571	0.222		ratio	0.522	0.235
	p-value		0.010		p-value		0.130
問題 C	freq	24	55	問題 C	freq	20	59
	switch	16	23		switch	12	23
	ratio	0.667	0.418		ratio	0.600	0.390
	p-value		0.053		p-value		0.123
問題 D	freq	37	39	問題 D	freq	42	38
	switch	21	20		switch	28	13
	ratio	0.568	0.513		ratio	0.667	0.342
	p-value		0.653		p-value		0.010

表 28: 選択肢の変更頻度 B1 : 大阪産業大学 (両側フィッシャー検定)

Part-fb (10)				Full-fb (10)			
		0 point	40 points			0 point	40 points
問題 A	freq	16	24	問題 A	freq	10	27
	switch	8	6		switch	3	18
	ratio	0.500	0.250		ratio	0.300	0.667
	p-value		0.176		p-value		0.067
問題 B	freq	18	22	問題 B	freq	18	22
	switch	10	7		switch	10	7
	ratio	0.556	0.318		ratio	0.556	0.318
	p-value		0.200		p-value		0.200
問題 C	freq	10	28	問題 C	freq	8	32
	switch	6	8		switch	4	12
	ratio	0.127	0.286		ratio	0.500	0.375
	p-value		<0.001		p-value		0.691
問題 D	freq	17	23	問題 D	freq	21	17
	switch	13	13		switch	16	4
	ratio	0.764	0.565		ratio	0.762	0.235
	p-value		0.166		p-value		0.010

問題 C では 40 ポイントを得たときの方が選択の変更頻度が高いため、Win-Stay-Lose-Shift 行動をとったとは認められない。

表 29: 選択肢の変更頻度 B8 : 関西大学 A (両側フィッシャー検定)

Part-fb (20)				Full-fb (20)			
		0 point	40 points			0 point	40 points
問題 A	freq	28	72	問題 A	freq	23	77
	switch	14	19		switch	12	13
	ratio	0.500	0.264		ratio	0.522	0.169
	p-value		0.033		p-value		0.010
問題 B	freq	45	55	問題 B	freq	47	53
	switch	13	10		switch	18	5
	ratio	0.289	0.182		ratio	0.383	0.094
	p-value		0.238		p-value		<0.001
問題 C	freq	34	66	問題 C	freq	24	76
	switch	14	21		switch	10	21
	ratio	0.412	0.318		ratio	0.417	0.276
	p-value		0.382		p-value		0.213
問題 D	freq	37	62	問題 D	freq	46	54
	switch	21	46		switch	21	13
	ratio	0.324	0.742		ratio	0.457	0.241
	p-value		<0.001		p-value		0.034

表 30: 選択肢の変更頻度 B8 : 大阪産業大学 (両側フィッシャー検定)

Part-fb (10)				Full-fb (10)			
		0 point	40 points			0 point	40 points
問題 A	freq	18	32	問題 A	freq	11	37
	switch	7	4		switch	5	7
	ratio	0.389	0.125		ratio	0.455	0.189
	p-value		0.041		p-value		0.113
問題 B	freq	21	29	問題 B	freq	18	32
	switch	6	6		switch	8	8
	ratio	0.286	0.207		ratio	0.444	0.250
	p-value		0.738		p-value		0.211
問題 C	freq	13	37	問題 C	freq	17	33
	switch	6	2		switch	8	13
	ratio	0.462	0.054		ratio	0.471	0.394
	p-value		0.002		p-value		0.764
問題 D	freq	27	23	問題 D	freq	17	33
	switch	12	9		switch	6	15
	ratio	0.444	0.391		ratio	0.353	0.455
	p-value		0.779		p-value		0.557

表 31: 選択肢の変更頻度 B9 : 関西大学 A (両側フィッシャー検定)

Part-fb (20)				Full-fb (20)			
		0 point	40 points			0 point	40 points
問題 A	freq	17	63	問題 A	freq	21	59
	switch	7	7		switch	10	6
	ratio	0.412	0.111		ratio	0.476	0.102
	p-value		0.008		p-value		<0.001
問題 B	freq	44	36	問題 B	freq	39	41
	switch	21	7		switch	24	14
	ratio	0.477	0.194		ratio	0.615	0.341
	p-value		0.010		p-value		0.025
問題 C	freq	25	55	問題 C	freq	26	53
	switch	8	10		switch	11	14
	ratio	0.320	0.182		ratio	0.423	0.264
	p-value		0.247		p-value		0.200
問題 D	freq	37	43	問題 D	freq	38	42
	switch	15	14		switch	20	11
	ratio	0.405	0.326		ratio	0.526	0.262
	p-value		0.492		p-value		0.022

表 32: 選択肢の変更頻度 B9 : 大阪産業大学 (両側フィッシャー検定)

Part-fb (10)				Full-fb (10)			
		0 point	40 points			0 point	40 points
問題 A	freq	12	28	問題 A	freq	13	27
	switch	5	7		switch	4	9
	ratio	0.416	0.250		ratio	0.307	0.333
	p-value		0.453		p-value		> 0.999
問題 B	freq	28	12	問題 B	freq	18	21
	switch	12	1		switch	12	3
	ratio	0.429	0.083		ratio	0.667	0.143
	p-value		0.318		p-value		<0.001
問題 C	freq	9	31	問題 C	freq	15	25
	switch	5	15		switch	7	11
	ratio	0.556	0.484		ratio	0.467	0.440
	p-value		>0.999		p-value		> 0.999
問題 D	freq	24	15	問題 D	freq	25	15
	switch	11	3		switch	13	5
	ratio	0.458	0.200		ratio	0.520	0.333
	p-value		0.171		p-value		0.332

表 33: 選択肢の変更頻度 B12 : 関西大学 A (両側フィッシャー検定)

Part-fb (20)				Full-fb (20)			
		0 point	40 points			0 point	40 points
問題 A	freq	25	75	問題 A	freq	22	78
	switch	5	10		switch	3	5
	ratio	0.200	0.133		ratio	0.136	0.064
	p-value		0.518		p-value		0.369
問題 B	freq	36	64	問題 B	freq	53	47
	switch	14	11		switch	16	8
	ratio	0.389	0.172		ratio	0.302	0.170
	p-value		0.029		p-value		0.161
問題 C	freq	30	70	問題 C	freq	22	78
	switch	11	20		switch	10	14
	ratio	0.367	0.286		ratio	0.455	0.179
	p-value		0.482		p-value		0.012
問題 D	freq	45	55	問題 D	freq	51	49
	switch	18	11		switch	18	7
	ratio	0.400	0.200		ratio	0.353	0.143
	p-value		0.045		p-value		0.021

表 34: 選択肢の変更頻度 B12 : 大阪産業大学 (両側フィッシャー検定)

Part-fb (10)				Full-fb (10)			
		0 point	40 points			0 point	40 points
問題 A	freq	9	41	問題 A	freq	16	34
	switch	4	12		switch	5	12
	ratio	0.444	0.293		ratio	0.313	0.353
	p-value		0.442		p-value		>0.999
問題 B	freq	22	28	問題 B	freq	23	26
	switch	9	3		switch	9	6
	ratio	0.409	0.107		ratio	0.391	0.231
	p-value		0.050		p-value		0.352
問題 C	freq	16	34	問題 C	freq	15	35
	switch	5	7		switch	5	9
	ratio	0.314	0.206		ratio	0.333	0.257
	p-value		0.486		p-value		0.733
問題 D	freq	27	23	問題 D	freq	21	29
	switch	8	4		switch	9	6
	ratio	0.296	0.174		ratio	0.429	0.207
	p-value		0.526		p-value		0.251

表 35: 選択肢の変更頻度 B1 : 関西大学 B (両側フィッシャー検定)

Part-fb				Full-fb			
		0 point	40 points			0 point	40 points
問題 A ($n=27$)	freq	38	68	問題 A ($n=29$)	freq	33	83
	switch	23	25		switch	20	33
	ratio	0.605	0.368		ratio	0.606	0.398
	p-value		0.025		p-value		0.062
問題 B ($n=26$)	freq	52	49	問題 B ($n=43$)	freq	89	81
	switch	38	14		switch	51	24
	ratio	0.731	0.286		ratio	0.573	0.296
	p-value		<0.001		p-value		<0.001

表 36: 選択肢の変更頻度 B1 : 同志社大学 (両側フィッシャー検定)

Part-fb				Full-fb			
		0 point	40 points			0 point	40 points
問題 A ($n=28$)	freq	34	77	問題 A ($n=20$)	freq	19	61
	switch	26	27		switch	12	18
	ratio	0.765	0.351		ratio	0.632	0.295
	p-value		<0.001		p-value		0.014
問題 B ($n=20$)	freq	42	38	問題 B ($n=20$)	freq	47	32
	switch	25	20		switch	28	13
	ratio	0.595	0.526		ratio	0.595	0.406
	p-value		0.653		p-value		0.114

表 37: 選択肢の変更頻度 B1 : 広島市立大学 (両側フィッシャー検定)

Part-fb				Full-fb			
		0 point	40 points			0 point	40 points
問題 A ($n=23$)	freq	30	62	問題 A ($n=21$)	freq	20	64
	switch	16	27		switch	12	28
	ratio	0.533	0.435		ratio	0.600	0.438
	p-value		0.504		p-value		0.305
問題 B ($n=20$)	freq	34	45	問題 B ($n=20$)	freq	50	29
	switch	22	21		switch	28	7
	ratio	0.647	0.467		ratio	0.560	0.241
	p-value		0.171		p-value		0.009

表 38: 選択肢の変更頻度 B8 : 関西大学 B (両側フィッシャー検定)

Part-fb				Full-fb			
		0 point	40 points			0 point	40 points
問題 A ($n=27$)	freq	45	90	問題 A ($n=29$)	freq	46	99
	switch	17	30		switch	21	24
	ratio	0.378	0.333		ratio	0.457	0.242
	p-value		0.702		p-value		0.012
問題 B ($n=26$)	freq	54	76	問題 B ($n=43$)	freq	87	123
	switch	25	8		switch	46	22
	ratio	0.463	0.105		ratio	0.529	0.179
	p-value		< 0.001		p-value		< 0.001

表 39: 選択肢の変更頻度 B8 : 同志社大学 (両側フィッシャー検定)

Part-fb				Full-fb			
		0 point	40 points			0 point	40 points
問題 A ($n=28$)	freq	37	103	問題 A ($n=20$)	freq	25	75
	switch	11	38		switch	10	20
	ratio	0.297	0.369		ratio	0.400	0.267
	p-value		0.547		p-value		0.218
問題 B ($n=20$)	freq	41	59	問題 B ($n=20$)	freq	46	52
	switch	17	12		switch	19	14
	ratio	0.415	0.203		ratio	0.413	0.269
	p-value		0.027		p-value		0.142

表 40: 選択肢の変更頻度 B8 : 広島市立大学 (両側フィッシャー検定)

Part-fb				Full-fb			
		0 point	40 points			0 point	40 points
問題 A ($n=23$)	freq	30	85	問題 A ($n=21$)	freq	29	76
	switch	11	34		switch	15	18
	ratio	0.367	0.400		ratio	0.517	0.237
	p-value		0.830		p-value		0.009
問題 B ($n=20$)	freq	48	49	問題 B ($n=20$)	freq	35	65
	switch	22	13		switch	11	14
	ratio	0.458	0.265		ratio	0.314	0.215
	p-value		0.059		p-value		0.335

表 41: 選択肢の変更頻度 B9 : 関西大学 B (両側フィッシャー検定)

Part-fb				Full-fb			
		0 point	40 points			0 point	40 points
問題 A ($n=26$)	freq	35	69	問題 A ($n=43$)	freq	48	120
	switch	17	12		switch	25	22
	ratio	0.486	0.174		ratio	0.521	0.183
	p-value		0.001		p-value		<0.001
問題 B ($n=27$)	freq	57	51	問題 B ($n=29$)	freq	58	58
	switch	29	12		switch	36	16
	ratio	0.509	0.235		ratio	0.621	0.276
	p-value		0.005		p-value		<0.001

表 42: 選択肢の変更頻度 B9 : 同志社大学 (両側フィッシャー検定)

Part-fb				Full-fb			
		0 point	40 points			0 point	40 points
問題 A ($n=20$)	freq	21	59	問題 A ($n=20$)	freq	23	56
	switch	7	21		switch	12	16
	ratio	0.333	0.356		ratio	0.522	0.286
	p-value		>0.999		p-value		0.069
問題 B ($n=28$)	freq	49	63	問題 B ($n=20$)	freq	33	47
	switch	21	13		switch	19	16
	ratio	0.429	0.206		ratio	0.576	0.340
	p-value		0.014		p-value		0.043

表 43: 選択肢の変更頻度 B9 : 広島市立大学 (両側フィッシャー検定)

Part-fb				Full-fb			
		0 point	40 points			0 point	40 points
問題 A ($n=20$)	freq	22	55	問題 A ($n=20$)	freq	26	54
	switch	14	25		switch	12	18
	ratio	0.636	0.455		ratio	0.462	0.333
	p-value		0.208		p-value		0.327
問題 B ($n=23$)	freq	45	47	問題 B ($n=21$)	freq	38	46
	switch	22	7		switch	20	10
	ratio	0.489	0.149		ratio	0.526	0.217
	p-value		<0.001		p-value		0.006

表 44: 選択肢の変更頻度 B12 : 関西大学 B (両側フィッシャー検定)

Part-fb				Full-fb			
		0 point	40 points			0 point	40 points
問題 A ($n=26$)	freq	32	98	問題 A ($n=43$)	freq	46	169
	switch	10	21		switch	21	35
	ratio	0.313	0.214		ratio	0.457	0.207
	p-value		0.339		p-value		0.001
問題 B ($n=27$)	freq	64	71	問題 B ($n=29$)	freq	58	87
	switch	24	12		switch	27	8
	ratio	0.375	0.169		ratio	0.466	0.092
	p-value		0.011		p-value		<0.001

表 45: 選択肢の変更頻度 B12 : 同志社大学 (両側フィッシャー検定)

Part-fb				Full-fb			
		0 point	40 points			0 point	40 points
問題 A ($n=20$)	freq	21	79	問題 A ($n=20$)	freq	34	66
	switch	6	19		switch	18	15
	ratio	0.286	0.241		ratio	0.529	0.227
	p-value		0.778		p-value		0.003
問題 B ($n=28$)	freq	67	73	問題 B ($n=20$)	freq	48	52
	switch	24	15		switch	25	13
	ratio	0.358	0.205		ratio	0.521	0.250
	p-value		0.059		p-value		0.007

表 46: 選択肢の変更頻度 B12 : 広島市立大学 (両側フィッシャー検定)

Part-fb				Full-fb			
		0 point	40 points			0 point	40 points
問題 A ($n=20$)	freq	31	65	問題 A ($n=20$)	freq	26	73
	switch	9	18		switch	14	16
	ratio	0.290	0.277		ratio	0.528	0.219
	p-value		>0.999		p-value		0.005
問題 B ($n=23$)	freq	48	67	問題 B ($n=21$)	freq	48	57
	switch	24	19		switch	16	13
	ratio	0.500	0.284		ratio	0.333	0.228
	p-value		0.021		p-value		0.276

Appendix D: Instruction

次のインストラクションは、レーヴンテストにおける正答数に関係なく、参加謝金を一律500円としたセッションで用いたものである。参加謝金を支払わず、レーヴンテストにおける正答数に応じて追加報酬を支払ってセッションでは、1問の正解につき50円を支払った。

インストラクション

ようこそ。

本日は実験に参加していただきありがとうございます。本日の謝金は、参加者共通の500円の参加謝金に加え、あなたの選択次第で得ることができる0から3200円の追加報酬の合計となります。

注意

- 実験実施者の注意及び指示に従って下さい。
- 実験中は、静かにして、他の参加者と会話したり、メモを交換したり、しないで下さい。また、他の参加者の行動を、のぞき見しないで下さい。
- 実験中は、背もたれによりかかるなど、姿勢をくずさないでください。指定された操作以外、絶対にしないで下さい。
- 携帯電話は電源を切り、使用しないで下さい。
- もし、質問がある場合や、手助けが必要な場合は、静かに手を挙げてください。

実験の説明：パート1

この実験では、画面に提示される二つの選択肢から一つを選んでもらいます。

次の状況を思い浮かべてください。あなたは4人からなる委員会のメンバーの1人です。この委員会では、4人のメンバー各々に120ポイントをどのように配分するかを投票によって決めています。各メンバーには数票の投票権が与えられていますが、メンバーごとに持っている票数が異なることもあります。各メンバーは、そのポイントの分割案を自由に提案することができ、あらかじめ定められた必要票数以上の賛成票を最初に獲得した分配案が採択され、それに基づいて120ポイントが4人に分配されます。ただし、複数の分割案が提案されたら

き、各メンバーは、持っている票を分割して複数の分割案に賛成することはできず、ただ一つの分割案を選択し、それに対して持っているすべての票を投じるものとします。

あなたを含む4人のメンバーそれぞれが持っている票数と、提案された分割案が採択されるために最低でも必要な票数の組み合わせを、委員会構成ということにしましょう。画面には2つの委員会構成が表示されます。あなたに割り当てられる票数は「あなた」という文字の下に表示されます。あなたは、表示された2つの委員会構成から、あなたにとってより便益をもたらすと思われるものを1つ選んでください。

あなたには、合計60回、上述の委員会構成の選択を行ってまいります。制限時間は各回30秒です。制限時間内に、選択を完了しなかった場合、その回のあなたのポイントはゼロです。制限時間内に、選択を完了した場合には、あなたが選んだ委員会構成において、ある理論に従って、コンピュータが自動的に120ポイントを4人に配分します。ただし、同じ委員会構成を選んだとしても、配分は回ごとに異なる可能性があります。

- フィードバックなしトリートメント：各期において、配分が決定されても、あなたは自分に配分されたポイントも、他のメンバーに配分されたポイントも閲覧することができません。しかし、実験の最後に、60回を通じてあなたが得た総ポイントが、1ポイント1円で換算され、謝金に追加されます。
- 部分フィードバックトリートメント：各期の最後に、10秒間、あなたに配分されたポイントのみが画面に表示されます。実験の最後に、60回を通じてあなたが得た総ポイントが、1ポイント1円で換算され、謝金に追加されます。
- 完全フィードバックトリートメント：各期の最後に、10秒間、あなたを含む委員会のメンバー全員に配分されたポイントが画面に表示されます。実験の最後に、60回を通じてあなたが得た総ポイントが、1ポイント1円で換算され、謝金に追加されます。

質問があれば、手を上げてください。

参考文献

- Arechar, A. A., Gächter, S., Molleman, L. 2018. Conducting interactive experiments online *Experimental Economics* 21, 99-131
- Azrieli, Y., Chambers, C.P., Healy, P.J. 2018. Incentives in experiments: A theoretical analysis. *Journal Political Economics* 126, 1472-1503.
- Banzhaf, J.F. 1965. Weighted Voting Doesn't Work: A Mathematical Analysis. *Rutgers Law Review* 19, 317-343.
- Basteck, C., Mantovani, M. (2018) Cognitive ability and games of school choice. *Games and Economic Behavior* 109, 156-183.
- Brunner, E., Munzel, U. 2000. The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal* 42, 17-25.
- Carpenter, P.A., Just, M.A., Shell, P. 1990. What one intelligence test measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review* 97, 404-413.
- Cason, T, N., Saijo, T., Yamato, T. 2002. Voluntary participation and spite in public good provision experiments: an international comparison. *Experimental Economics* 5, 133-153.
- Deegan, J., Packel, E. 1978. A New Index of Power for Simple n-person Games. *International Journal of Game Theory* 7, 113-123.
- Ernst, M.D. 2004. Permutation methods: A basis for exact inference. *Statistical Science* 19, 676-685.
- Felsenthal, D.S., Machover, M. 1998. *The Measurement of Voting Power : Theory and Practice, Problems and Paradoxes*. Edward Elgar, London.
- Fischbacher, U. 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics* 10, 171-178.
- Fisher, R.A. 1935. *The Design of Experiments*, 1st ed., (Hafner Pub. Co., 1971).
- Gill, D., Prowse, V. 2016. Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economics* 124, 1619-1676.

- Greiner, B. 2003. Online Recruitment System for Economic Experiments (ORSEE). <http://www.orsee.org/web>
- Guerci, E., Hanaki, N., Watanabe, N. 2017. Meaningful Learning in Weighted Voting Games: An Experiment. *Theory and Decision* 83, 131-153.
- Guerci, E., Hanaki, N., Watanabe, N., Esposito, G., Lu, X. 2014. A Methodological Note on a Weighted Voting Experiment. *Social Choice and Welfare* 43, 827-850.
- Hanaki, N. Cognitive ability and observed behavior in laboratory experiments: Implications for macroeconomic theory. forthcoming in *Japanese Economic Review*.
- Hanaki, N., Akiyama, E., Funaki, Y., Ishikawa, R. 2017. Diversity in cognitive ability and mispricing in experimental asset markets. GREDEG Working Paper No. 2017-08
- Hanaki, N., Jacquemet, N., Luchini, S., Zylbersztejn, A. 2016. Cognitive ability and the effect of strategic uncertainty. *Theory and Decision* 81, 101-121.
- Hanaki, N., Jacquemet, N., Luchini, S., Zylbersztejn, A. 2016. Fluid intelligence and cognitive reflection in a strategic environment: evidence from dominance-solvable game. *Frontiers in Psychology: Personality and Social Psychology* 10 August 2016. <http://dx.doi.org/10.3389/fpsyg.2016.01188>
- Heckman, J. J., Ichimura, H., Smith, J., Todd, P. 1998. Characterizing selection bias using experimental data. *Econometrica* 66, 1017-1098.
- Kawamura, T., Ogawa, K. 2019. Cognitive ability and human behavior in experimental ultimatum games. *Research in Economics* 73, 97-106.
- Montero, M., Sefton, M., Zhang, P. 2008. Enlargement and the Balance of Power: An Experimental Study. *Social Choice and Welfare* 30, 69-87.
- Neubert, K., Brunner, E. 2007. A studentized permutation test for the non-parametric Behrens-Fisher problem. *Computational Statistics & Data Analysis* 51, 5192-5204.
- Neugebauer, T., Perote, J., Schmidt, U., Loos, M. (2008). Selfish-biased conditional cooperation: on the decline of contributions in repeated public goods experiments. *Journal of Economic Psychology* 30, 52-60.

- Nowak, M.A., Sigmund, K. (1993). A strategy of win stay, lose shift that outperforms tit-for-tat in the prisoner's-dilemma game. *Nature* 364, 56-58.
- Proto, E., Rustichini, A., Sofianos, A. 2019. Intelligence, personality and gains from cooperation in repeated interactions. *Journal Political Economy* 127; 1351-1390.
- Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S. 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *American Economic Review* 81, 1068-1095.
- Rick, S., Weber, R.A. 2010. Meaningful Learning and Transfer of Learning in Games Played Repeatedly without Feedback. *Games and Economic Behavior* 68, 716-730.
- Rosser, J.B. Jr., Eckel, C. 2010. Special Issue on the Methodology of Experimental Economics. Rosser, J.B. Jr., Eckel, C. eds., *Journal of Economic Behavior and Organization* 73, 1-132.
- Shapley, L.S., Shubik, M. 1954. A Method for Evaluating the Distribution of Power in a Committee System. *American Political Science Review* 48, 787-792.
- Watanabe, N. 2018. Supplementary Results for "Meaningful Learning in Weighted Voting Games: An Experiment". Kansai University RISS Discussion Paper Series No. 62, June 2018
- Watanabe, N. 2019. A Mouse-tracking Experiment for Learning in Weighted Voting. mimeo.
- Watanabe, N., Kawamura, T., Ogawa, K. 2020. A School Choice Experiment: Cognitive Ability and Information. mimeo.
- Weber, R.A. 2003. Learning with no Feedback in a Competitive Guessing Game. *Games and Economic Behavior* 44, 134-144.