ソシオネットワーク戦略ディスカッションペーパーシリーズ　ISSN 2434-9445
第 97 号　2021 年 12 月
RISS Discussion Paper Series
No.97　December , 2021

# Behavioral change and cognitive ability in repeated prisoner’s dilemma game experiments

Tetsuya Kawamura, Takahiro Miura,
Misato Inaba & Kazuhito Ogawa

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: https://www.kansai-u.ac.jp/riss/index.html
e-mail: riss@ml.kandai.jp
tel. 06-6368-1228
fax. 06-6330-3304

# Behavioral change and cognitive ability in repeated prisoner's dilemma game experiments*

Tetsuya Kawamura[†], Takahiro Miura[‡], Misato Inaba[§] & Kazuhito Ogawa[¶]

**Abstract**

We examine whether cognitive ability affects the cooperative behavior of non-student participants in repeated prisoner's dilemma experiments. The results show that high cognitive ability participants (1) cooperate more in the presence of the infinitely repeated condition (IFR) than the finitely repeated condition (FR), (2) cooperate more as the expected number of stage games in a round increases under the IFR, and (3) do not cooperate more as the number of stage games in a round increases under the FR. The strategy frequency estimation in the case of IFR suggests that for high cognitive ability, when the expected number of stage games increases, the share of grim trigger strategy increases while the defective strategy is less employed. Thus, high cognitive ability participants change their behavior according to the situation. Additionally, the behavior of low cognitive ability participants do not seem to change; the cooperation rate remains unchanged despite the conditions. However, the strategy frequency estimation in the case of IFR reveals a slight change in the participants ' behavior, based on the situation. Overall, this reveals a low cooperation rate.

Keywords: Behavioral change, (in)finitely repeated prisoner's dilemma, laboratory experiment, cognitive ability, non-student participants

JEL codes: C91, C65, D91

---

*This study was supported by the MEXT-Supported Program for Strategic Research Foundation at Private Universities, 2014-2018 (S1411035) and 18H00869 from the Japan Society for the Promotion of Science. We thank Prof. Hajime Kobayashi (Kansai U.) and Prof. Kazumi Shimizu (Waseda U.)for their helpful comments at the conference on Experimental Social Science in 2018. We would also like to thank Prof. Nobuyuki Hanaki (Osaka U.)for the suggestions with regard to the analysis conducted in the study at the conference in WEHIA'19.

†Research Institute for Socionetwork Strategies (RISS), Kansai University, 3-3-35, Yamate-cho, Suita, Osaka, Japan, 564-8680, Faculty of Economics and Business Management, Tezukayama University, 7-1-1, Tezukayama, Ikoma, Nara, Japan, 150-0031.

‡RISS & PERSOL CAREER CO., LTD., Marunouchi Building, 27F, 2-4-1 Marunouchi, Chiyoda, Tokyo, 100-6328, Japan.

§RISS & Faculty of Economics, Kindai University, 228-3 Shin-Kami-Kosaka, Higashi-Osaka, Osaka, 577-8502, Japan.

¶Corresponding author, RISS & Faculty of Sociology, Kansai University, 3-3-35, Yamate-cho, Suita, Osaka, Japan, 564-8680. Email: kz-ogawa@kansai-u.ac.jp

# 1 Introduction

A growing number of experimental studies (Brañas-Garza et al., 2012; Carpenter et al., 2013; Gill and Prowse, 2016; Kawamura and Ogawa, 2019; Ben-Ner et al., 2004; Chen et al., 2013; Yamagishi et al., 2016; Inaba et al., 2018; Hanaki et al., 2016; Proto et al., 2019) reveal the significant effect of cognitive ability on economic decision-making. Here, cognitive ability is defined as a fluid intelligence that is related to "the capacity to think logically, analyze and solve novel problems, independent of background knowledge" (Mullainathan and Shafir, 2013, p.48).

Prior studies exploring this effect are generally grouped into two categories. Studies in the first category examine the effect of cognitive ability on single agent decision-making, such as time and risk preferences. Frederick (2005) find that a Cognitive Reflection Test (CRT) score predicts time and risk preferences. Benjamin et al. (2013) find that small-stakes risk aversion and short-run discounting are less common among those with higher standardized test scores, and elementary-school GPA predicts the preferences measured at the end of high school.

Studies in the second category investigate the relationship between cognitive ability and decision-making in economic games. This category consists of two sub-categories. The first sub-category focuses on games with unique subgame perfect equilibrium. In $p$-beauty contest game experiments, a participant with higher cognitive ability chooses a closer number to the sub-game perfect equilibrium (Brañas-Garza et al., 2012; Carpenter et al., 2013; Gill and Prowse, 2016). In ultimatum game (UG) experiments, the increase in cognitive ability lowers the minimum acceptance offer of a responder when he/she decides with the strategy method (Kawamura and Ogawa, 2019). In dictator game (DG) experiments, the higher the cognitive ability, the less a dictator donates (Ben-Ner et al., 2004; Chen et al., 2013; Yamagishi et al., 2016; Inaba et al., 2018; Ogawa et al., 2020), but Brandstätter and Güth (2002) do not observe this relationship.

In finitely repeated PD game experiments,[1], which belongs to the first sub-category, Al-Ubaydli et al. (2016)[2] and Barreda-Tarrazona et al. (2017) find that cognitive ability does not affect individual cooperation or, even if it affects, the effect is weak. Al-Ubaydli et al. (2016) also find that paired cooperation is positively correlated with the average cognitive ability of the two players.

[1] Selten and Stoecker (1986) find that experienced participants cooperate until shortly before the end of a super game and then deviate. Andreoni and Miller (1993) experimentally investigate cooperation in this setting, and consider a series of finitely repeated PD games, in which they manipulate participants' beliefs about their opponent's type. Their results indicate that the sequential equilibrium reputation model predicts cooperative behavior well in a finitely repeated PD setting. Bó (2005) finds that the cooperation rate is lower in this case than in the infinitely repeated setting.

[2] To measure cognitive ability, Al-Ubaydli et al. (2016) employed Raven's Standard Progressive Matrices Plus, a version between the Standard Progressive Matrices and the Advanced Progressive Matrices.

The second sub-category focuses on games with multiple equilibrium, especially, infinitely repeated PD game experiments[3]. Jones (2008) performs meta-analysis of repeated PD games and finds, as SAT scores increase by 100, the cooperation rate increases by 5–8%.

Proto et al. (2019) studied university students and experimentally investigated the effects of cognitive ability on cooperative behavior in infinitely repeated PD games where cooperative strategy is supported as a subgame perfect equilibrium and risk dominance. The scores of Raven's Advanced Progressive Matrices (APM) test (Raven, 1936), which measures abstract reasoning and is regarded as a non-verbal estimate of fluid intelligence, was used to measure cognitive ability. Participants were divided into high and low score groups, and they played infinitely repeated PD games with a partner from the same group. Their results indicate that the cooperation rate of the high score group is higher than that of the low score group when the continuation probability, $\delta_H = 0.75$. When $\delta_L = 0.5$, the cognitive ability does not affect behavior. The cooperation rate of each group is higher in $\delta_H$ than in $\delta_L$. Their strategy estimation in $\delta_L$ shows that the major strategy is "always defect" (all-D)in both groups. The same estimation in $\delta_H$ shows that the major strategies employed by the high score group are tit-for-tat (TFT) and a grim trigger, while the major ones among low score group are "always defect" and a grim trigger.

Considering few studies have compared whether the effects of cognitive ability on cooperative behavior differ between finitely repeated and infinitely repeated conditions, we experimentally investigate the relationship between cognitive ability measured by scores of the Raven's APM test, which is also employed by recent studies (Proto et al., 2019; Hanaki et al., 2016; Kawamura and Ogawa, 2019; Ogawa et al., 2020) and cooperative behavior in repeated PD games. We studied non-student participants because we wanted to collect a wide range of cognitive ability in order to find effects of their cognitive ability on their cooperative behavior.

Focusing on finitely and infinitely repeated PD games, we first confirm the robustness of the results in prior studies. The infinitely repeated condition has cooperative equilibria and the existence of risk dominance should be considered for equilibrium selection (Blonski et al., 2011; Fudenberg et al., 2012; Blonski and Spagnolo, 2015). On the contrary, the equilibrium of the finitely repeated condition is always defection. Second, comparing

[3]There are two waves of infinitely repeated PD games experiments with perfect monitoring (Bó and Fréchette, 2018). In the first wave, the possible benefits of cooperation are not fully exploited because the increase in cooperation is small. Roth and Murnighan (1978) find that the higher the continuation probability is, the higher the cooperation rate. Holt (1985) conducts a Cournot duopoly experiment and compares decision-making in a treatment where the random-stopping rule ($\delta = 5/6$) is introduced from the end of the seventh stage and a one-shot game treatment. Holt (1985) finds that the former treatment is more collusive than the latter. Feinberg and Husted (1993) set $\delta = 5/6$ and conduct the PD game experiment with a payoff decline, finding that cooperation is less likely to emerge when this decline is high.

The second wave begins with Bó (2005) providing more positive evidence that the shadow of the future can increase cooperation by conducting one-shot, finitely repeated, and infinitely repeated PD game experiments with $\delta = 0.5$ and $0.75$. Duffy and Ochs (2009) find that the cooperation rate in an infinitely repeated game increases more under fixed matching than under random matching.

the results in both conditions enables us to examine how non-students with high (low) cognitive ability change their behavior according to a repetition condition.

From our results, compared to the finitely repeated condition, high score participants behave as the standard repeated game theory expects. Therefore, they change their behavior according to conditions. However, considering that the cooperative rate of those with low scores remains unchanged, they do not seem to change their behavior according to conditions.

To investigate the behavioral changes further, we estimated the participants' strategies in the infinitely repeated condition. The estimation indicates that the participants with high scores use a Tit-For-Tat (TFT) strategy when continuation probability ($\delta$) is high and the all-D strategy when this probability is low. The participants with low scores employ all-D when $\delta$ is low, while they employ trigger or all-D when $\delta$ is high. This suggests that they partially change their behavior according to circumstances but the combination of major strategies could not reach cooperation.

Our study adds two contributions to the literature. The first contribution is a robustness check. In most experimental studies of repeated game theory, university students from Western societies were participants. In this study, we used Japanese adults as participants to conduct this kind of experiment. Using them enables us to examine the effects of age, gender, personality trait and cognitive ability. Our results incorporate these effects and confirm the robustness of previous research results.

Our second contribution is that we add new insights into the relationship between cognitive ability and cooperative behavior. Whereas, Proto et al. (2019) used the infinitely repeated setting and Al-Ubaydli et al. (2016) used the finitely repeated setting to investigate whether cognitive ability affects the cooperation rate, we used both repeated settings. By comparing the infinitely and finitely repeated settings, it will be clearer how those with high (or low) cognitive ability cooperate.

The remainder of this paper is organized as follows: section 2 develops our hypotheses, section 3 explains the experimental procedure, section 4 shows the experimental results, and section 5 explores the behavioral differences between participants with low and high cognitive abilities. Section 6 discusses these results. Section 7 concludes the paper.

## 2 Hypotheses

We calculate the theoretical conditions that the trigger strategy is the sub-game perfect equilibrium in the infinitely repeated PD game with a discount factor of $\delta$. In the experiments, we employ the payoff matrix in

Table 1, which is PD2 in Bó (2005).

<Table 1 here >

When $\delta \geq 5/11$, the trigger strategy is the sub-game perfect equilibrium in the infinitely repeated PD game.

We then introduce risk dominance as an equilibrium selection criterion, according to Blonski et al. (2011), Fudenberg et al. (2012), and Blonski and Spagnolo (2015). Using the reduced form of the payoff matrix (Table 2), when

$$\delta \geq \frac{2}{3},$$

the trigger strategy is a subgame perfect equilibrium and risk dominant, while the all-D strategy is a sub-game perfect equilibrium. If

$$\delta \leq \frac{5}{11},$$

the all-D strategy is a sub-game perfect equilibrium and risk dominance. If

$$\frac{5}{11} \leq \delta \leq \frac{2}{3},$$

the trigger strategy is a sub-game perfect equilibrium, but the all-D strategy is a sub-game perfect equilibrium and risk dominant.

<Table 2 here>

In the infinitely repeated (IFR) treatment, we set the discount factor as $\delta = 0.5$ (IFR2) or $\delta = 0.75$ (IFR4). In IFR2 and IFR4, a round has two and four-stage games on average, respectively. In IFR2, employing the trigger strategy is the sub-game perfect equilibrium but not risk dominant. In IFR4, this strategy is the sub-game perfect equilibrium and risk dominant.

We define the cooperation rate of a participant $i(= 1, ..., N)$ in a treatment as

$$C(\text{treatment})_i = \frac{\text{the number of choosing cooperation in a session by } i}{\text{total number of decision-makings by } i.}.$$

The average cooperation rate in a treatment is

$$C(\text{treatment}) = \frac{\sum_{i=1}^{N} C(\text{treatment})_i}{N}.$$

The discount factor level and the risk dominance (Bó, 2005; Bó and Fréchette, 2011; Blonski et al., 2011; Blonski and Spagnolo, 2015) affect decision-making. As the discount factor increases, people will cooperate more. If the cooperative equilibrium is risk dominant, people will cooperate more. Additionally, in IFR4, this strategy is also risk dominant.

We also conduct finitely repeated PD game experiments (FR2 and FR4). In treatments FR2 and FR4, a round ends at the second and the fourth stage, respectively.

According to the findings of Proto et al. (2015, 2019), higher cognitive ability brings cooperative behavior in the infinitely repeated PD game experiments. The cooperative behavior of the participants with lower cognitive ability remains the same. Applying the findings, participants with high cognitive ability will change their behavior according to the situation. This indicates they will cooperate more in IFR4 than in IFR2 treatments. Those with low cognitive ability will not change their behavior according to the situation, indicating that their cooperation rate will remain unchanged between IFR2 and IFR4.

We now propose the hypotheses for IFR treatments. According to the equilibrium prediction of the repeated game, $C(\text{IFR4}) \geq C(\text{IFR2})$ is a natural hypothesis. Theoretically, this hypothesis will be supported regardless of the cognitive ability level.

Therefore, we examine the following hypotheses.

- 1(a) For a participant with high cognitive ability, $C(\text{IFR4}) \geq C(\text{IFR2})$
- 1(b) For a participant with low cognitive ability, $C(\text{IFR4}) \geq C(\text{IFR2})$

In FR treatments, participants will not change their behavior regardless of the cognitive ability level, because the sub-game perfect equilibrium is always defection. Considering Al-Ubaydli et al. (2016) and Barreda-Tarrazona et al. (2017) report that cognitive ability does not affect individual cooperative behavior in the finitely repeated PD game experiments, the cooperation rate will remain unchanged between FR2 and FR4 treatments regardless of the cognitive ability level.

Thus, we propose hypotheses 2(a) and 2(b).

- 2(a) For a participant with high cognitive ability, $C(\text{FR2}) = C(\text{FR4})$
- 2(b) For a participant with low cognitive ability, $C(\text{FR2}) = C(\text{FR4})$

Here, we propose hypotheses about the comparison of cooperation rates between FR and IFR treatments. When the average number of stage games in a round is the same between FR and IFR treatments, the cooperation rate will be higher in IFR than in FR. This hypothesis will be applied to all the participants, regardless of the cognitive ability level.

Thus, we propose the following hypotheses.

- 3(a) For a participant with high cognitive ability, $C(\text{IFR4}) > C(\text{FR4})$
- 3(b) For a participant with low cognitive ability, $C(\text{IFR4}) > C(\text{FR4})$
- 4(a) For a participant with high cognitive ability, $C(\text{IFR2}) > C(\text{FR2})$
- 4(b) For a participant with low cognitive ability, $C(\text{IFR2}) > C(\text{FR2})$

## 3 Experimental procedure

To recruit participants, we asked a leaflet distributing company to distribute a flyer to recruit participants from residential areas, within a 20 km radius from the Center for Experimental Economics at Kansai University, which is in the northern part of Osaka prefecture, Japan. Those who wanted to participate in the experiment applied in advance through an online application form, telephone, email, or letter. Additionally, we sent an e-mail and/or direct mail to participants from experiments other than PD and invited them to participate. Each participant took part in this type of experiment only once. Recruiting non-student participants enabled us to collect behavioral data from participants with various cognitive ability levels.

Each participant provided informed, written consent before beginning the experiment. A series of experiments was approved by the Institutional Review Board of the Center for Experimental Economics at Kansai University (No.2016018, No.2017002).

In each treatment,[4] all participants were provided with instructions on the game structure. Each treatment consisted of multiple identical rounds. Each round consisted of one or more stage games. In each stage, participants independently and simultaneously chose A or B. For each round, participants were put into pairs so that they never played with participants they had previously been paired with to omit contagion effects as much as possible, which could be caused if a pair was fixed throughout the experiment.

[4]The experiment was computerized using z-Tree (Fischbacher, 2007).

In IFR treatments, the following information was provided with the instructions. The experimenter chose a card randomly from four playing cards at the end of each stage. When $\delta = 0.5$ $(0.75)$, two jokers (one joker) and two (three) aces were used for the random choice; there was a probability of $1/2$ $(3/4)$ that all participants would proceed to the next stage and a probability of $1/2$ $(1/4)$ that they would stop at the current stage. These probabilities remained unchanged throughout the session. A round lasted until the experimenter chose a joker. The result of the random choice is common information as the experimenter draws a card in front of all participants.

In FR treatments, a participant received instructions that the number of stage games in a round was two (four) and the number of rounds was ten (five).

After the experiment, participants were given 10 minutes to answer a questionnaire that measured their cognitive ability, which contained 16 of the 48 questions in Raven's APM test (Raven, 1936) version.12. We used the APM version, composed of 48 questions in total (in Set I and Set II). In this experiment, the participants were required to complete the answers in about 10 minutes. Accordingly, we randomly selected 16 of 48 questions that they were asked to complete after conducting the RPD experiment. In this paper, the Raven score refers to the number of correct answers to the 16 questions on the APM version. We included questions 1, 4, 7, and 10 from Set I and questions 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, and 34 from Set II.

After that, they also answered ten Big Five personality questions (Oshio et al., 2012) in 10 minutes.

Participants received a participation reward of JPY 1,000 [5] and a performance-based reward, which was the total payoff in the experiment, where 1 point = JPY 1. The scores for Raven's APM test did not carry a monetary reward. Tables 3 and 4 show session profiles.

<Table 3 here>

We defined a participant with high cognitive ability if his or her score is higher than the median score, which is 8 in Table 4. Otherwise, a participant is categorized as having low cognitive ability. We referred to them as high score group (HSG) or low score group (LSG), respectively.

<Table 4 here>

[5] At the time of the experiment, 1 USD = 110 JPY.

## 4 Experimental Results

### 4.1 Overview

First, we observe and investigate the predictions from the repeated PD games. Table 5 indicates the individual cooperation rate in each treatment. (1) The difference in cooperation rates between IFR2 and IFR4 is 8% (the Mann-Whitney's U and two-sided t-tests, for both tests $p < 0.001$), (2) the difference in the cooperation rates between FR2 and FR4 is approximately 3% (for both tests, $p < 0.05$), (3) the difference in the cooperation rate of FR2 and IFR2 is 4.5 % (for both tests, $p < 0.001$), and (4) the difference between the cooperation rate of FR4 and IFR4 is about 10% (for both tests, $p < 0.001$).

Second, we explain the logit regression results. In the regressions shown in Tables 12, 13, 14 and 15, we employ the following equations:

$$
\begin{aligned}
\log\left[\frac{Prob(Choice_{i,r,s}=1)}{1-Prob(Choice_{i,r,s}=1)}\right] &= \beta_0 + \beta_1\text{Treatment dummy} + \beta_2\text{Stage} + \beta_3\text{Round} + \beta_4\text{Age} \\
&+ \beta_5\text{Gender dummy} + \beta_6\text{Age} \times \text{Gender} + \beta_7\text{Extroversion} \\
&+ \beta_8\text{Neuroticism} + \beta_9\text{Openness} + \beta_{10}\text{Agreeableness} \\
&+ \beta_{11}\text{Conscientiousness} + S \times \beta_{12}\text{POC} + \epsilon_{t,r,s}
\end{aligned} \tag{1}
$$

Here, $i$, $r$ and $s$ indicate the ID of participant, the round number and the stage number, respectively. Treatment dummy is Random or Four in Table 11. The POC is the choice in the previous stage of the counterpart. $S$ is 1 if we consider the effect of the previous choice by the counterpart, otherwise, 0. Extroversion, Neuroticism, Openness, Agreeableness and Conscientiousness are personality traits referred to as "Big five" (Oshio et al., 2012).

Model (1) in Table 12 indicates that the odds ratio of the dummy variable 'Four' is significant. Thus, the cooperation rate is significantly higher in IFR4 than in IFR2 when we employ all the observations in IFR2 and IFR4. This is true when we consider the decision-making in the previous stage of the couterpart (Model (4) in Table 12). Model (1) in Table 13 indicates that the odds ratio of 'Four' is not significant when we employ all the observations in FR2 and FR4. This holds true when we consider the decision-making in the previous stage of the couterpart (Model (4) in Table 13). Models (1) and (4) in Table 14 indicate that the odds ratio of 'Random' is significant. The Models (1) and (4) in Tables 15 indicate that the odds ratio of 'Random' is significant. These results indicate that the predictions from the repeated PD games are supported.

<Table 5 here>

<Table 6 here>

Table 6, which focuses on the 1st stage cooperation, indicates the following results: (1) the cooperation rate in the 1st stage is significantly higher in IFR4 than in IFR2 ($p < 0.01$, the Mann-Whitney test), (2) the cooperation rate in the 1ˆst stage is significantly higher in FR4 than in FR2 ($p < 0.01$, the Mann-Whitney test), (3) the cooperation rate in the 1st stage is significantly higher in IFR4 than in FR4 ($p < 0.01$, the Mann-Whitney test), and (4) the cooperation rate in the 1st stage is significantly higher in IFR2 than in FR2 ($p = 0.054$, the Mann-Whitney test). Except for (2), the predictions from the repeated PD games are supported.

<Table 7 here>

<Table 8 here>

<Table 9 here>

<Table 10 here>

### 4.2 Hypotheses test [6]

Before we investigate the hypotheses, we compare the cooperation difference between the HSG and LSG in each treatment. In IFR treatments (Table 7), the cooperation rate of the HSG is significantly higher than that of the LSG (two-sided t-test and Mann-Whitney's U test, $p < 0.01$). In IFR4, the same result holds (both tests, $p < 0.01$). In IFR2, the same result mildly holds (both tests, $p < 0.10$). This is in line with Proto et al. (2015). In FR treatments (Table 9), the cooperation rate of the LSG is significantly higher than that of the HSG (two-sided t-test and Mann-Whitney's U test, $p < 0.01$). In FR2, the same result holds. While in FR4, this rate of the LSG is not significantly higher than that of the HSG.

[6]In this study, we classify participants according to their scores; a participant with a score higher (lower) than the median (=8) belongs to a high (low)score group. However, if we classify a participant who has a score larger than 10, top 25% score, into a high score group, or otherwise a low score group, the investigation results remain almost unchanged. See Appendix B.

Tables 7 show the cooperation rate of the HSG is higher in IFR4 than in IFR2 and that this rate of the LSG is almost the same between IFR2 and IFR4. We examine hypotheses 1(a) and 1(b) by using simple statistical tests. The cooperation rate of the HSG is significantly higher in IFR4 than in IFR2 (Mann-Whitney's U and t-tests, $p < 0.05$), indicating hypothesis 1(a) is supported. The cooperation rate of the LSG is not significantly higher in IFR4 than in IFR2, indicating hypothesis 1(b) is not supported.

Tables 8 indicate that the 1st stage average cooperation rate of the HSG is higher in IFR4 than in IFR2 ($p < 0.01$, Mann-Whitney's U and t-tests) and that this rate of the LSG is (1) higher in IFR4 than in IFR2 ($p < 0.05$, Mann-Whitney's U and t-tests). Focusing on the 1st stage behavior, hypothesis 1(a) and 1(b) are supported.

<Table 11 here>

<Table 12 here>

In this section, we investigate the results of logit regressions (equation (1)). Models (2) and (5) in Table 12 show the result of decision-making by the HSG. This indicates that hypothesis 1(a) is supported; the cooperation rate is significantly higher in IFR4 than in IFR2 (the odds ratio of "Four" is greater than 1 and significant).

Models (3) and (6) in Table 12 show the result of decision-making by the LSG. Hypothesis 1(b) is not supported; the cooperation rate is not significantly higher in IFR4 than in IFR2 (the odds ratio of "Four" is not significant).

Using Table 9, we examine hypotheses 2(a) and 2(b) by using simple statistical tests; The cooperation rate of the HSG is significantly higher in FR4 than in FR2 (Mann-Whitney's U and t-tests, $p < 0.01$), indicating that hypothesis 2(a) is not supported. Further, the cooperation rate of the LSG is not significantly higher in FR4 than in FR2 (the same tests), indicating that hypothesis 2(b) is supported.

We now examine the cooperation rate of the 1st stage. Table 10 shows that the rate of the HSG is higher in FR4 than in FR2 (Mann-Whitney's U and t-tests, $p < 0.01$) and that the rate of the LSG is not significantly higher in FR4 than in FR2. Focusing on the 1st stage behavior, hypothesis 2(a) is not supported while 2(b) is supported.

Models (2) and (5) in Table 13 show the logit regression (equation (1)) results of the HSG. Model (2) indicates that hypothesis 2(a) is not supported (the odds ratio of "Four" is significant), while Model (5) indicates that hypothesis 2(a) is supported. Therefore, this hypothesis is supported if we consider the previous choice of the counterpart.

Models (3) and (6) in Table 13 show the logit regression (equation (1)) results of the LSG. Hypothesis 2(b) is supported because the odds ratio of "Four" is not significant.

<Table 13 here>

Using Tables 7 and 9, we examine hypotheses 3(a) and 3(b) by using simple statistical tests. The cooperation rate of the HSG is significantly higher in IFR2 than in FR2 (Mann-Whitney's U and t-tests, $p < 0.05$), indicating that hypothesis 3(a) is supported. The cooperation rate of the LSG is not significantly higher in IFR2 than in FR2 (both tests), indicating that hypothesis 3(b) is not supported.

Using tables 8 and 10, we check hypotheses 3(a) and 3(b). The 1st stage cooperation rate of the HSG is significantly higher in IFR2 than in FR2 (Mann-Whitney's U and t-tests, $p < 0.01$). This rate of the LSG is not significantly higher in IFR2 than in FR2. Focusing on the 1st stage behavior, hypothesis 3(a) is supported while 3(b) is not.

Models (2) and (5) in Table 14 indicate that the odds ratio of "Random" is greater than 1 and significant in both models. Thus, hypothesis 3(a) is supported.

For the cooperative behavior of the LSG, Models (3) and (6) in Table 14 indicate that the odds ratio of "Random" is not significant.

<Table 14 here>

Using Tables 7 and 9, we examine hypotheses 4(a) and 4(b) by using simple statistical tests. The cooperation rate of the HSG is significantly higher in IFR4 than in FR4 (Mann-Whitney's U and t-tests, $p < 0.01$), indicating hypothesis 4(a) is supported. The cooperation rate of the LSG is significantly higher in IFR4 than in IFR2, indicating hypothesis 4(b) is supported.

Using tables 8 and 10, we check hypotheses 4(a) and 4(b). The 1st stage cooperation rate of the HSG is significantly higher in IFR2 than in FR2 (Mann-Whitney's U and t-tests, $p < 0.01$). This rate of the LSG is not significantly higher in IFR2 than in FR2. Focusing on the 1st stage behavior, hypothesis 4(a) is supported while 4(b) is not.

<Table 15 here>

Models (2) and (5) in Table 15 indicate that hypothesis 4(a) is supported for the HSG (the odds ratio of "Random" is greater than 1 and significant), while Models (4) and (6) in the same table indicate that hypothesis 4(b) is not supported for the LSG.

Finally, we summarize the examination results of the hypotheses. When we analyze all the observations without distinguishing the level of cognitive ability, the differences in the cooperation rate supports the predictions of the repeated game theory.

We then divide participants into the HSG or LSG and investigate the hypotheses. For the HSG, all the hypotheses are supported when we consider the choice of the counterpart in the previous stage. Focusing on the first stage behavior, most of the hypotheses, except for 2(a), are supported.

Therefore, those in the HSG behave according to the expectations of the standard repeated game theory, namely "finite or infinite" condition and the expected number of stage games in a round affects the choice of the HSG.

Further, all the hypotheses but the hypothesis 2(b) are not supported for the LSG. This suggests that the behavior of those in the LSG is not affected by the repetition condition or the expected number of stage games in a round.

## 5 Additional Analyses

We conduct additional analyses to examine deeply why the cooperation rate of the HSG is in line with the basic repeated game theory but the cooperation rate of the LSG is not. To investigate the reason partially, we investigate how the cooperation rate in each group evolves in all the treatments and the kind of strategies each group employs in IFR treatments.

For the strategy estimation in IFR treatments, the strategy set was all-D, all-C (always choosing Cooperation), Grim trigger, Tit-For-Tat (TFT), WSLS (win-stay-lose-shift), and trigger with two periods punishment (Table 17) according to the maximum likelihood estimation developed by Bó and Fréchette (2011) and Fudenberg et al. (2012). This estimation[7] which enables us to investigate why the cooperation rate of the HSG increases from IFR2 to IFR4 but that of the LSG remains unchanged between IFR2 and IFR4.

Our findings are as follows: as a round continues, the LSG tends to choose defection in FR2, IFR2, and IFR4,

[7] Table A.1 is the estimation result with all the observations and indicates that 76%, 10.5%,0.9%, and 11% of choices were all-D, all-C, grim trigger, and TFT in IFR2, respectively. While 51%, 21%, 15%, and 7% of choices were all-D, grim trigger, TFT, and all-C in IFR4, respectively. The percentages of grim trigger and TFT increased in IFR4. Therefore, as $\delta$ increases, the percentage of the cooperative strategies increases.

while the HSG tends to choose defection in FR2 and IFR2. The strategy estimation reveals that the major strategy the HSG employs is the defective strategy in both IFR treatments, while the share of cooperative strategies such as grim trigger and TFT is much larger in IFR4 than in IFR2. This confirms that the HSG changes their strategy between IFR2 and IFR4. The major strategy the LSG employs is the defective strategy in both IFR treatments. The share of grim trigger strategy is larger in IFR4 than in IFR2; this partially suggests the behavioral change of the LSG. However, the combination of the defective strategy and grim trigger strategy yields mutual defection, implying that the cooperation rate remains unchanged.

## 5.1 High score group

First, we consider the evolution of the cooperation rate of the HSG in FR treatments (Table 16). In FR2, as the round progresses, the cooperation rate is lower (Mann-Whitney's U test; $p < 0.05$). In FR4, as the round goes on, there is little change in the cooperation rate.

Second, we consider the evolution of the cooperation rate in IFR treatments (Table 16). Considering defective strategy is a subgame perfect equilibrium strategy and risk dominant in IFR2, the HSG tends to choose defection as rounds continue, although the difference is non-significant (Mann-Whitney's U test, n.s.). In IFR4, where a cooperative strategy is a subgame perfect equilibrium strategy and risk dominant, the cooperation rate increases slightly (Mann-Whitney's U test, n.s.). This result suggests that the HSG chooses to cooperate from the first round. This may be due to the fact that the grim trigger strategy is a subgame perfect equilibrium strategy and risk dominant in IFR4.

<Table 16 here>

<Table 17 here>

Third, we investigate what kind of strategies the HSG employed, as shown in Table 17. The most popular strategy is all-D in IFR2 (74.7%) and IFR4 (43.2%). However, the share of all-D is clearly larger in IFR2 than in IFR4. Further, the shares of grim trigger and Tit-for-Tat is higher in IFR4 (grim trigger, 13.7%, TFT, 32.0%) than in IFR2 (grim trigger, 0.3%, TFT, 17.9%). The difference in the strategy distribution between IFR2 and IFR4 indicates that the cooperation rate of HSG is higher in IFR4 than in IFR2.

### 5.2 Low score group

First, we investigate how the cooperation rate of the LSG evolves. In FR2, as a round progresses, the cooperation rate is mildly lower (Table 16, Mann-Whitney's U test; LSG, $p < 0.10$). In FR4, as a round goes on, there is little change in the cooperation rate.

We consider the evolution of the cooperation rate in IFR treatments (Table 16). In IFR2, the cooperation rate is high in the first round. However, it decreases as a round continues (Mann-Whitney's U test, $p < 0.05$). This may be because the defective strategy is the subgame perfect equilibrium and risk dominant.

In IFR4, the cooperation rate decreases as a round goes on (Kruskal-Wallis test, $p < 0.01$). In this treatment, a cooperative strategy is a subgame perfect equilibrium strategy and risk dominant. However, the LSG does not tend to cooperate.

To investigate the behavior of the LSG in IFR treatments more in depth, we estimate what strategy they employed. The estimation method is the same as introduced for the HSG (section 5.1).

Table 17 indicates the strategy distribution of the LSG. The share of all-D is different between the two treatments, but this strategy is most popular in both treatments (75.7% in IFR2, 60.7% in IFR4). The share of grim trigger strategy is much larger in IFR4 (19.3%) than in IFR2 (2.9%). Despite the difference in strategy distribution between IFR2 and IFR4, more than half the strategy employed by the LSG is all-D in both treatments. This is because the cooperation rate of the LSG is almost the same between IFR4 and IFR2.

## 6 Discussion

We examined whether the cognitive ability of the non-students affects their cooperative behavior in repeated PD game experiments. When we do not consider the effect of cognitive ability, the regression results indicate that all the predictions of the repeated game theory were supported.

For the HSG, all the hypotheses were supported. Further, for the LSG, most of the hypotheses were not supported; the cooperation rate remains unchanged among treatments. The result of the HSG is close to the prediction of the repeated game theory, while that of the LSG is far from the prediction.

To examine the choice behavior of each group more in depth, we first examined how the cooperation rate evolves in each treatment. In most of the treatments, the cooperation rate of the LSG decreased as a round continued. That of the HSG decreased in FR2 and IFR2 but remained unchanged in FR4 and IFR4.

Second, we investigated the strategy estimation in IFR treatments. The strategy distribution of the HSG was largely different between IFR2 and IFR4. In both treatments, the most popular strategy was all-D but the share largely decreased from IFR2 to IFR4. The share of this strategy is less than 50% in IFR4. At the same time, the share of grim trigger and TFT increased from IFR2 to IFR4. The above-mentioned change yielded the increase in the cooperation rate from IFR2 to IFR4.

For the LSG, the all-D was most popular in both IFR treatments and the change in the share was small from IFR2 to IFR4. In both treatments, the share is larger than 50%. Further, the share of grim trigger increased from IFR2 to IFR4. This explains why the increase in cooperative strategy of the LSG in IFR4 did not lead to promote cooperation. Assume that a participant "X" has the low cognitive ability, if "X" plays with the participant in the same group, probably both keep on choosing defection. If "X" plays with the participant in the HSG, he or she probably chooses defection in the 1st stage. The counterpart reacts to "X"'s first choice and keeps on choosing defection from the second stage. Therefore, the strategy distribution of the LSG in IFR4 will still yield a low cooperation rate.

On the relationship between prior studies and our study, the result of the HSG in FR2 and in FR4 seems to be similar to the result of Bó (2005). Although Bó (2005) does not investigate closely, from the result in his finitely repeated treatment, the cooperation rate seemed to increase from two- to four-stage games.

Proto et al. (2019) studied university students, separated them into HSG or LSG according to the level of cognitive ability, and investigated the behavioral differences between the HSG and LSG in the wider classes of infinitely repeated games. In contrast, we recruited non-student adults and investigated the behavioral differences between HSGs and LSGs in finitely and infinitely repeated PD games. We conducted a series of experiments without separating them by the level of cognitive ability. Although the experimental method is different between Proto et al. (2019) and ours, our findings add new insights into the investigation between cognitive ability and behavioral differences in repeated games.

Our findings on the individual behavior of the LSG in finitely repeated PD treatments is in line with Al-Ubaydli et al. (2016) and Barreda-Tarrazona et al. (2017). However, our finding on the cooperation rate of the HSG in FR treatments is different. This might be due to the length of the stage games in a round. Both prior studies conducted ten-stage games in a round in their experiments, while we conducted two- and four-stage game experiments. Ten-stage games might be enough for participants to learn a certain choice.

Although the game structure is different, the behavioral change of the HSG is similar to the results of Yamagishi et al. (2016) and Inaba et al. (2018). Yamagishi et al. (2016) experimentally found that a sender with a thick prefrontal cortex and those with high cognitive abilities act selfishly in DG and reciprocally in UG. Inaba et al. (2018) found that high CRT participants give significantly smaller amounts of money in DG than in UG. In addition, the behavioral change of the HSG is in line with the result of Hanaki et al. (2016). They compare the situations characterized by strategic uncertainty; those with high cognitive ability change their behavior depending on the opponent type (human or robot), whereas those with low cognitive ability do not.

Additionally, Yamagishi et al. (2016) and Inaba et al. (2018) find that participants with low cognitive abilities do not change their behaviors. This is different from our result that those with low cognitive ability partially change cooperative behavior.

## 7 Conclusions

We examined how cognitive ability affects behavior in repeated PD experiments. The experimental results indicate that participants with high cognitive ability change their behavior as standard game theory expects. The strategy estimation results indicate that they change their strategy from IFR2 to IFR4.

On the contrary, participants with low cognitive ability seem not to change their behavior according to the conditions; the cooperation rate remains unchanged among treatments. However, the strategy estimation in IFR treatments shows that they change their cooperative behavior from IFR2 to IFR4, but this change does not increase cooperation. All-D is the most popular strategy and the share of all-D is more than 50% in both IFR treatments while the share of grim trigger increases from IFR2 to IFR4 but the share is at most 20%.

We conclude our paper with mentioning potential future studies. We did not divide the participants by their cognitive ability level in advance as Proto et al. (2019) did. A future study is to consider whether the same result in Proto et al. (2019) can be obtained when we employ their experimental procedure. A second proposal for a future study is to find out the effect of risk dominance on the cooperation rate of the HSG more clearly by conducting IFR$x$ $(x > 4)$ treatments. The difference in the expected number of stage games in a round brings about the difference in the risk dominant strategy. Our study does not fully capture the reason why their cooperation rate increases from IFR2 to IFR4. To find out the reason for this increase, conducting IFR$x$ will be required. This also helps us to examine whether the LSG could cooperate more when the expected number of stage games increases. These further studies would enable us to examine the importance of cognitive ability on decision-making in repeated PD experiments.

## References


**Al-Ubaydli, Omar, Garett Jones, and Jaap Weel**, "Average player traits as predictors of cooperation in a repeated prisoner's dilemma," *Journal of Behavioral and Experimental Economics*, oct 2016, *64*, 50–60.

**Andreoni, James and John H. Miller**, "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence," *The Economic Journal*, may 1993, *103* (418), 570.

**Barreda-Tarrazona, Iván, Ainhoa Jaramillo-Gutiérrez, Marina Pavan, and Gerardo Sabater-Grande**, "Individual Characteristics vs. Experience: An Experimental Study on Cooperation in Prisoner's Dilemma," *Frontiers in Psychology*, apr 2017, *8*.

**Ben-Ner, Avner, Fanmin Kong, and Louis Putterman**, "Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving," *Journal of Economic Psychology*, oct 2004, *25* (5), 581–589.

**Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro**, "WHO IS 'BEHAVIORAL'? COGNITIVE ABILITY AND ANOMALOUS PREFERENCES," *Journal of the European Economic Association*, nov 2013, *11* (6), 1231–1255.

**Blonski, Matthias and Giancarlo Spagnolo**, "Prisoners' other Dilemma," *International Journal of Game Theory*, Feb 2015, *44* (1), 61–81.

**_, Peter Ockenfels, and Giancarlo Spagnolo**, "Equilibrium Selection in the Repeated Prisoner's Dilemma: Axiomatic Approach and Experimental Evidence," *American Economic Journal: Microeconomics*, aug 2011, *3* (3), 164–192.

**Bó, Pedro Dal**, "Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games," *American Economic Review*, nov 2005, *95* (5), 1591–1604.

**_ and Guillaume R Fréchette**, "The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence," *American Economic Review*, feb 2011, *101* (1), 411–429.

**_ and Guillaume R. Fréchette**, "On the Determinants of Cooperation in Infinitely Repeated Games: A Survey," *Journal of Economic Literature*, mar 2018, *56* (1), 60–114.

**Brañas-Garza, Pablo, Teresa García-Muñoz, and Roberto Hernán González**, "Cognitive effort in the Beauty Contest Game," *Journal of Economic Behavior & Organization*, jul 2012, *83* (2), 254–260.

**Brandstätter, H. and W. Güth**, "Individuality in dictator and ultimatum games.," *Central European Journal of Operation Research*, 2002, *3* (10), 191–215.

**Carpenter, Jeffrey, Michael Graham, and Jesse Wolf**, "Cognitive ability and strategic sophistication," *Games and Economic Behavior*, jul 2013, *80*, 115–130.

**Chen, Chia-Ching, I-Ming Chiu, John Smith, and Tetsuji Yamada**, "Too smart to be selfish? Measures of cognitive ability, social preferences, and consistency," *Journal of Economic Behavior & Organization*, jun 2013, *90*, 112–122.

**Duffy, John and Jack Ochs**, "Cooperative behavior and the frequency of social interaction," *Games and Economic Behavior*, jul 2009, *66* (2), 785–812.

**Feinberg, Robert M. and Thomas A. Husted**, "An Experimental Test of Discount-Rate Effects on Collusive Behaviour in Duopoly Markets," *The Journal of Industrial Economics*, jun 1993, *41* (2), 153.

**Fischbacher, U.**, "Z-Tree: Zurich toolbox for ready-made economic experiments.," *Experimental Economics*, 2007, *10* (2), 171–178.

**Frederick, Shane**, "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, nov 2005, *19* (4), 25–42.

**Fudenberg, Drew, David G Rand, and Anna Dreber**, "Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World," *American Economic Review*, apr 2012, *102* (2), 720–749.

**Gill, David and Victoria Prowse**, "Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A Level-k Analysis," *Journal of Political Economy*, dec 2016, *124* (6), 1619–1676.

**Hanaki, Nobuyuki, Nicolas Jacquemet, Stéphane Luchini, and Adam Zylbersztejn.**, "Cognitive ability and the effect of strategic uncertainty.," *Theory and Decision*, 2016, *81* (1), 101^^e2^^80^^93121.

**Holt, Charles A.**, "An Experimental Test of the Consistent-Conjectures Hypothesis," *The American Economic Review*, 1985, *75* (3), 314–325.

**Inaba, Misato, Yumi Inoue, Satoshi Akutsu, Nobuyuki Takahashi, and Toshio Yamagishi**, "Preference and strategy in proposer's prosocial giving in the ultimatum game," *PLOS ONE*, mar 2018, *13* (3), e0193877.

**Jones, Garett**, "Are smarter groups more cooperative? Evidence from prisoner's dilemma experiments, 1959–2003," *Journal of Economic Behavior & Organization*, dec 2008, *68* (3-4), 489–497.

**Kawamura, Tetsuya and Kazuhito Ogawa**, "Cognitive Ability and Human Behavior in Experimental Ultimatum Games," *Research in Economics*, M 2019, *73* (1), 97–106.

**Kusakawa, Takao, Kazuhito Ogawa, and Tatsuhiro Shichijo**, "An experimental investigation of a third-person enforcement in a prisoner's dilemma game," *Economics Letters*, dec 2012, *117* (3), 704–707.

**Mullainathan, S. and E. Shafir**, *Scarcity: Why Having Too Little Means So Much*, Macmillan, New York, 2013.

**Ogawa, Kazuhito, Tetsuya Kawamura, and Keiichiro Matsushita**, "Effects of cognitive ability and age on giving in dictator game experiments," *Research in Economics*, 2020, *74* (4), 323–335.

**Oshio, Atsushi, Shingo Abe, and Pino Cutrone**, "Development, Reliability, and Validity of the Japanese Version of Ten Item Personality Inventory (TIPI-J)," *The Japanese Journal of Personality*, 2012, *21* (1), 40–52.

**Proto, Eugenio, Aldo Rustichini, and Andis Sofianos**, "Higher Intelligence Groups Have Higher Cooperation Rates in the Repeated Prisoner ' s Dilemma," CAGE Working Paper No: 255 October 2015.

**_ , _ , and _** , "Intelligence, Personality, and Gains from Cooperation in Repeated Interactions," *Journal of Political Economy*, 2019, *127* (3), 1351–1390.

**Raven, J.C.**, "Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive," Master's thesis, University of London 1936.

**Roth, Alvin E and J.Keith Murnighan**, "Equilibrium behavior and repeated play of the prisoner's dilemma," *Journal of Mathematical Psychology*, apr 1978, *17* (2), 189–198.

**Selten, Reinhard and Rolf Stoecker**, "End behavior in sequences of finite Prisoner's Dilemma supergames A learning theory approach," *Journal of Economic Behavior & Organization*, mar 1986, *7* (1), 47–70.

**Yamagishi, Toshio, Haruto Takagishi, Alan de Souza Rodrigues Fermin, Ryota Kanai, Yang Li, and Yoshie Matsumoto**, "Cortical thickness of the dorsolateral prefrontal cortex predicts strategic choices in economic games," *Proceedings of the National Academy of Sciences*, may 2016, *113* (20), 5582–5587.

**Yang, Junho, Tetsuya Kawamura, and Kazuhito Ogawa**, "Experimental Multimarket Contact Inhibits Co-operation," *Metroeconomica*, jul 2015, *67* (1), 21–43.

Table 1: Payoff matrix used in the experiment

| | A | B |
|---|---|---|
| A | 75 pt | 10 pt |
| B | 100 pt | 45 pt |

Table 2: Reduced payoff matrix

| | Trigger | All-D |
|---|---|---|
| Trigger | $\frac{75}{1-\delta}$ | $10+\frac{45\delta}{1-\delta}$ |
| All-D | $100+\frac{45\delta}{1-\delta}$ | $\frac{45}{1-\delta}$ |

Table 3: Session profile

| Date | Treatment | Exp. no. of stage | No. of participants |
|---|---|---|---|
| 160609_1245 | FR | 4 | 24 |
| 160617_1310 | FR | 4 | 28 |
| 160702_1336 | FR | 4 | 12 |
| 160908_1243 | IFR | 4 | 16 |
| 160913_1204 | IFR | 4 | 28 |
| 160930_0939 | FR | 2 | 22 |
| 160930_1338 | FR | 2 | 16 |
| 161015_1019 | IFR | 2 | 26 |
| 161015_1350 | FR | 2 | 16 |
| 161101_1002 | IFR | 2 | 22 |
| 161101_1327 | IFR | 2 | 12 |
| 161217_0953 | IFR | 2 | 24 |
| 161217_1352 | IFR | 4 | 18 |
| 170317_0952 | FR | 4 | 20 |
| 170317_1418 | IFR | 4 | 18 |
| 170519_1004 | IFR | 4 | 14 |
| 170519_1418 | FR | 4 | 14 |
| 170520_0948 | FR | 2 | 20 |
| 170520_1323 | FR | 2 | 18 |

Table 4: Profiles of participants

| | FR2 ($n$ =92) | IFR2 ($n$ =84) | FR4 ($n$ =98) | IFR4 ($n$ =94) | All |
|---|---|---|---|---|---|
| Age | 55.2(16.2) | 55.0 (16.6) | 58.3 (15.1) | 55.9 (15.8) | 56.2 (15.9) |
| | min 24 | min 24 | min 18 | min 22 | min 18 |
| | max 82 | max 84 | max 82 | max 85 | max 85 |
| Average Raven's score | 8.0 (3.7) | 7.6 (3.1) | 7.71 (3.2) | 8.1 (3.3) | 7.9 (3.3) |
| Median Raven's score | 8 | 7 | 8 | 9 | 8 |
| Min Raven's score | 1 | 2 | 0 | 0 | 0 |
| Max Raven's score | 15 | 15 | 15 | 15 | 15 |
| The % of male | 51% | 41% | 46% | 49% | 47% |

The numbers between parentheses are standard deviations.

Table 5: Descriptive statistics of the cooperation rate of each treatment The numbers between parentheses represent the average number of realized stages.

| | | Average cooperation rate | |
|---|---|---|---|
| Treatment | No. obs. | mean | Standard dev. |
| FR2 | 1808 | 0.207 | 0.405 |
| FR4 | 1960 | 0.239 | 0.427 |
| Total | 3768 | 0.224 | 0.417 |
| IFR2 ( 1.83 ) | 1746 | 0.254 | 0.435 |
| IFR4 ( 3.42 ) | 2138 | 0.338 | 0.473 |
| Total | 3884 | 0.300 | 0.458 |

Table 6: Descriptive statistics of the average cooperation rate in the first stage

| | | Average cooperation rate in the first stage | |
|---|---|---|---|
| Treatment | No. obs. | mean | Standard dev. |
| FR2 | 904 | 0.218 | 0.413 |
| FR4 | 490 | 0.292 | 0.455 |
| Total | 1394 | 0.244 | 0.430 |
| IFR2 | 840 | 0.257 | 0.437 |
| IFR4 | 470 | 0.421 | 0.494 |
| Total | 1310 | 0.316 | 0.465 |

Table 7: Mean cooperation rate and cognitive ability in IFR treatments

| | Raven $< 9$ | | | Raven $\geq 9$ | | |
|---|---|---|---|---|---|---|
| | Coop. rate | Std. dev. | Obs. | Coop. rate | Std. dev. | Obs. |
| Total | 0.28 | 0.01 | 2086 | 0.324 | 0.01 | 1798 |
| IFR2 | 0.269 | 0.444 | 1078 | 0.229 | 0.421 | 668 |
| IFR4 | 0.292 | 0.455 | 1008 | 0.380 | 0.486 | 1130 |

Table 8: Mean cooperation rate in the first stage and cognitive ability in IFR treatments

| | Raven $< 9$ | | | Raven $\geq 9$ | | |
|---|---|---|---|---|---|---|
| | Coop. rate | Std. dev. | Obs. | Coop. rate | Std. dev. | Obs. |
| Total | 0.292 | 0.455 | 750 | 0.348 | 0.477 | 560 |
| IFR2 | 0.267 | 0.443 | 520 | 0.241 | 0.428 | 320 |
| IFR4 | 0.348 | 0.477 | 230 | 0.492 | 0.501 | 240 |

Table 9: Mean cooperation rate and cognitive ability in FR treatments

| | Raven $< 9$ | | | Raven $\geq 9$ | | |
|---|---|---|---|---|---|---|
| | Coop. rate | Std. dev. | Obs. | Coop. rate | Std. dev. | Obs. |
| Total | 0.251 | 0.01 | 2062 | 0.191 | 0.01 | 1706 |
| FR2 | 0.253 | 0.435 | 922 | 0.158 | 0.365 | 886 |
| FR4 | 0.249 | 0.433 | 1140 | 0.226 | 0.442 | 820 |

Table 10: Mean cooperation rate in the first stage and cognitive ability in FR treatments

| | Raven $< 9$ | | | Raven $\geq 9$ | | |
|---|---|---|---|---|---|---|
| | Coop. rate | Std. dev. | Obs. | Coop. rate | Std. dev. | Obs. |
| Total | 0.276 | 0.447 | 746 | 0.207 | 0.405 | 648 |
| FR2 | 0.260 | 0.439 | 461 | 0.174 | 0.379 | 443 |
| FR4 | 0.302 | 0.460 | 285 | 0.278 | 0.449 | 205 |

Table 11: Variable list

| Variable | Description |
|---|---|
| Choice | Dependent variable, 1 if a participant cooperates, 0 otherwise in a stage. |
| POC | The choice of counterpart in the previous stage. 1 if he or she cooperated, 0 otherwise. |
| Random | 1 if IFR treament, 0 otherwise. |
| Fix | 1 if FR treament, 0 otherwise. |
| Four | 1 if the (expected) number of stages is 4, 0 otherwise. |
| Two | 1 if the (expected) number of stages is 2, 0 otherwise. |
| stage | stage number |
| round | round number |
| age | The age of a participant. |
| gender | 1 if male, 0 otherwise. |
| Raven | The score of Raven's test, min. 0 and max. 16. |
| High Raven | 1 if a participant has the Raven score higher than 8, 0 otherwise. |
| Low Raven | 1 if a participant has the Raven score lower than 9, 0 otherwise. |
| Extroversion | Personality trait. The average score of the answers of two questions on Extroversion in Oshio et al. (2012). |
| Neuroticism | Personality trait. The average score of the answers of two questions on Neuroticism in Oshio et al. (2012). |
| Openness | Personality trait. The average score of the answers of two questions on Openness in Oshio et al. (2012). |
| Agreeableness | Personality trait. The average score of the answers of two questions on Agreeableness in Oshio et al. (2012). |
| Conscientiousness | Personality trait. The average score of the answers of two questions on Conscientiousness in Oshio et al. (2012). |

Table 12: Hypothesis 1 IFR2 vs. IFR4.

| | (1) All data | (2) Raven $\geq 9$ | (3) Raven $< 9$ | (4) All data | (5) Raven $\geq 9$ | (6) Raven $< 9$ |
|---|---|---|---|---|---|---|
| Four | 1.484*** | 2.953*** | 0.812 | 1.258 | 2.787*** | 0.686** |
| | (0.204) | (0.741) | (0.124) | (0.191) | (0.977) | (0.126) |
| round | 0.945** | 1.027 | 0.894*** | 0.985 | 1.128** | 0.911*** |
| | (0.023) | (0.043) | (0.023) | (0.027) | (0.059) | (0.031) |
| stage | 0.923* | 0.889** | 0.971 | 0.983 | 0.965 | 1.022 |
| | (0.043) | (0.051) | (0.042) | (0.043) | (0.047) | (0.052) |
| age | 0.997 | 0.996 | 1.014 | 0.997 | 0.992 | 1.007 |
| | (0.005) | (0.011) | (0.009) | (0.006) | (0.017) | (0.012) |
| gender | 1.180 | 1.018 | 6.371** | 0.951 | 0.735 | 5.517* |
| | (0.495) | (0.666) | (5.227) | (0.495) | (0.720) | (5.692) |
| gender×age | 1.003 | 1.003 | 0.979* | 1.006 | 1.014 | 0.981 |
| | (0.007) | (0.014) | (0.012) | (0.009) | (0.021) | (0.015) |
| Extroversion | 1.057 | 1.107 | 0.997 | 1.000 | 1.248** | 0.859 |
| | (0.068) | (0.102) | (0.091) | (0.072) | (0.124) | (0.089) |
| Neuroticism | 1.067 | 1.172** | 0.985 | 1.074 | 1.152 | 0.979 |
| | (0.057) | (0.093) | (0.062) | (0.062) | (0.102) | (0.072) |
| Openness | 0.944 | 0.995 | 0.902 | 1.007 | 1.100 | 0.995 |
| | (0.068) | (0.099) | (0.093) | (0.085) | (0.135) | (0.125) |
| Agreeableness | 0.947 | 0.792** | 1.157 | 0.912 | 0.666*** | 1.259* |
| | (0.067) | (0.080) | (0.105) | (0.080) | (0.080) | (0.150) |
| Conscientiousness | 1.047 | 1.124 | 0.960 | 1.053 | 1.124 | 0.897 |
| | (0.081) | (0.140) | (0.114) | (0.093) | (0.159) | (0.129) |
| POC | | | | 4.283*** | 13.888*** | 1.606** |
| | | | | (0.754) | (3.321) | (0.298) |
| Observations | 3884 | 1798 | 2086 | 2574 | 1238 | 1336 |
| Pseudo $R^2$ | 0.019 | 0.049 | 0.025 | 0.090 | 0.268 | 0.032 |

The odds ratio is reported.

Robust standard errors clustered by pair in each round are presented in parentheses.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 13: Hypothesis 2 FR2 vs. FR4

| | (1) All data | (2) Raven $\geq 9$ | (3) Raven $< 9$ | (4) All data | (5) Raven $\geq 9$ | (6) Raven $< 9$ |
|---|---|---|---|---|---|---|
| Four | 1.154 | 1.588$^{**}$ | 0.924 | 1.015 | 1.447 | 0.800 |
| | (0.153) | (0.342) | (0.154) | (0.168) | (0.395) | (0.168) |
| round | 0.927$^{***}$ | 0.923$^{**}$ | 0.927$^{**}$ | 0.954$^{*}$ | 0.946 | 0.957 |
| | (0.022) | (0.034) | (0.027) | (0.027) | (0.045) | (0.032) |
| stage | 0.863$^{***}$ | 0.842$^{**}$ | 0.875$^{***}$ | 0.967 | 0.924 | 0.986 |
| | (0.035) | (0.056) | (0.045) | (0.066) | (0.102) | (0.092) |
| age | 1.009 | 0.982$^{**}$ | 1.031$^{***}$ | 1.001 | 0.972$^{***}$ | 1.021$^{**}$ |
| | (0.006) | (0.008) | (0.008) | (0.007) | (0.010) | (0.009) |
| gender | 0.064$^{***}$ | 0.066$^{***}$ | 0.238$^{*}$ | 0.044$^{***}$ | 0.038$^{***}$ | 0.175$^{*}$ |
| | (0.037) | (0.057) | (0.205) | (0.030) | (0.040) | (0.167) |
| gender×age | 1.039$^{***}$ | 1.042$^{***}$ | 1.018 | 1.047$^{***}$ | 1.052$^{***}$ | 1.025$^{*}$ |
| | (0.009) | (0.016) | (0.013) | (0.011) | (0.020) | (0.014) |
| Extroversion | 0.848$^{***}$ | 0.786$^{***}$ | 0.915 | 0.832$^{***}$ | 0.809$^{**}$ | 0.875$^{*}$ |
| | (0.046) | (0.069) | (0.059) | (0.052) | (0.083) | (0.066) |
| Neuroticism | 0.944 | 0.899 | 0.979 | 0.954 | 0.924 | 0.963 |
| | (0.046) | (0.071) | (0.062) | (0.055) | (0.095) | (0.072) |
| Openness | 1.058 | 1.039 | 1.028 | 0.962 | 0.898 | 0.963 |
| | (0.056) | (0.091) | (0.076) | (0.059) | (0.108) | (0.082) |
| Agreeableness | 1.171$^{**}$ | 1.317$^{***}$ | 1.009 | 1.233$^{***}$ | 1.271$^{**}$ | 1.164 |
| | (0.072) | (0.121) | (0.089) | (0.091) | (0.140) | (0.127) |
| Conscientiousness | 0.871$^{*}$ | 0.922 | 0.929 | 0.918 | 1.103 | 0.882 |
| | (0.062) | (0.094) | (0.101) | (0.074) | (0.131) | (0.114) |
| POC | | | | 2.963$^{***}$ | 3.859$^{***}$ | 2.574$^{***}$ |
| | | | | (0.448) | (0.788) | (0.525) |
| Observations | 3748 | 1686 | 2062 | 2364 | 1048 | 1316 |
| Pseudo $R^2$ | 0.044 | 0.055 | 0.054 | 0.074 | 0.110 | 0.071 |

The odds ratio is reported.

Robust standard errors clustered by pair in each round are presented in parentheses.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 14: Hypothesis 3 IFR4 vs. FR4

| | (1) All data | (2) Raven $\geq 9$ | (3) Raven $< 9$ | (4) All data | (5) Raven $\geq 9$ | (6) Raven $< 9$ |
|---|---|---|---|---|---|---|
| Random | 1.781*** | 2.492*** | 1.038 | 1.536*** | 2.084*** | 0.966 |
| | (0.234) | (0.486) | (0.155) | (0.186) | (0.436) | (0.157) |
| round | 0.969 | 1.117 | 0.853*** | 0.970 | 1.109 | 0.861*** |
| | (0.043) | (0.077) | (0.042) | (0.041) | (0.080) | (0.046) |
| stage | 0.912** | 0.889** | 0.927* | 0.984 | 0.946 | 1.008 |
| | (0.042) | (0.049) | (0.043) | (0.042) | (0.046) | (0.050) |
| age | 0.985*** | 0.975*** | 1.002 | 0.983*** | 0.963*** | 1.003 |
| | (0.006) | (0.008) | (0.010) | (0.007) | (0.011) | (0.013) |
| gender | 0.294*** | 0.280** | 0.366 | 0.153*** | 0.077*** | 0.367 |
| | (0.134) | (0.177) | (0.300) | (0.079) | (0.063) | (0.357) |
| gender×age | 1.022*** | 1.023* | 1.016 | 1.033*** | 1.048*** | 1.017 |
| | (0.008) | (0.013) | (0.013) | (0.009) | (0.017) | (0.015) |
| Extroversion | 0.897* | 0.941 | 0.789*** | 0.862** | 0.935 | 0.771*** |
| | (0.054) | (0.090) | (0.060) | (0.055) | (0.094) | (0.066) |
| Neuroticism | 1.075 | 1.155 | 0.967 | 1.100 | 1.109 | 0.994 |
| | (0.067) | (0.102) | (0.077) | (0.067) | (0.097) | (0.083) |
| Openness | 0.956 | 0.827* | 1.178* | 0.945 | 0.863 | 1.135 |
| | (0.056) | (0.086) | (0.106) | (0.063) | (0.105) | (0.114) |
| Agreeableness | 1.107 | 0.963 | 1.610*** | 1.099 | 0.944 | 1.639*** |
| | (0.092) | (0.094) | (0.180) | (0.099) | (0.095) | (0.207) |
| Conscientiousness | 0.956 | 1.311** | 0.564*** | 0.998 | 1.341** | 0.580*** |
| | (0.077) | (0.156) | (0.080) | (0.081) | (0.155) | (0.083) |
| POC | | | | 4.730*** | 8.798*** | 2.458*** |
| | | | | (0.773) | (1.908) | (0.487) |
| Observations | 4098 | 1950 | 2148 | 3138 | 1505 | 1633 |
| Pseudo $R^2$ | 0.025 | 0.061 | 0.054 | 0.110 | 0.222 | 0.080 |

The odds ratio is reported.

Robust standard errors clustered by pair in each round are presented in parentheses.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 15: Hypothesis 4 IFR2 vs. FR2

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | All data | Raven $\geq 9$ | Raven $< 9$ | All data | Raven $\geq 9$ | Raven $< 9$ |
| Random | 1.377*** | 1.764*** | 1.095 | 1.384** | 1.551* | 1.124 |
| | (0.147) | (0.303) | (0.142) | (0.200) | (0.398) | (0.203) |
| round | 0.927*** | 0.931** | 0.921*** | 0.964* | 0.993 | 0.954* |
| | (0.016) | (0.027) | (0.020) | (0.021) | (0.038) | (0.025) |
| stage | 0.970 | 0.915 | 1.006 | 1.045 | 1.107 | 1.040 |
| | (0.052) | (0.080) | (0.068) | (0.102) | (0.171) | (0.128) |
| age | 1.023*** | 1.002 | 1.046*** | 1.024*** | 1.006 | 1.033*** |
| | (0.005) | (0.010) | (0.008) | (0.006) | (0.014) | (0.010) |
| gender | 1.022 | 0.488 | 4.345* | 1.555 | 0.828 | 3.616 |
| | (0.480) | (0.314) | (3.655) | (0.914) | (0.785) | (3.591) |
| gender×age | 1.000 | 1.014 | 0.977* | 0.994 | 1.006 | 0.981 |
| | (0.008) | (0.013) | (0.012) | (0.010) | (0.020) | (0.015) |
| Extroversion | 1.023 | 0.876 | 1.100 | 0.974 | 0.966 | 0.982 |
| | (0.052) | (0.075) | (0.073) | (0.068) | (0.120) | (0.081) |
| Neuroticism | 0.895** | 0.964 | 0.817*** | 0.888** | 0.979 | 0.799*** |
| | (0.039) | (0.070) | (0.044) | (0.052) | (0.096) | (0.059) |
| Openness | 1.260*** | 1.374*** | 1.171* | 1.286*** | 1.361*** | 1.171 |
| | (0.071) | (0.112) | (0.099) | (0.095) | (0.150) | (0.124) |
| Agreeableness | 0.872** | 1.086 | 0.773*** | 0.873* | 0.969 | 0.822** |
| | (0.049) | (0.118) | (0.052) | (0.067) | (0.153) | (0.073) |
| Conscientiousness | 0.993 | 0.749*** | 1.193** | 1.006 | 0.796 | 1.224* |
| | (0.060) | (0.075) | (0.107) | (0.081) | (0.111) | (0.145) |
| POC | | | | 2.183*** | 4.242*** | 1.502** |
| | | | | (0.318) | (0.929) | (0.278) |
| Observations | 3534 | 1534 | 2000 | 1800 | 781 | 1019 |
| Pseudo $R^2$ | 0.041 | 0.035 | 0.065 | 0.057 | 0.101 | 0.050 |

The odds ratio is reported.

Robust standard errors clustered by pair in each round are presented in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 16: The cooperation rate: round$= 1$ & round$\geq 2$

| | Round $= 1$ | | | Round$\geq 2$ | | |
|---|---|---|---|---|---|---|
| | Coop. rate | Std. Err. | Obs. | Coop. rate | Std. Err. | Obs. |
| FR2 (All) | 0.283 | 0.033 | 184 | 0.198 | 0.010 | 1,624 |
| FR2 (Raven$<$ 9) | 0.330 | 0.049 | 94 | 0.245 | 0.015 | 828 |
| FR2 (Raven$\geq$ 9) | 0.233 | 0.045 | 90 | 0.150 | 0.012 | 796 |
| FR4 (All) | 0.260 | 0.022 | 392 | 0.234 | 0.11 | 1,568 |
| FR4 (Raven$<$ 9) | 0.272 | 0.030 | 228 | 0.243 | 0.014 | 912 |
| FR4 (Raven$\geq$ 9) | 0.244 | 0.034 | 164 | 0.221 | 0.016 | 656 |
| IFR2 (All) | 0.330 | 0.034 | 194 | 0.244 | 0.011 | 1,552 |
| IFR2 (Raven$<$ 9) | 0.368 | 0.045 | 117 | 0.257 | 0.014 | 961 |
| IFR2 (Raven$\geq$ 9) | 0.273 | 0.051 | 77 | 0.223 | 0.017 | 591 |
| IFR4 (All) | 0.374 | 0.023 | 452 | 0.329 | 0.011 | 1,686 |
| IFR4 (Raven$<$ 9) | 0.398 | 0.033 | 226 | 0.261 | 0.016 | 782 |
| IFR4 (Raven$\geq$ 9) | 0.350 | 0.032 | 226 | 0.387 | 0.016 | 904 |

Table 17: Individual strategies in IFR treatments for different Raven scores

| | | **Raven Score $\geq 9$** | | | **Raven Score $< 9$** | | |
|---|---|---|---|---|---|---|---|
| | | **Total** | $\delta = 0.5$ | $\delta = 0.75$ | **Total** | $\delta = 0.5$ | $\delta = 0.75$ |
| **ALL-D** | **Coef** | 0.583*** | 0.747*** | 0.432*** | 0.694*** | 0.757*** | 0.607*** |
| | **S.E.** | 0.021 | 0.029 | 0.034 | 0.026 | 0.035 | 0.039 |
| | **CI** | [0.549,0.621] | [0.695,0.793] | [0.38,0.489] | [0.652,0.739] | [0.705,0.825] | [0.531,0.663] |
| **ALL-C** | **Coef** | 0.068*** | 0.065*** | 0.072*** | 0.103*** | 0.136*** | 0.071*** |
| | **S.E.** | 0.017 | 0.019 | 0.028 | 0.019 | 0.027 | 0.022 |
| | **CI** | [0.042,0.099] | [0.036,0.097] | [0.027,0.118] | [0.073,0.138] | [0.089,0.184] | [0.041,0.109] |
| **TRIGGER** | **Coef** | 0.071*** | 0.003 | 0.137*** | 0.098*** | 0.029 | 0.193*** |
| | **S.E.** | 0.022 | 0.009 | 0.043 | 0.03 | 0.035 | 0.057 |
| | **CI** | [0.034,0.107] | [0.0,0.028] | [0.064,0.206] | [0.044,0.153] | [0.0,0.116] | [0.115,0.294] |
| **TFT** | **Coef** | 0.259*** | 0.179*** | 0.32*** | 0.061** | 0.048 | 0.074 |
| | **S.E.** | 0.031 | 0.031 | 0.045 | 0.03 | 0.041 | 0.056 |
| | **CI** | [0.206,0.312] | [0.128,0.237] | [0.256,0.408] | [0.018,0.112] | [0.0,0.133] | [0.0,0.176] |
| **WSLS** | **Coef** | 0.014 | 0.004 | 0.026 | 0.033* | 0.018 | 0.041 |
| | **S.E.** | 0.012 | 0.009 | 0.025 | 0.017 | 0.021 | 0.028 |
| | **CI** | [0.0,0.034] | [0.0,0.025] | [0.0,0.072] | [0.007,0.068] | [0.0,0.061] | [0.0,0.089] |
| **T2** | **Coef** | 0.005 | 0.002 | 0.013 | 0.011 | 0.012 | 0.014 |
| | **S.E.** | 0.009 | 0.009 | 0.022 | 0.014 | 0.019 | 0.018 |
| | **CI** | [0.0,0.026] | [0.0,0.029] | [0.0,0.058] | [0.0,0.042] | [0.0,0.052] | [0.0,0.046] |
| $\gamma$ | **Coef** | 0.474*** | 0.447*** | 0.456*** | 0.635*** | 0.646*** | 0.604*** |
| | **S.E.** | 0.015 | 0.027 | 0.022 | 0.022 | 0.032 | 0.033 |
| | **CI** | [0.45,0.497] | [0.405,0.495] | [0.42,0.49] | [0.599,0.676] | [0.594,0.699] | [0.552,0.658] |

# A Strategy estimation for all the participants in IFR treatments

Table A.1: Individual strategies in IFR treatments

| | | **Total** | $\delta = 0.5$ | $\delta = 0.75$ |
|---|---|---|---|---|
| **ALL-D** | **Coef** | 0.648*** | 0.76*** | 0.513*** |
| | **S.E.** | 0.016 | 0.023 | 0.025 |
| | **CI** | [0.622,0.675] | [0.718,0.802] | [0.476,0.56] |
| **ALL-C** | **Coef** | 0.087*** | 0.105*** | 0.072*** |
| | **S.E.** | 0.012 | 0.017 | 0.019 |
| | **CI** | [0.07,0.108] | [0.075,0.136] | [0.043,0.111] |
| **TRIGGER** | **Coef** | 0.077*** | 0.009 | 0.152*** |
| | **S.E.** | 0.017 | 0.013 | 0.036 |
| | **CI** | [0.05,0.11] | [0.0,0.034] | [0.08,0.215] |
| **TFT** | **Coef** | 0.155*** | 0.11*** | 0.211*** |
| | **S.E.** | 0.022 | 0.022 | 0.036 |
| | **CI** | [0.121,0.193] | [0.076,0.146] | [0.156,0.271] |
| **WSLS** | **Coef** | 0.023* | 0.009 | 0.037* |
| | **S.E.** | 0.012 | 0.013 | 0.019 |
| | **CI** | [0.0,0.041] | [0.0,0.031] | [0.001,0.066] |
| **T2** | **Coef** | 0.01 | 0.006 | 0.015 |
| | **S.E.** | 0.009 | 0.01 | 0.016 |
| | **CI** | [0.0,0.027] | [0.0,0.03] | [0.0,0.05] |
| $\gamma$ | **Coef** | 0.553*** | 0.561*** | 0.522*** |
| | **S.E.** | 0.013 | 0.02 | 0.016 |
| | **CI** | [0.532,0.579] | [0.53,0.594] | [0.496,0.549] |

Each "Coef" represents the probability estimated using the ML of the corresponding strategy.
S.E. is reported in brackets.
$\gamma$ is the error coefficient estimated for the choice function used in the ML.
We test equality to 0 using the Wald test. $^{*}$, $p < 0.1$, $^{**}$. $p < 0.05$, $^{***}$. $p < 0.01$.

## B Instruction IFR4 [8]

In this experiment, you will make economic decisions. Your cash holdings awarded at the end of the experiment depend not only on your own decision-making but also on the other participants' decisions. The content of this experiment is briefly explained in the "Flowchart of the Experiment." Please listen to the following explanations while referring to the chart.

### Overview of the experiment

Note the "Overall flow" in the upper half of the "Flowchart of the Experiment." The experiment proceeds from the left to the right and consists of five rounds. In each round, you are required to make a decision about an economic choice. In each round, your reward depends on your and the other participants' decisions. Below is the procedure of this experiment.

### Preparation for the experiment

Before the experiment begins, you are assigned a participant ID. This ID remains the same throughout the experiment. Then, you are assigned 1000 yen as the participation fee. Note the "Content of a round '' in the lower half of the flowchart. Because the content of a round is the same from the first to the fifth rounds, the explanation here applies to all rounds.

### First round (1): Preparation

At the start of the first round, groups of two participants are formed. You will not know who the member in your group is. A round consists of one "stage" or "several stages." The number of stages in a round is randomly determined. The first stage is called "stage 1," the second stage is called "stage 2", etc. We explain below what you should do in the first stage of the first round.

### First round (2): Procedure in the first stage

In the first stage, you are assigned the following role. At the beginning of the stage, you and your group member are given four alternatives, two under each A and B. You and your group member choose alternative between A or B within 30 seconds. After both of you finish choosing an alternative, only your actual decisions are revealed to your group member. Your total points change according to the decision results in the following way.

[8]This instruction is based on those of Kusakawa et al. (2012) and Yang et al. (2015) and was originally written in Japanese.

### First round (3): How to change the cash holdings in the first stage

The sheet "Change in point in each stage" indicates the change in the points of each player. We explain how to read the table when you choose A or B. You pick a line in this table based on your decision. If you choose A, you check the line in which A is written in the leftmost cell. That is, you check the line in which +75 yen and +10 yen are written. Based on the decision of your group member, you can understand which cell you read. If your group member chooses A, you read the left cell. Your reward in this stage is +75 yen. If your group member chooses B, you read the right cell. Your reward in this stage is +10 yen.

Assume you choose B. If your group member chooses A, you read the left cell. Your reward in this stage is +100 yen. If the other person chooses B, you read the right cell. Your reward in this stage is +45 yen.

### First round (4): At the end of the first stage

After all the groups have made their decisions, the experimenter randomly chooses whether there is another stage, namely the second stage.

To do so, we have four cards, one is the joker and the others are kings.

First, with probability $1/4$, the first round ends with the first stage. The points of each participant are carried over to the next round. With probability $3/4$, you proceed to the second stage in the first round. The points of each participant are carried over to the next stage. If the experimenter chooses the joker, the first round ends with the first stage. If he/she chooses the king, the second stage begins.

Therefore, a round proceeds until the experimenter chooses the joker. The action of choosing a card is projected onto a large screen so that all the participants can confirm the action.

*When we conducted FR treatments, we change "First round (4): At the end of the first stage" in the following way.*

*After all the participant made decision in the first stage,they proceed the second stage. Their decision making in the second stage is the same as they do in the first stage.In each round, they have four stages.*

### First round (5): How decision-makers are chosen in the second round

*After choosing the joker, the second round begins. The procedure in this round is the same as in the first round. The total points are the end of the first round are carried to the second round.*

### Procedure after the first round

*As previously mentioned, the procedure in the first to the fifth rounds is the same. However, there is an important point to be made regarding how groups of three are created. From the second round, three participants who have not been in the same group before become a group. Therefore, the participants who have already been in the same group with you will not be in a group with you again.*

### How to finalize the experiment

*When the fifth round ends, the experiment also ends. The cash holdings at that point are paid to you ( 1 point = 1 yen). The larger the amount of points, the more you are paid.*

### Other important points

*The decisions made during the experiment are done on an individually assigned PC. Read "How to read the computer displays" and make sure you understand it. You have 10 minutes from then to formulate your strategy. You can read this document in these 10 minutes.*

*That is all for today's experiment. Please do not talk to the other participants and please follow the experimenter's instructions.*

*If you have questions about the experiment, please raise your hand and ask your question loud enough for everyone to hear. The experimenter will answer your question aloud.*

# C Additional regression analyses 1

*Here, we redefine that a participant with the Raven's score is higher than nine belongs to high score group. The results are almost the same found in Tables 12 to 15.*

Table C.2: Hypothesis 1 IFR2 vs. IFR4.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Raven $\geq 10$ | Raven $< 10$ | Raven $\geq 10$ | Raven $< 10$ |
| Four | 2.325*** | 1.089 | 2.492** | 0.945 |
| | (0.613) | (0.166) | (1.022) | (0.169) |
| round | 1.006 | 0.920*** | 1.134* | 0.951 |
| | (0.046) | (0.023) | (0.074) | (0.029) |
| stage | 0.890* | 0.937* | 0.974 | 0.971 |
| | (0.057) | (0.036) | (0.055) | (0.045) |
| age | 1.027 | 1.006 | 1.032 | 0.997 |
| | (0.018) | (0.008) | (0.034) | (0.009) |
| gender | 2.642 | 2.482 | 2.231 | 1.470 |
| | (2.299) | (1.684) | (3.568) | (1.124) |
| gender×age | 0.978 | 0.992 | 0.985 | 1.000 |
| | (0.020) | (0.010) | (0.036) | (0.012) |
| Extroversion | 1.169 | 1.020 | 1.313** | 0.912 |
| | (0.120) | (0.080) | (0.151) | (0.085) |
| Neuroticism | 1.221** | 0.948 | 1.187* | 0.962 |
| | (0.096) | (0.059) | (0.116) | (0.070) |
| Openness | 0.998 | 1.008 | 1.163 | 1.098 |
| | (0.103) | (0.102) | (0.158) | (0.134) |
| Agreeableness | 0.768** | 1.040 | 0.665*** | 1.064 |
| | (0.090) | (0.090) | (0.099) | (0.120) |
| Conscientiousness | 1.115 | 0.958 | 1.050 | 0.919 |
| | (0.148) | (0.104) | (0.172) | (0.117) |
| POC | | | 20.271*** | 2.027*** |
| | | | (5.483) | (0.360) |
| Observations | 1328 | 2556 | 908 | 1666 |
| Pseudo $R^2$ | 0.056 | 0.019 | 0.332 | 0.031 |

The odds ratio is reported.

Robust standard errors clustered by pair in each round are presented in parentheses.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table C.3: Hypothesis 2 FR2 vs. FR4

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Raven$\geq 10$ | Raven$< 10$ | Raven$\geq 10$ | Raven$< 10$ |
| Four | 1.331 | 1.066 | 1.342 | 0.864 |
| | (0.354) | (0.159) | (0.462) | (0.161) |
| round | 0.921** | 0.929*** | 0.943 | 0.956 |
| | (0.038) | (0.026) | (0.052) | (0.031) |
| stage | 0.840** | 0.869*** | 0.905 | 0.982 |
| | (0.064) | (0.041) | (0.119) | (0.078) |
| age | 0.977** | 1.023*** | 0.964*** | 1.010 |
| | (0.010) | (0.007) | (0.013) | (0.009) |
| gender | 0.057*** | 0.112*** | 0.023*** | 0.062*** |
| | (0.053) | (0.092) | (0.029) | (0.062) |
| gender_age | 1.050*** | 1.029** | 1.074*** | 1.039*** |
| | (0.018) | (0.012) | (0.024) | (0.015) |
| Extroversion | 0.790** | 0.870** | 0.803 | 0.838** |
| | (0.094) | (0.053) | (0.110) | (0.062) |
| Neuroticism | 0.973 | 0.956 | 1.054 | 0.916 |
| | (0.105) | (0.055) | (0.148) | (0.061) |
| Openness | 1.094 | 1.025 | 0.920 | 0.941 |
| | (0.110) | (0.068) | (0.137) | (0.073) |
| Agreeableness | 1.271** | 1.057 | 1.311** | 1.185* |
| | (0.135) | (0.086) | (0.165) | (0.122) |
| Conscientiousness | 0.890 | 0.943 | 0.980 | 0.954 |
| | (0.106) | (0.088) | (0.133) | (0.106) |
| POC | | | 3.498*** | 2.695*** |
| | | | (0.902) | (0.477) |
| Observations | 1228 | 2520 | 739 | 1625 |
| Pseudo $R^2$ | 0.053 | 0.050 | 0.113 | 0.075 |

The odds ratio is reported.

Robust standard errors clustered by pair in each round are presented in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.4: Hypothesis 3 IFR4 vs. FR4

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Raven $\geq 10$ | Raven $< 10$ | Raven $\geq 10$ | Raven $< 10$ |
| Random | 2.786*** | 1.257* | 2.346*** | 1.204 |
| | (0.678) | (0.169) | (0.675) | (0.171) |
| round | 1.084 | 0.927* | 1.068 | 0.934 |
| | (0.097) | (0.039) | (0.102) | (0.042) |
| stage | 0.899* | 0.906*** | 0.959 | 0.975 |
| | (0.054) | (0.035) | (0.051) | (0.042) |
| age | 0.964*** | 0.991 | 0.953** | 0.987 |
| | (0.012) | (0.008) | (0.019) | (0.009) |
| gender | 0.218* | 0.285** | 0.021*** | 0.215** |
| | (0.193) | (0.182) | (0.028) | (0.155) |
| gender×age | 1.028 | 1.021** | 1.075*** | 1.027** |
| | (0.019) | (0.010) | (0.029) | (0.012) |
| Extroversion | 0.866 | 0.864** | 0.837 | 0.841** |
| | (0.094) | (0.060) | (0.101) | (0.067) |
| Neuroticism | 1.248** | 0.904 | 1.240** | 0.929 |
| | (0.118) | (0.069) | (0.121) | (0.076) |
| Openness | 0.796* | 1.099 | 0.834 | 1.052 |
| | (0.099) | (0.082) | (0.120) | (0.088) |
| Agreeableness | 1.020 | 1.376*** | 1.064 | 1.329** |
| | (0.113) | (0.157) | (0.121) | (0.182) |
| Conscientiousness | 1.408** | 0.701*** | 1.270* | 0.778** |
| | (0.209) | (0.079) | (0.184) | (0.093) |
| POC | | | 15.007*** | 2.594*** |
| | | | (3.959) | (0.443) |
| Observations | 1302 | 2796 | 1007 | 2131 |
| Pseudo $R^2$ | 0.093 | 0.034 | 0.311 | 0.062 |

The odds ratio is reported.

Robust standard errors clustered by pair in each round are presented in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.5: Hypothesis 4 IFR2 vs. FR2

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Raven $\geq 10$ | Raven $< 10$ | Raven $\geq 10$ | Raven $< 10$ |
| Random | 2.110*** | 1.122 | 1.614* | 1.160 |
| | (0.382) | (0.141) | (0.460) | (0.201) |
| round | 0.936** | 0.919*** | 0.992 | 0.956* |
| | (0.029) | (0.019) | (0.041) | (0.024) |
| stage | 0.850* | 1.032 | 1.174 | 1.011 |
| | (0.080) | (0.067) | (0.198) | (0.122) |
| age | 1.007 | 1.049*** | 1.003 | 1.035*** |
| | (0.013) | (0.007) | (0.021) | (0.009) |
| gender | 0.383 | 5.226** | 0.537 | 4.084 |
| | (0.272) | (4.328) | (0.580) | (3.975) |
| gender×age | 1.017 | 0.975** | 1.017 | 0.979 |
| | (0.016) | (0.012) | (0.025) | (0.014) |
| Extroversion | 1.023 | 1.075 | 1.048 | 0.985 |
| | (0.099) | (0.067) | (0.146) | (0.081) |
| Neuroticism | 0.973 | 0.831*** | 0.977 | 0.826*** |
| | (0.074) | (0.043) | (0.102) | (0.060) |
| Openness | 1.281*** | 1.211** | 1.276** | 1.219* |
| | (0.105) | (0.099) | (0.139) | (0.127) |
| Agreeableness | 1.015 | 0.753*** | 0.991 | 0.786*** |
| | (0.131) | (0.050) | (0.174) | (0.071) |
| Conscientiousness | 0.736*** | 1.200** | 0.792 | 1.191 |
| | (0.078) | (0.103) | (0.113) | (0.139) |
| POC | | | 3.687*** | 1.769*** |
| | | | (0.911) | (0.303) |
| Observations | 1254 | 2280 | 640 | 1160 |
| Pseudo $R^2$ | 0.045 | 0.072 | 0.094 | 0.059 |

The odds ratio is reported.

Robust standard errors clustered by pair in each round are presented in parentheses.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

# D Additional regression analyses 2

*We report the results of logit regressions when we consider the Raven's score of a participant directly and test the hypotheses 1(a) to 4(b). In the results, "raven" is the score of Raven's test of a participant (Table 11).*

Table D.6: Hypotheses 1(a) and 1(b): Analysis of the odds ratio

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | All data | Raven≥9 | Raven<9 | All data | Raven≥9 | Raven<9 |
| Four | 1.432*** | 3.076*** | 0.838 | 1.219 | 2.892*** | 0.715* |
| | (0.191) | (0.735) | (0.125) | (0.182) | (0.962) | (0.129) |
| stage | 0.918* | 0.884** | 0.973 | 0.976 | 0.958 | 1.023 |
| | (0.041) | (0.050) | (0.043) | (0.042) | (0.045) | (0.052) |
| round | 0.946** | 1.028 | 0.893*** | 0.988 | 1.139** | 0.910*** |
| | (0.023) | (0.042) | (0.023) | (0.027) | (0.059) | (0.031) |
| age | 1.007 | 1.009 | 1.003 | 1.008 | 1.011 | 0.999 |
| | (0.005) | (0.008) | (0.007) | (0.006) | (0.010) | (0.009) |
| gender | 1.337*** | 1.039 | 1.589*** | 1.281* | 1.269 | 1.579*** |
| | (0.145) | (0.197) | (0.218) | (0.168) | (0.298) | (0.266) |
| raven | 1.070*** | 1.322*** | 0.992 | 1.056* | 1.276*** | 1.032 |
| | (0.026) | (0.082) | (0.037) | (0.030) | (0.099) | (0.047) |
| Extroversion | 1.061 | 1.179* | 1.005 | 1.006 | 1.319*** | 0.863 |
| | (0.067) | (0.105) | (0.093) | (0.072) | (0.129) | (0.091) |
| Neuroticism | 1.066 | 1.293*** | 0.984 | 1.071 | 1.246** | 0.985 |
| | (0.057) | (0.108) | (0.065) | (0.062) | (0.119) | (0.076) |
| Openness | 0.945 | 0.905 | 0.914 | 1.010 | 1.050 | 1.014 |
| | (0.070) | (0.100) | (0.092) | (0.087) | (0.130) | (0.125) |
| Agreeableness | 0.944 | 0.737*** | 1.114 | 0.911 | 0.611*** | 1.212* |
| | (0.067) | (0.080) | (0.096) | (0.081) | (0.075) | (0.135) |
| Conscientiousness | 1.044 | 1.100 | 0.976 | 1.044 | 1.098 | 0.905 |
| | (0.082) | (0.144) | (0.119) | (0.092) | (0.149) | (0.132) |
| POC | | | | 4.231*** | 14.014*** | 1.608** |
| | | | | (0.739) | (3.364) | (0.303) |
| Observations | 3884 | 1798 | 2086 | 2574 | 1238 | 1336 |
| Pseudo $R^2$ | 0.024 | 0.073 | 0.022 | 0.092 | 0.281 | 0.030 |

Robust standard errors clustered by pair in each round are presented in parentheses.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table D.7: Hypotheses 2(a) and 2(b)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | All data | Raven≥9 | Raven<9 | All data | Raven≥9 | Raven<9 |
| Four | 1.152 | 1.526** | 0.907 | 1.012 | 1.406 | 0.778 |
| | (0.151) | (0.320) | (0.150) | (0.164) | (0.373) | (0.166) |
| stage | 0.864*** | 0.843** | 0.875*** | 0.968 | 0.920 | 0.987 |
| | (0.035) | (0.056) | (0.045) | (0.065) | (0.101) | (0.091) |
| round | 0.928*** | 0.924** | 0.928** | 0.954* | 0.949 | 0.957 |
| | (0.022) | (0.034) | (0.027) | (0.026) | (0.045) | (0.032) |
| age | 1.020*** | 0.992 | 1.041*** | 1.016*** | 0.985* | 1.035*** |
| | (0.005) | (0.007) | (0.007) | (0.006) | (0.009) | (0.008) |
| gender | 0.670*** | 0.588*** | 0.764* | 0.694*** | 0.553** | 0.853 |
| | (0.075) | (0.120) | (0.116) | (0.091) | (0.147) | (0.148) |
| raven | 0.969* | 0.936 | 1.018 | 0.961* | 0.928 | 1.024 |
| | (0.018) | (0.047) | (0.044) | (0.022) | (0.061) | (0.051) |
| Extroversion | 0.853*** | 0.767*** | 0.914 | 0.836*** | 0.781** | 0.875* |
| | (0.046) | (0.066) | (0.059) | (0.051) | (0.078) | (0.067) |
| Neuroticism | 0.958 | 0.922 | 0.989 | 0.970 | 0.967 | 0.977 |
| | (0.046) | (0.074) | (0.063) | (0.055) | (0.101) | (0.072) |
| Openness | 1.082 | 1.006 | 1.039 | 0.995 | 0.868 | 0.975 |
| | (0.056) | (0.084) | (0.075) | (0.060) | (0.101) | (0.082) |
| Agreeableness | 1.147** | 1.376*** | 0.990 | 1.212*** | 1.319** | 1.139 |
| | (0.071) | (0.139) | (0.088) | (0.088) | (0.160) | (0.122) |
| Conscientiousness | 0.873* | 0.911 | 0.943 | 0.909 | 1.100 | 0.897 |
| | (0.062) | (0.095) | (0.103) | (0.073) | (0.133) | (0.115) |
| POC | | | | 2.934*** | 3.867*** | 2.587*** |
| | | | | (0.449) | (0.791) | (0.530) |
| Observations | 3748 | 1686 | 2062 | 2364 | 1048 | 1316 |
| Pseudo $R^2$ | 0.035 | 0.047 | 0.052 | 0.062 | 0.099 | 0.068 |

The odds ratio is reported.

Robust standard errors clustered by pair in each round are presented in parentheses.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table D.8: Hypotheses 3(a) and 3(b)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | All data | Raven≥9 | Raven<9 | All data | Raven≥9 | Raven<9 |
| Random | 1.752*** | 2.500*** | 1.105 | 1.502*** | 2.052*** | 1.019 |
| | (0.229) | (0.488) | (0.163) | (0.182) | (0.430) | (0.166) |
| stage | 0.906** | 0.888** | 0.922* | 0.975 | 0.945 | 0.999 |
| | (0.040) | (0.048) | (0.039) | (0.041) | (0.044) | (0.046) |
| round | 0.973 | 1.117 | 0.853*** | 0.976 | 1.116 | 0.862*** |
| | (0.044) | (0.077) | (0.042) | (0.041) | (0.080) | (0.045) |
| age | 1.004 | 0.989 | 1.021*** | 1.005 | 0.989 | 1.022*** |
| | (0.005) | (0.007) | (0.007) | (0.006) | (0.008) | (0.008) |
| gender | 0.986 | 0.803 | 1.046 | 0.954 | 0.754 | 1.107 |
| | (0.112) | (0.159) | (0.147) | (0.117) | (0.162) | (0.173) |
| raven | 1.065*** | 1.051 | 1.191*** | 1.041* | 1.031 | 1.157*** |
| | (0.022) | (0.051) | (0.051) | (0.024) | (0.058) | (0.053) |
| Extroversion | 0.914 | 0.960 | 0.753*** | 0.886* | 0.964 | 0.744*** |
| | (0.055) | (0.091) | (0.054) | (0.056) | (0.098) | (0.062) |
| Neuroticism | 1.082 | 1.169* | 0.999 | 1.100 | 1.123 | 1.024 |
| | (0.066) | (0.102) | (0.070) | (0.065) | (0.097) | (0.077) |
| Openness | 0.934 | 0.837* | 1.175* | 0.933 | 0.899 | 1.127 |
| | (0.057) | (0.088) | (0.106) | (0.065) | (0.112) | (0.114) |
| Agreeableness | 1.121 | 0.923 | 1.744*** | 1.108 | 0.867 | 1.754*** |
| | (0.092) | (0.090) | (0.188) | (0.100) | (0.086) | (0.216) |
| Conscientiousness | 0.946 | 1.319** | 0.531*** | 0.987 | 1.370*** | 0.552*** |
| | (0.074) | (0.154) | (0.070) | (0.081) | (0.158) | (0.075) |
| POC | | | | 4.597*** | 8.395*** | 2.398*** |
| | | | | (0.745) | (1.800) | (0.469) |
| Observations | 4098 | 1950 | 2148 | 3138 | 1505 | 1633 |
| Pseudo $R^2$ | 0.025 | 0.059 | 0.067 | 0.103 | 0.211 | 0.088 |

The odds ratio is reported.

Robust standard errors clustered by pair in each round are presented in parentheses.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table D.9: Hypotheses 4(a) and 4(b):

| | (1) All data | (2) Raven≥9 | (3) Raven<9 | (4) All data | (5) Raven≥9 | (6) Raven<9 |
|---|---|---|---|---|---|---|
| Random | 1.378*** | 1.911*** | 1.182 | 1.381** | 1.634* | 1.185 |
| | (0.148) | (0.331) | (0.153) | (0.200) | (0.420) | (0.216) |
| stage | 0.970 | 0.915 | 1.005 | 1.042 | 1.123 | 1.033 |
| | (0.052) | (0.079) | (0.068) | (0.102) | (0.171) | (0.128) |
| round | 0.927*** | 0.930** | 0.921*** | 0.964* | 0.998 | 0.953* |
| | (0.016) | (0.027) | (0.020) | (0.021) | (0.038) | (0.025) |
| age | 1.023*** | 1.024*** | 1.027*** | 1.018*** | 1.021* | 1.019** |
| | (0.005) | (0.008) | (0.008) | (0.007) | (0.011) | (0.009) |
| gender | 1.010 | 0.696* | 1.023 | 1.094 | 0.834 | 1.087 |
| | (0.113) | (0.135) | (0.146) | (0.167) | (0.246) | (0.203) |
| raven | 1.002 | 1.269*** | 0.911** | 0.982 | 1.213** | 0.946 |
| | (0.021) | (0.080) | (0.033) | (0.027) | (0.106) | (0.047) |
| Extroversion | 1.023 | 0.891 | 1.112 | 0.979 | 0.987 | 0.990 |
| | (0.053) | (0.074) | (0.073) | (0.069) | (0.118) | (0.082) |
| Neuroticism | 0.895*** | 1.027 | 0.811*** | 0.887** | 1.036 | 0.795*** |
| | (0.038) | (0.079) | (0.046) | (0.052) | (0.110) | (0.061) |
| Openness | 1.260*** | 1.300*** | 1.144 | 1.288*** | 1.304** | 1.149 |
| | (0.072) | (0.109) | (0.094) | (0.096) | (0.149) | (0.119) |
| Agreeableness | 0.872** | 1.046 | 0.769*** | 0.873* | 0.932 | 0.820** |
| | (0.049) | (0.128) | (0.052) | (0.066) | (0.159) | (0.075) |
| Conscientiousness | 0.994 | 0.767*** | 1.198** | 0.994 | 0.805 | 1.231* |
| | (0.059) | (0.076) | (0.107) | (0.080) | (0.110) | (0.147) |
| POC | | | | 2.180*** | 4.427*** | 1.478** |
| | | | | (0.317) | (0.996) | (0.276) |
| Observations | 3534 | 1534 | 2000 | 1800 | 781 | 1019 |
| Pseudo $R^2$ | 0.041 | 0.049 | 0.066 | 0.056 | 0.109 | 0.050 |

The odds ratio is reported.

Robust standard errors clustered by pair in each round are presented in parentheses.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$