Economics Working Paper Series

2020/017

Theories of reasoning and focal point play with a matched non-student sample

Zhixin Dai, Jiwei Zheng and Daniel John Zizzo

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

Theories of reasoning and focal point play with a matched non-student sample

Zhixin Dai
China Financial Policy Research Center, School of Finance, Renmin University of China, 59 Zhongguancun Street, Beijing 100872, China

Jiwei Zheng[1]
Lancaster University Management School, Lancaster University, Lancaster, LA1 4YX, UK.

Daniel John Zizzo
School of Economics, University of Queensland, St Lucia Qld 4072, Australia

September 2020

Abstract

We present a coordination game experiment testing the robustness of the predictive power of *level-k reasoning* and *team reasoning* in a sample of Chinese tax administrators that is matched for likely socio-economic characteristics with our student sample. We show how the incidence of coordination game play is virtually identical between Chinese tax administrators and university students. However, relatively to non-students, students are comparatively more attracted by the focal point under team reasoning when this has equal payoffs and the other outcomes do not.

Keywords: external validity, non-student sample, focal points, team reasoning, level-k, coordination games.

JEL Classification: C72, C78, C91.

---

[1] Corresponding author. Jiwei Zheng's work on the project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No. 670103. The project has received financial support from the National Natural Science Foundation of China (Grant No. 71703161). We thank Bob Sugden, Ted Turocy and seminar participants at the UEA CBESS group meetings for valuable feedback. The usual disclaimer applies.

1. Introduction

There has been no research so far to show whether the comparative predictive power of different theories of coordination games, supported by experiments with student samples, generalize to a non-student sample. This paper presents the first experiment attempting to test this. Specifically, we test the robustness of the predictive power of the two most established theories of coordination games with focal points, namely *level-k reasoning* (Stahl and Wilson, 1994; Nagel, 1995; Costa-Gomes et al., 2001) and *team reasoning* (Sugden, 1993; Bacharach, 1999, 2006), in a sample of Chinese tax administrators that is likely to be matched in socio-economic characteristics with our university student sample.

Our experiment contained four tacit coordination games, using both student and non-student subjects. The student subjects were from Renmin University of China and non-student subjects were Chinese tax administrators. Tax administrators in China are a highly educated professional sample, and, as such, a good match in socio-economic characteristics to that of university students from the good universities that normally make up the subject pool for coordination games experiments. Renmin University students are an especially good match for tax administrators since it has a long-standing tradition of producing graduates who end up in highly qualified public sector positions such as that of tax administrators. Consequently, using tax administrators allows us to test the robustness of the predictive power of level-k and team reasoning using a lower boundary methodology.

Specifically, if systemic behavioral change is observed using non-student subjects who broadly have similar socio-economic characteristics as student ones, then it is reasonable to infer that the non-robustness conclusion will maintain in experiments using ‘more different’ non-student samples. Assume instead that systemic behavioral change is not observed. This would be reassuring for the value of laboratory experiments but also opens up the way for further non-student sample experiments identifying to what extent the external validity of laboratory experiments inferences can be pushed further within this domain, and what further factors may lead to its failure.

Traditional game theory based on the assumption of best-response reasoning cannot help individuals identify a unique pure strategy to maximize their chance of coordination in coordination games with more than one Nash equilibrium. However, various experimental evidence has shown that players often manage to use some salient properties of the game (i.e. *cues*) to converge their expectations on a unique equilibrium (i.e. the *focal point*); as a result, they achieve higher coordination success than what the traditional game theory predicts (Schelling, 1960; Mehta et al., 1994; Bacharach and Bernasconi, 1997; Crawford et al., 2008; Bardsley et al., 2010).

In recent years, there has been an increasing interest in theories that can explain behavior in coordination games. Team reasoning theory postulates that individuals treat themselves and their partners as a team. When dealing with coordination problems, team reasoners try to find out the best solution (or the best rule) from the viewpoint of the team, aiming to either maximize team utility or achieve mutually beneficial outcomes. Level-k theory divides players into different thinking levels. It assumes that the lowest level – level-0 – players make decisions non-strategically, whereas higher level players best respond to their beliefs about the behavior of the players who are one level below them. In different models of level-k, level-0 players' non-strategic behavior is sometimes assumed to be different. In some models, it refers to choosing randomly (Stahl and Wilson, 1994), while in others, it refers to choosing the strategy favoring themselves with a probability greater than 0.5, and when there are more than one such strategy, level-0 players will bias towards the strategy with some salient but payoff-irrelevant labels (if there is any) attached to it (Crawford et al., 2008).

Experimental evidence suggests that the predictive power of team reasoning and level-k is context-dependent (Bardsley et al., 2010; Faillo et al., 2017). Individuals are more likely to use team reasoning in coordination games in which players' interests are perfectly aligned, or in games in which the Nash equilibrium suggested by team reasoning has more equal payoffs than the payoffs in other equilibria. Individuals are more likely to use level-k reasoning, for example, in coordination games involving conflicts of interest, such Battle of the Sexes (van Elten and Penczynski, 2020; Crawford et al., 2008; Faillo et al., 2017; Isoni et al., 2013, 2020).

All the experimental findings in relation to the predictive power of level-k and team reasoning are based on experiments using student subjects. Economists have questioned whether using student subjects can lead academic experiments to generate systematically biased results. For example, Henrich et al. (2010) view student subjects as a major hindrance to generalizing results derived from experimental studies, as students are sometimes believed to be psychologically unusual and not representative to the general population. Conversely, Gächter (2010) suggests that, although the right choice of the subject pool depends on the research questions, at least in the domain of experimental economics, students are often the best subject pool, especially for studies aiming to test theories assuming cognitive sophistication. Belot et al. (2015) contains a review of experimental evidence on student vs. non-student samples in a range of games, though not coordination games; they present the results of an experiment showing that non-students are more selfish and less rational than students from the two universities based in Oxford. They have a beauty contest game where they test level-k and find that students tend to have higher levels of reasoning, an effect that disappears once one controls for age. Bosch-Domenech et al. (2002) instead found comparable level-k results with student and non-sample samples in beauty contest games.

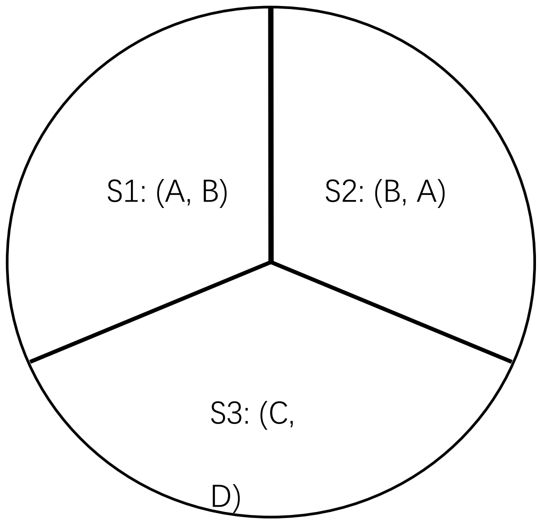
The debate about student subjects has started to affect subject recruitments in focal point studies. Jackson and Xing (2014) stated that they recruited subjects using Amazon Mechanical Turk because they believe university students may not be a representative subject pool for their research involving coordination and bargaining games with focal points. Despite the concern, and as noted earlier, ours is the first paper investigating the comparative predictive power of different theories of behavior in coordination games with non-students.

Our main finding is one of same coordination rates and robustness of predictive power of different theories between students and non-students, with one exception that will be discussed later. Sections 2 and 3 describe the experimental design and results, respectively. A brief discussion and conclusions are in section 4.

2. Experimental design

We employ four two-player coordination games in our experiment. In each game, subjects see a pie (see Figure 1) with three slices of equal size. We denote the three slices as S1, S2, and S3. Each slice contains two numbers separated by a comma. Each subject needs to choose a pie slice without having any communication with her partner. If she and her partner choose the same slice, they will earn a positive amount shown on that slice. From each subject's perspective, the amount she could earn is always the number shown on the left side of the comma, and the amount her partner could earn is the number shown on the right. If the two subjects choose different slices, they earn nothing. The pie is randomly rotated across subjects, so the position of the slices could be different between a pair of subjects. This setting allows us to minimize the possibility of bringing payoff-irrelevant cues (e.g. the position of the slice) into the games.

Figure 1: An example of a coordination game



Notes: ½ A, ½ B <C, D < A, B.

The payoffs shown on each slice in the four games are reported in Table 1. They are variations of payoff pairs used in Faillo et al. (2017). In each game, the payoffs on S3 are always lower but more than half than those on the other two slices. Under alternative interpretations of level-k theory, this feature allows us to distinguish level-k reasoners and team reasoners. Level-k theory commonly assumes that level-0 players think non-strategically. Specifically, level-0 players are biased towards choosing whichever strategy gives them the highest material payoff without considering their partners' behavior. In our experiment, going for the highest material payoff leads to a prediction that level-0 players will not choose S3, since it is strongly Pareto dominated by S1 and S2 (i.e. C, D < A, B). Consequently, higher-level players who anchor their beliefs on the behavior of lower-level players will not choose S3, either. Alternatively, we can reach the same conclusion if we follow Crawford et al.'s (2008) assumption that level-0 players only exist in players' minds, in which case, even if level-0 players are assumed to *bias towards* the higher payoff (i.e. choose the higher-payoff options with a probability greater than 0.5) or play randomly, higher-level players will choose S1 or S2 rather than S3 to maximize their payoff, and S3 is never played. However, team reasoning predicts exactly the opposite. This is because, from the team's perspective, S1 and S2 are isomorphic. In different games, the payoff pairs on these two slices are either the same or symmetric between players. Since team reasoners are unable to distinguish S1 and S2, the ex-ante expected payoffs for these two slices equal half of the payoffs shown on them. Consequently, the best rule to follow, from a team reasoner's perspective, is to coordinate on S3 since, ex-ante, its expected payoffs for the team dominate the payoffs on the other two slices (i.e. ½ A, ½ B <C, D). Accordingly, S3 is the *team-optimal slice*, and (S3, S3) is the focal point.

The main difference across the four games is whether or not the payoffs are equal between players. In Game 1, payoffs between players are equal in S1 and S2, but unequal in S3. In

Games 2 and 3, payoffs are unequal and equal in all slices, respectively. In Game 4, payoffs are unequal in S1 and S2, but equal in S3.

Table 1: Payoffs on each pie slice

	S1 (A, B)	S2 (B, A)	S3 (C, D)
Game 1	100，100	100，100	90，80
Game 2	100，90	90，100	80，70
Game 3	100，100	100，100	90，90
Game 4	100，90	90，100	80，80

Notes: ½ A, ½ B < C, D < A, B. Payoffs were in Chinese Yuan (RMB; 1 RMB ≅ 0.15 USD at the time of the experiment).

Experimental instructions can be found in Appendix A. We had two treatments. In the *Tax-admin* treatment, conducted in April 2019, subjects were tax administrators who volunteered to take part while attending a 7-day training school in Renmin university of China (n = 62). In the *Student* treatment, conducted in May 2019, subjects were university students at the Renmin University of China, Beijing (n = 42). The experiment was computerized by using zTree (Fischbacher, 2007). Upon arrival, subjects were randomly assigned to a terminal by drawing a tag from a bag. The experimenter read the instructions aloud. Subjects were asked to go through the instructions with the experimenter and answer a brief questionnaire to make sure that instructions were correctly understood. Subjects were not given feedback until the end of the experiment. At the end of the experiment, one game was randomly picked for each subject. Subjects' final payments equaled their earnings of that picked game plus a 50 RMB participation fee. They had to complete a demographic questionnaire before getting paid. Sessions lasted approximately 60 minutes. The average payment was 94.33 RMB (S.D. 7.54), equivalent to approximately 14 USD at the time of the experiment.

Both level-k and team reasoning are silent on how their predictive power will comparatively change in predicting students' and non-students' behavior in tacit coordination games, and, as noted, experimental evidence on this has been non-existent so far. We make a general prediction of external validity of laboratory experiments, which leads us to predict that the comparative explanatory power of level-k and team reasoning will not change between our student and non-student samples, particularly in the presence of samples with comparable socio-economic characteristics.

3. Results

Figure 1 reports the proportion of times the team-optimal slice, S3, is chosen, and Table 2 shows the same data as Figure 1, plus the proportions of S1 and S2 choices. Results in both Figure 1 and Table 2 are broken down by game and treatment. In all games, except Game 4 in the Student treatment, only a small proportion of subjects chose S3. This finding is consistent with Faillo et al. (2017), who used Italian university students as subjects and find that team reasoning is inhibited in games in which the team-optimal equilibrium is dominated by the other two equilibria.

Result 1: *Only a small proportion of subjects chose S3. Subjects' behavior is more in line with level-k rather than team reasoning.*

In Game 1, payoffs are equal in S1 and S2 but not in S3. Only 18% of students and 19% of tax administrators chose S3. These proportions increase to 21% and 26% respectively when payoffs in S3 also become equal (Game 3). However, the changes are not statistically significant (McNemar test $p = 0.317$ and $p = 0.180$, respectively). In Game 2, in which payoffs in all three slices are unequal, only 10% and 7% of the tax administrators and students, respectively, chose S3. These proportions increase to 24% in the Tax-admin treatment and 48% in the Student treatment when the payoffs in S3 are equal (Game 4). The differences are statistically significant (McNemar test, $p = 0.003$, $p<0.001$, respectively). The equal payoffs in the team-optimal slice can increase the choice of S3 only when the payoffs in the other two slices are unequal.

Result 2: *The predictive power of team reasoning is facilitated if the team-optimal equilibrium gives more equal payoffs than the others.*

Figure 1: Proportion of the team-optimal slice (S3) choices by game and treatment

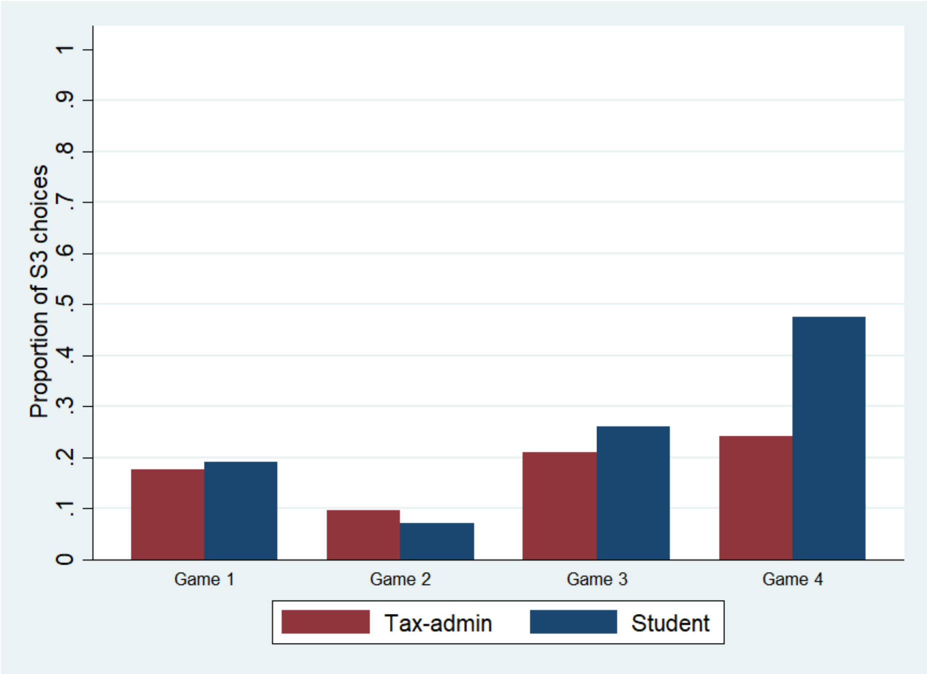


Table 2: Proportion of each slice's choices by game and treatment

	Tax admin (62 obs.)			Student (42 obs.)		
	S1 (LK)	S2 (LK)	S3 (TR)	S1 (LK)	S2 (LK)	S3 (TR)
Game 1	0.48	0.34	0.18	0.43	0.38	0.19
Game 2	0.39	0.52	0.10	0.43	0.50	0.07
Game 3	0.37	0.42	0.21	0.33	0.40	0.26
Game 4	0.37	0.39	0.24	0.29	0.24	0.48
Average	0.40	0.42	0.18	0.37	0.38	0.25

*LR and TR represent the choice(s) predicted by level-k and team reasoning, respectively.

In Games 1-3, tax administrators and students do not behave differently in the respect of choosing the team-optimal slice. (χ^2 test, $p = 0.866$, $p = 0.652$, $p = 0.535$, in Games 1-3, respectively). But in Game 4, in which the team-optimal slice has equal payoffs but the other two slices do not, nearly half (48%) of the students chose S3, and this proportion is twice as many as the proportion of tax administrators choosing the same slice (24%) (χ^2 test, p=0.013).

Students' behavior is more consistent with team reasoning than tax administrators' in a game in which only the payoffs in the team-optimal slice are equal. These results are replicated in regression analysis controlling for *period* and *gender* (see Appendix B for more details). Since most tax administrators are older than students, to avoid multicollinearity caused by *treatment* and *age*, the latter was excluded from the regressions (the correlation between *treatment* and *age* is 0.9). To test how age affects coordination behavior, we added *age* to the regression using Tax-admin data only[2]. We do not find any evidence showing that *age* has an effect on choosing the team-optimal slice.

Result 3: *Students and tax administrators do not behave differently in three out of four games. Students' behavior can be better explained by team reasoning than tax administrators in a game in which the team-optimal equilibrium gives more equal payoffs than the others.*

Table 3 reports the expected coordination rate (ECR) in each game. ECR is calculated as follows.

In Game 1 and Game 2, if a subject's payoff on S3 is higher than her partner's payoff, then we call this subject Player 1 and her partner Player 2. In Game 4, since the payoffs in S3 are the same, we define the player whose payoff is higher in S1 as Player 1, and the other Player 2. In each of these games, the ECR for each slice is calculated by multiplying the proportion of Player 1s who chose that slice by the proportion of Player 2s who chose that slice in that game. The total ECR is the sum of the ECR for each slice. We do not define Player 1 and Player 2 in

[2] The variable *age* records subjects' choices among six options. 1="18-24"; 2="25-34"; 3="35-49"; 4="50-64"; 5="65+"; 6= "Prefer not to say". In Student treatment, 97.62% subjects chose option 1, so we do not have enough variability to run the same regression using Student data.

Game 3, since payoffs in Game 3 are equal in all slices. In Game 3, we calculate ECR using the same way as Mehta et al. (1994) and Sitzia and Zheng (2019) showing below:

$$ECR = \sum_i ECR_i = \sum_i \frac{n_i(n_i - 1)}{N(N-1)}$$

where N represents the total number of subjects, n_i represents the number of subjects who chose that slice, and ECR_i represents the expected coordination rate for slice i.[3]

Table 3: Expected coordination rate by game and treatment

	Tax-admin	Student
Game 1	0.37	0.35
Game 2	0.42	0.41
Game 3	0.35	0.33
Game 4	0.34	0.36

The values of ECR in Tax-admin and Student treatments are nearly the same. In all games, the ECR differences across treatments are negligible (0.01-0.02). Although students chose S3 more frequently than tax administrators in Game 4, this difference does not translate into higher coordination success.

[3] In our experiment, subjects were randomly paired with one another, and no feedback was given until the end of the experiment. The coordination success/failure in an 'actual' pair is only used as a tool to calculate subjects' earnings in a specific game. In this paper, following Mehta et al. (1994) and Sitzia and Zheng (2019), we compare ECR instead of the 'actual' coordination rates, since the former contains more information than the latter.

Result 4: *In all games, including Game 4, the expected coordination rate in Tax-admin and Student treatments do not differ*.

4. Discussion and Conclusion

In three coordination games out of four, we find the same predictive power of different models of play in coordination games with students and tax administrators. The exception is the game where the team-optimal equilibrium gives more equal payoffs than the others; in this game, students' behavior can be better explained by team reasoning than tax administrators'. Note that this result is not incompatible with Cooper and Kagel's (2016) literature review conclusion that social preferences are at least as large among non-students then they are among students; none of the papers they rely on use coordination games, which is not surprising as coordination game play cannot be explained by social preference models. Age does not appear to be an explanatory factor of the difference, for within the sample of tax administrators (where there is age variability) it has no predictive power. It might be that payoff equality can facilitate team reasoning more for students than for tax administrators. Another possibility is that the reason does not have to do with students but rather with our tax administrator sample. Specifically, it may have to do with how our tax administrators perceive that colleagues will perceive sacrificing payoff to go to an equal outcome. The Chinese income tax brackets are progressive in income (like those in many other countries), but do not imply that a wealthier person is taxed to have the same income as less wealthy person. Having noted the difference, we also note that in this game, like in the others, coordination rates are the same with both students and tax administrators. Clearly, further research is needed.

With the qualification above, our lower boundary methodology generally implies that, for given socio-economic characteristics, we do not observe differences in predictive power of team reasoning relative to level-k when moving from a student to a non-student sample. Future research can extend this testing process to other samples to identify the limits to the external validity of inferences from laboratory experiments, and, specifically, what factors affect this limit. We believe that this is a good methodology to go beyond traditional dichotomies between student and non-student samples, as non-student samples can themselves be very different and generalizations (from laboratory as well as from non-laboratory experiments) need to be done cautiously.

While our results are with Chinese samples, our coordination rates and comparative predictive power of level-k and team reasoning are comparable to those found in the corresponding games of Faillo et al. (2017) with a Western (specifically, Italian) sample. Obviously, future research may wish to look more explicitly and systematically at cross-cultural comparisons. Taken at its face value, our key finding of mostly robustness of the respective predictive power of team reasoning and level-k theories with a non-student sample, speaks to the potential generality of these theories and their underpinning cognitive mechanisms. However, as noted, our tax administrators are comparatively less attracted than students by the focal point under team reasoning when this has equal payoffs and the other outcomes do not.

References

Bacharach, M. (1993). *Variable universe games*. In: Binmore, Ken, Kinmar, Alan, Tani, Piero (Eds), Frontiers of Game Theory. MIT Press, Cambridge.

Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory*. (N. Gold & R. Sugden, Eds.). Princeton University Press, Princeton.

Bardsley, N., Mehta, J., Starmer, C., & Sugden, R. (2010). Explaining focal points: Cognitive hierarchy theory versus team reasoning. *Economic Journal*, *120*(543), 40–79.

Belot, M., Duch, R., & Miller, L. (2015). A comprehensive comparison of students and non-students in classic experimental games. *Journal of Economic Behavior and Organization*, *113*, 26–33.

Bosch-Domenech, A., Montalvo, J. G., Nagel, R., & Satorra, A. (2002). One, Two, (Three), Infinity , ... : Newspaper and lab beauty-contest experiments. *American Economic Review*, *92*(5), 1687–1701.

Cooper, D. J., & Kagel, J. (2016). Other regarding preferences: A survey of experimental results. In J. Kagel and A. Roth (eds.), The Handbook of Experimental Economics, Vol. 2, Princeton: Princeton University Press, 217-289.

Costa-Gomes, M. A., & Crawford, V. P. (2006). Cognition and behavior in two-person guessing games : An experimental study. *American Economic Review*, *96*(5), 1737–1768.

Crawford, V. P., Gneezy, U., & Rottenstreich, Y. (2008). The power of focal points is limited: Even minute payoff asymmetry may yield large coordination failures. *American Economic Review*, *98*(4), 1443–1458.

Elten, J. V., & Penczynski, S. P. (2020). Coordination games with asymmetric payoffs: An experimental study with intra-group communication. *Journal of Economic Behavior and Organization*, *169*, 158–188.

Faillo, M., Smerilli, A., & Sugden, R. (2017). Bounded best-response and collective-optimality reasoning in coordination games. *Journal of Economic Behavior and Organization*, *140*, 317–335.

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

Gächter, S. (2010). (Dis)advantages of student subjects: What is your research question? *Behavioural and Brain Sciences*, *33*(2–3), 92–93.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83.

Isoni, A., Poulsen, A., Sugden, R., & Tsutsui, K. (2013). Focal points in tacit bargaining problems: Experimental evidence. *European Economic Review*, *59*, 167–188.

Isoni, A., Sugden, R., & Zheng, J. (2020). The pizza night game: conflict of interest and payoff inequality in tacit bargaining games with focal points. *European Economic Review*, *127*, 267–272.

Jackson, M. O., & Xing, Y. (2014). Culture-dependent strategies in coordination games. *Proceedings of the National Academy of Sciences*, *111*, 10889–10896.

Mehta, J., Starmer, C., & Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. *American Economic Review*, *84*(3), 658–673.

Nagel, B. R. (1995). Unraveling in guessing games : An experimental study. *American Economic Review*, *85*(5), 1313–1326.

Schelling, T. C. (1960). *The strategy of conflict*. Harvard University Press, Cambridge.

Sitzia, S., & Zheng, J. (2019). Group behaviour in tacit coordination games with focal points – an experimental investigation. *Games and Economic Behavior*, *117*, 461–478.

Stahl, D. O., & Wilson, P. W. (1994). Experimental evidence on players' models of other players. *Journal of Economic Behavior and Organization*, *25*(3), 309–327.

Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behavior. *Social Philosophy and Policy*, *10*(1), 69–89.

Appendix A: Experimental instructions (translated from Chinese)

Introduction

[oral: I will now take you through the instructions, and I will read them out.]

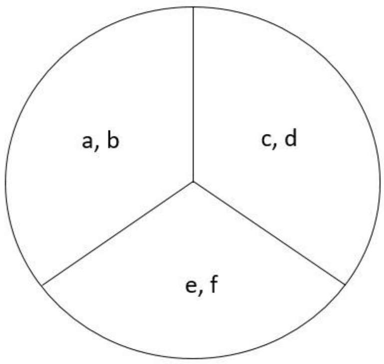
Welcome and thank you for taking part in this experiment. Everyone in the room has exactly the same instructions.

It is important that you remain silent and do not look at what other participants are doing. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. We expect and appreciate your cooperation.

The Pie task

At the beginning of the experiment, you will be matched with another person in the room. You and the other person will not be told each other's identity. Your earnings will depend both on your decision and the decision of the other person.

You and the other person will need to play four Pie tasks in this experiment. An example of a Pie task is shown below.

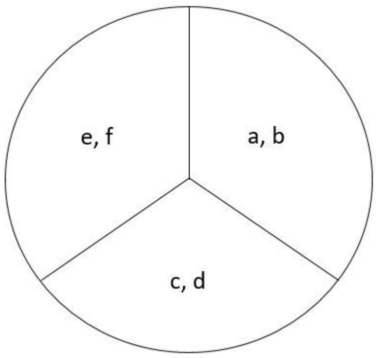


In the pie task, you and the other person will be presented with a three-slice pie, and asked to choose one slice.

There are two amounts shown on each slice, represented by two letters. For simplicity reasons, we use the letters written on each slice to name those slices. We call the slice on the top left slice AB, on the top right slice CD, and at the bottom slice EF. **If you and the other person choose the same slice, you will earn the amount on the left of the comma of the chosen slice, and the other person will earn the amount on the right**. **But if you and the other person choose difference slices, neither of you will earn anything in that task.**

For example, if you and the other person both choose the slice AB, you will earn amount ‘a’, and the other person will earn amount ‘b’. But if you choose the slice AB and the other person chooses slice EF, then you and the other person will earn nothing.

The orientation of the pie is randomly decided. This means that, although you and the other person will see the same pie, its orientation will vary. For example, you may see the pie shown above, while the other person may see the pie shown below.



There is no way for you to know what orientation the pie of the other person sees.

<u>Your earnings</u>

At the end of the experiment, the computer will randomly choose one of the four tasks, and your earnings in that randomly chosen task will be realised. In addition to whatever you have earned in that task, you will be given a participation fee of ¥50.

Comprehension questionnaire

The following questions are meant to check your understanding of the basic rules of the experiment.   If anything is unclear, please raise your hand and an experimenter will come to assist you.

<u>Questions 1 of 4</u>

In the Pie task shown above, if you and the other person both chose slice CD, how much you and the other person earn in that task?

A: You earn c and the other person earns d.

B: You earn d and the other person earns c.

C: You earn a and the other person earns b.

D: You and the other person both earn nothing.

Questions 2 of 4

In the Pie task shown above, if you chose slice CD, and the other person chose slice AB, how much will you and the other person earn in that task?

A: You earn c and the other person earns d.

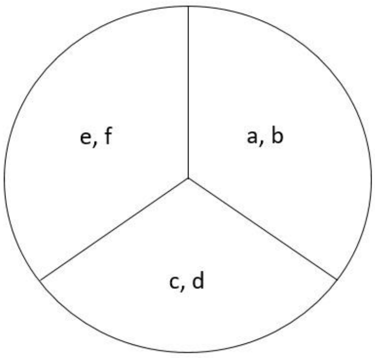
B: You earn d and the other person earns c.

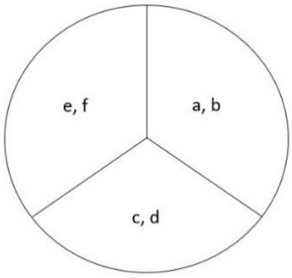
C: You earn a and the other person earns b.

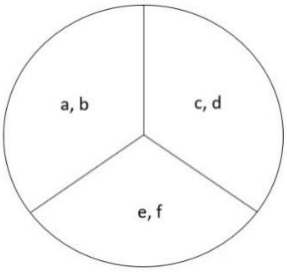
D: You and the other person both earn nothing.

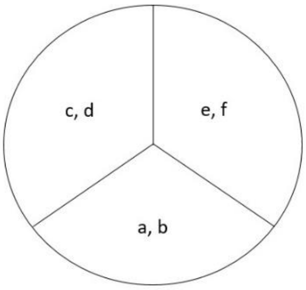
Questions 3 of 4

You see a pie task shown below in a pie task. What will the other person's pie task look like?



A: 

B: 

C: 

D: All three above are possible.

Questions 4 of 4

If you are a chosen subject, your earnings of the experiment will be:

A: The sum of the earnings from all the tasks.

B: The sum of the earnings from all the tasks, plus the participation fee of ¥50.

C: The amount you earn in the randomly selected task, plus the participation fee of ¥50.

D: The amount you earn in the randomly selected task.

Correct answers:

A D D C

Appendix B: Regression results (Marginal effects)

Model: logit model clustering subjects (One regression for each game)

Game	Period (Standard error)	Treatment (Standard error)	Gender (Standard error)
Game 1	-0.102***	0.022	0.051
	(0.04)	(0.81)	(0.08)
Game 2	0.028	-0.041	0.038
	(0.03)	(0.05)	(0.06)
Game 3	0.052	0.023	0.082
	(0.04)	(0.09)	(0.09)
Game 4	-0.031	0.208**	0.042

	(0.04)	(0.10)	(0.10)

n = 104 for each regression.

The dependent variable takes value 1 if the choice is S3 and 0 otherwise.

Period. Variable taking value from 1 to 4.

Treatment: dummy variable taking value 1 for Student treatment and 0 for Tax-admin treatment.

Gender: dummy variable taking value 0 if male and 1 if female.

*** significant at 1%; ** significant at 5%; * significant at 10%

Model: logit model clustering subjects (One regression for each game, using only Tax-admin treatment data)

Game	Period (Standard error)	Age (Standard error)	Gender (Standard error)
Game 1	-0.053	-0.056	0.010
	(0.04)	(0.08)	(0.10)
Game 2	0.022	0.003	0.004
	(0.05)	(0.08)	(0.08)
Game 3	0.084	0.003	0.037
	(0.05)	(0.08)	(0.11)
Game 4	-0.042	-0.013	0.010
	(0.05)	(0.09)	(0.12)

n = 62 for each regression.

The dependent variable takes value 1 if the choice is S3 and 0 otherwise.

Period. Variable taking values from 1 to 4.

Age: 1="18-24"; 2="25-34"; 3="35-49"; 4="50-64"; 5="65+"; 6= "Prefer not to say". No subject chose option 6, and so age was used as a variable between 1 and 5 in each regression, with higher values implying higher age.

Gender: dummy variable taking value 0 if male and 1 if female.

*** significant at 1%; ** significant at 5%; * significant at 10%