

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

Zhixin Dai¹

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

Jiwei Zheng

Lancaster University Management School, Lancaster University, Lancaster, LA1 4YX, UK.

Daniel John Zizzo

School of Economics, The University of Queensland, St Lucia Qld 4072, Australia

September 2023

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, compared to non-students, students are more attracted by the focal point when this offers equal payoffs while the other equilibria do not. Evidence from an additional experiment suggests that this could be because payoff equality promotes team reasoning more effectively among students.

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 demonstrating whether the comparative predictive power of different theories of coordination games, supported by experiments with student samples, generalizes to a non-student sample. This paper presents the first experiment attempting to test this using a Chinese experimental dataset. 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 tax administrators that is likely to be matched in socio-economic characteristics with our university student sample.²

Our experiment featured four tacit coordination games, using both student and non-student subjects. The student subjects were from Renmin University of China, while the non-student subjects were Chinese tax administrators. As highly educated professionals, tax administrators closely match the socio-economic characteristics of students from good universities who typically comprise the subject pool for coordination games experiments. Renmin University students are an especially good match for tax administrators as the university has a long-standing tradition of producing graduates who end up in highly qualified public sector positions such as tax administration. Consequently, the use of tax administrators as subjects enables us to test the robustness of the predictive power of level-k and team reasoning using a ‘lower bound’ approach. By a ‘lower bound’ approach we mean that, if a systematic behavioral change is observed in 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 persist in experiments using ‘more different’ non-student samples. However, if systematic 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 inferences from laboratory experiment apply to them, and what further factors may lead to its failure. Our study provides

² There is a small number of experiments that have looked at coordination games in China (e.g., Stoddard and Leibbrandt, 2014; Mizrahi et al., 2018; Li, 2007; Hu et al., 2020; Peyman and Kelsey, 2021). They are not directly comparable to this paper due to differences in experimental games and objectives.

a crucial initial step in exploring the applicability of the findings regarding team reasoning and level-k in coordination games beyond student populations.

There is a debate about the generalizability of findings with student populations to non-student populations. While Cooper and Kagel's (2016) believe, based on experiments with Western samples, that social preferences are similar among students and non-students, Belot et al. (2015) note that in trust games, dictator games and public good games students tend to be more selfish than their convenience non-student sample. They also find that, in beauty-contest and second price auctions, students tend to exhibit higher levels of reasoning, with a roughly 2 to 1 ratio of equilibrium choices between students and non-students. Snowberg and Yariv (2021) argue that, while there are behavioral differences consistent with Belot et al. (2015), the correlations are similar between MTurk samples and a representative sample of the US population, with some more differences with a student sample. Exadaktylos et al. (2013) emphasize similarities between the two types of samples in relation to dictator, ultimatum, and trust games (see also Falk and Zehnder., 2013 for trust games). The debate about the use of student subjects has begun to influence subject recruitment in focal point studies. Jackson and Xing (2014) have explained their choice to recruit subjects via MTurk, stating that university students may not represent a suitable subject pool for their research, which involves coordination and bargaining games with focal points.

Despite these concerns, our study is the first to investigate the comparative predictive power of different theories of behavior in coordination games with both student and non-student samples. This investigation is particularly valuable for three reasons. First, the mixed nature of the evidence suggests that it may depend on the psychological mechanism involved. Second, while doing tests with representative samples of a country's population is ideal, this is not always possible. This makes a 'lower bound' methodology of the kind proposed here a useful alternative. Third, many economic or managerial decisions are based on specific population samples. These samples may behave differently from a representative sample of the population as a whole, whether because of sample selection or of the effect of experience on particular cohorts.

Ample experimental evidence has shown that players often manage to use some salient properties, or cues, in a coordination game to converge their expectations on a unique equilibrium, often referred to as the *focal point*. Consequently, they achieve a level of coordination success that exceeds the predictions of traditional game theory based on the best response analysis (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 with focal points. The most relevant theories in this field are team reasoning and level-k³. Team reasoning theory postulates that individuals view themselves and their partners as a team. When confronted with coordination problems, these team reasoners strive to identify the best solution (or rule) from the team's perspective, aiming to either maximize team utility or attain mutually beneficial outcomes. In contrast, level-k theory categorizes players into different thinking levels. Players at the lowest level – level-0 – are assumed to make non-strategic decisions, while higher level players best respond to their beliefs about the behaviors of the players one level below them⁴.

Experimental evidence with student samples suggests that the predictive power of team reasoning and level-k is context-dependent (Bardsley et al., 2010; Faillo et al., 2017). Individuals' behavior is better explained by team reasoning in coordination games in which players' interests are perfectly aligned, or when the equilibrium suggested by team reasoning has more equal payoffs. Conversely, their behavior is better explained by level-k in coordination games involving conflicts of interest, such as the 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 concerning the predictive power of level-k and team reasoning are from experiments using student subjects, predominantly from Western countries. Some economists have

³ In the focal point literature, coordination games employed in experiments are structured in a way that certain refinements of the Nash equilibrium solution concept, such as risk dominance, are incapable of providing valuable predictions.

⁴ In different level-k models, the assumptions regarding the non-strategic behavior of level-0 players can be different. In some models, it is assumed to be random choice (Stahl and Wilson, 1994). In others, it is suggested that these players choose the strategy favoring themselves with a probability greater than 0.5. In case where more than one such strategy exists, level-0 players will bias towards the strategy (if there is any) that has a salient, albeit payoff-irrelevant label attached to it (Crawford et al., 2008).

raised concerns about whether the use of student subjects can lead academic experiments to generate systematically biased results. For example, Henrich et al. (2010) perceive student subjects as a major hindrance to generalizing results from experimental studies, considering them to be psychologically atypical and not representative of the general population. Conversely, Gächter (2010) suggests that the choice of a suitable subject pool depends on the research questions researchers aim to investigate. He believes that within the field of experimental economics, students are often the best subjects, especially for experiments designed to test theories that assume 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.

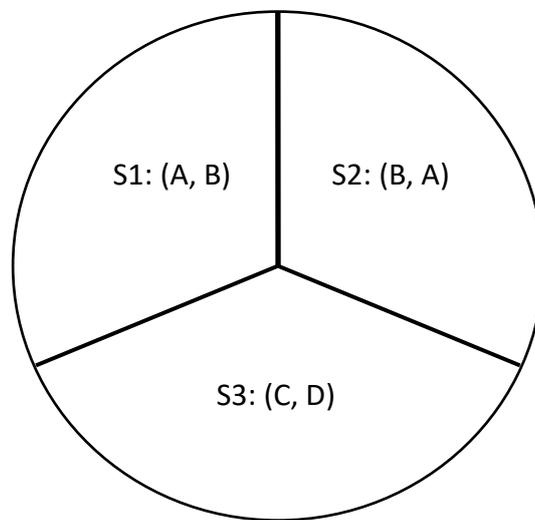
Our main finding reveals a consistency in the coordination rates and the robustness of the predictive power of the two theories between students and non-students, with one exception that we will discuss later. We have successfully replicated findings from previous literature using a Chinese sample. Sections 2 and 3 describe the main experiment's (Experiment 1) design and results, respectively. Section 4 presents the design and results of an additional experiment (Experiment 2) aimed at further elucidating the findings from the original experiment. A brief discussion and conclusions can be found in section 5.

2. Experiment 1: 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: $\frac{1}{2} A, \frac{1}{2} B < C, D < A, B$.

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 payoffs, 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 see these outcomes from their teams' perspective, they 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., $\frac{1}{2} A, \frac{1}{2} B < C, D$). 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: $\frac{1}{2} A, \frac{1}{2} B < C, D < A, B$. Payoffs were in Chinese Yuan (RMB; 1 RMB \cong 0.15 USD at the time of the experiment).

Experimental instructions can be found in Appendix A. The experiment consisted of 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. Experiment 1: 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, along with 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 the results of the corresponding games in 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 equilibria⁵.

Result 1: *Only a small proportion of subjects chose S3. Subjects' behavior is more in line with level-k rather than team reasoning. This result is consistent with the result of a similar experiment using Italian university students as subjects.*

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. In Game 3, in which the payoffs in S3 also become equal, these proportions increase to 21% and 26% for students and tax administrators, respectively. However, these increases 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 become equal (Game 4). The differences in S3 choices between Game 2 and Game 4 are statistically significant (McNemar test, $p = 0.003$, and $p < 0.001$, 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.*

⁵ Games 1-4 in Faillo et al. (2017) correspond to Games 1-4 in our study. The proportions of S3 choices in the four games in their study are 11.9% in Game 1, 1.6% in Game 2, 13.5% in Game 3 and 43.7% in Game 4, respectively.

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

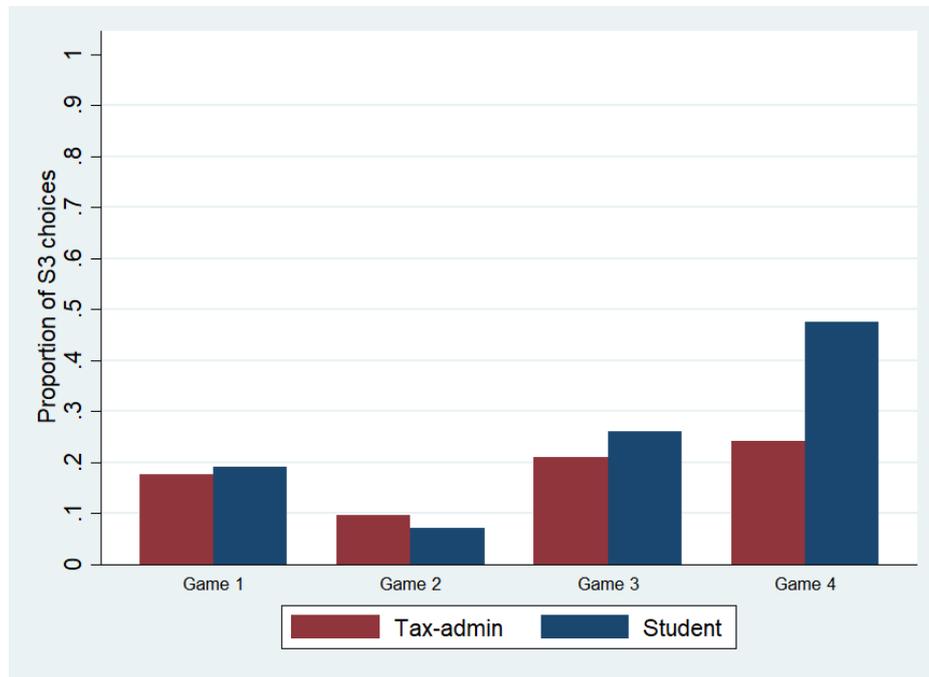


Table 2: Proportion of each slice's choices by game and treatment

	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 whereas 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 only⁶. 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 Game 3, since payoffs in Game 3 are equal in all slices. In Game 3, we calculate ECR using the same approach as Mehta et al. (1994) and Sitzia and Zheng (2019), as shown 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 .⁷

⁶ 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.

⁷ 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.

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.

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

4. Experiment 2: Why did students choose the team-optimal slice more than tax administrators in Game 4?

Even though the behavior difference between students and tax administrators in Game 4 does not result in a big change in ECR, it remains important to understand the potential reasons behind such a difference. Given that the team-optimal equilibrium in Game 4 is also the only equilibrium that provides equal payoffs between players, there could be two possible reasons for the behavior difference. Firstly, it might be that payoff equality promotes team reasoning more effectively among students than tax administrators. Secondly, students could exhibit a much stronger degree of inequality aversion than tax administrators. They might prefer coordinating on the only equal outcome, albeit Pareto dominated, or choose not to coordinate at all.⁸

⁸ A third explanation of the difference could be age related, but this does not appear to be an explanatory factor of the difference, since, within the sample of tax administrators (where there is age variability), it has no predictive power.

To distinguish between these two possible reasons, we conducted Experiment 2, recruiting 36 student subjects from Renmin University of China⁹. The logistics of this experiment were the same as the first one, with the exception that some additional tasks were added before the coordination games.¹⁰

At the beginning of the experiment, subjects were paired, and one member of each pair was randomly selected as an *allocator* to complete four allocation tasks. In each allocation task, the allocator was required to choose between two payoff distributions for themselves and the other subject (i.e., *receiver*), with one distribution more equal but less efficient than the other. The payoff distributions in each allocation task are shown in Table 4. The first number in each distribution indicates the amount of Chinese RMB the allocator could earn, while the second number indicates the amount that the receiver could earn.

Table 4: Payoff distributions in allocation tasks

Allocation task	Payoff distribution 1	Payoff distribution 2
Task 1	(115, 105)	(95, 95)
Task 2	(85, 75)	(65, 65)
Task 3	(105, 115)	(95, 95)
Task 4	(75, 85)	(65, 65)

The payoff distributions used in allocation tasks were selected to correspond to the payoffs in Game 4. Specifically, the first payoff distributions in allocation tasks 1 and 3 were generated by adding 15 RMB to the payoffs in S1 and S2 of Game 4 (viewed from Player 1’s perspective, or S2 and S1 from Player 2’s perspective), and in allocation tasks 2 and 4 were generated by subtracting 15 RMB from the payoffs in S1 and S2 of Game 4 (viewed from Player 1’s perspective, or S2 and S1 from Player 2’s perspective).

⁹ Due to COVID-19 policy in China, the additional experiment was conducted in 2023, a time when Renmin University of China had stopped organizing training classes for tax administrators. Consequently, we were unable to recruit tax administrators as our subjects. However, as will be shown later, student subjects in the additional experiment exhibited a very minimal degree of inequality aversion. This allows us to reject the inequality aversion explanation for the behaviour differences we observed between students and tax administrators in Game 4.

¹⁰ A reviewer noted a potential source of confusion in Experiment 1 related to a comprehension question where we did not display the game pie from the other player’s perspective, unlike what we did in the main instructions; we fixed this potential source of confusion in Experiment 2 and found it did not seem to matter, as results are similar between the two experiments (see footnote 10).

The same approach was applied to S3 to generate the second payoff distribution in each allocation task. In tasks 1 and 3, the second payoff distribution corresponds to the payoffs in S3 of Game 4 increased by 15 RMB. In tasks 2 and 4, it corresponds to the payoffs in S3 of Game 4 decreased by 15 RMB.

The order of the four allocation tasks was randomized. To minimize the potential spillover effect from the allocation tasks to the coordination games, a Cognitive Reflection Test (CRT) and two risk elicitation tasks were added between them. Further details regarding the test and the tasks can be found in the instructions in Appendix C. Moreover, receivers were not informed the payoff distributions selected by their corresponding allocators before the end of the experiment. All subjects were informed that their co-players in the coordination game were different from the subject with whom they were paired in the allocation tasks. The allocation tasks, CRT test, risk elicitation tasks, and corresponding games were all incentivized¹¹. The average payment per subject was around 129 RMB.

Table 5 reports the proportion of each payoff distribution chosen by the allocators, broken down by allocation task. In each task, only a small proportion of allocators (averaging around 11%) chose the payoff distribution that was equal but less efficient. Overall, 72.2% of allocators never chose the equal payoff distribution across the four allocation tasks, 16.7% of allocators chose the equal distribution once, 5.6% chose it twice, another 5.6% chose it three times, and no allocator chose it in all four tasks. Student subjects did not exhibit a high degree of inequality aversion that outweighs efficiency. This result does not support the explanation suggesting that students could have a much stronger degree of inequality aversion than tax administrators.

In contrast, in coordination Game 4, over one-third (38.9%) of these allocators chose the corresponding slice, S3. The frequency with which allocators chose the equal payoff distribution in the allocation tasks does not show a positive correlation with their S3 choices in Game 4 (correlation coefficient -0.17). The presence of payoff equality in Game 4 appears to foster the use of team reasoning among students, resulting in a larger proportion of S3 choices. The observed behavior difference, coupled with the

¹¹ At the end of the experiment, one task/game each from the four allocation tasks, two risk elicitation tasks, and four coordination games was randomly selected. A subject's final payment equals the sum their earnings from these randomly picked tasks in addition to their earnings from the CRT test, where each correct answer was valued at 1 RMB. There was no participation fee, as the allocation task guaranteed a positive payoff greater than 65RMB.

absence of correlation between allocation tasks and Game 4 offers support to the explanation that payoff equality facilitates team reasoning more effectively among students than tax administrators. More details regarding the proportion of choices for each slice in the coordination games can be found in Appendix D¹².

Table 5: Proportion of each payoff distribution’s choices by allocation tasks (allocators only)

Allocation task	Payoff distribution 1	Payoff distribution 2
Task 1	0.94	0.06
Task 2	0.83	0.17
Task 3	0.89	0.11
Task 4	0.89	0.11

Result 6: *Inequality aversion cannot explain why a greater proportion of student went for the payoff equal outcome in Game 4 than tax administrators.*

5. Discussion and Conclusion

In Experiment 1, 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 in China. The exception is Game 4, where the team-optimal equilibrium gives more equal payoffs than the others. In this game, a significantly higher proportion of students chose the team-optimal slice compared to tax administrators. Our results from Experiment 2 rule out inequality aversion as an explanation of the Game 4 behavioral difference. They suggest that payoff equality facilitates team reasoning more effectively among students than tax administrators.

With this qualification, our ‘lower bound’ approach 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

¹² There is no statistically significant difference in student subjects’ behavior in coordination games across the two experiments. Our findings from the first experiment remain consistent when pooling together students’ choices from both experiments.

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. Another interesting direction for future research would be to use a game frame that reflected the tax administrators' work environment, rather than an abstract frame. A further direction of future research could be to verify how the degree that each player benefits from a change in reference point impacts on team reasoning.

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, and this cannot be ascribed to a differential level of inequality aversion between our two samples.

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.

- Bacharach, M., & Bernasconi, M. (1997). The variable frame theory of focal points: An experimental study. *Games and economic Behavior*, *19*(1), 1-45.
- 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.
- Exadaktylos, F., Espín, A. M., & Branas-Garza, P. (2013). Experimental subjects are not different. *Scientific reports*, *3*(1), 1213.
- 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.

- Falk, A., & Zehnder, C. (2013). A city-wide experiment on trust discrimination. *Journal of Public Economics*, 100, 15-27.
- 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.
- Hu, Y., Kagel, J., Yang, H., & Zhang, L. (2020). The effects of pre-play communication in a coordination game with incomplete information. *Journal of Economic Behavior & Organization*, 176, 403-415.
- 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.
- Li, T. (2007). Are there timing effects in coordination game experiments? *Economics Bulletin*, 3(13), 1-9.
- 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.

- Mizrahi, D., Laufer, I., Zuckerman, I., & Zhang, T. (2018). The effect of culture and social orientation on Player's performances in tacit coordination games. In *Brain Informatics: International Conference, BI 2018, Arlington, TX, USA, December 7–9, 2018, Proceedings 11*, 437-447.
- Nagel, B. R. (1995). Unraveling in guessing games : An experimental study. *American Economic Review*, 85(5), 1313–1326.
- Peryman, J., & Kelsey, D. (2021). Ambiguity when playing coordination games across cultures. *Theory and Decision*, 90(3-4), 485-505.
- 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.
- Snowberg, E., & Yariv, L. (2021). Testing the waters: Behavior across participant pools. *American Economic Review*, 111(2), 687-719.
- 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.
- Stoddard, O., & Leibbrandt, A. (2014). An experimental study on the relevance and scope of nationality as a coordination device. *Economic Inquiry*, 52(4), 1392-1407.
- 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.]

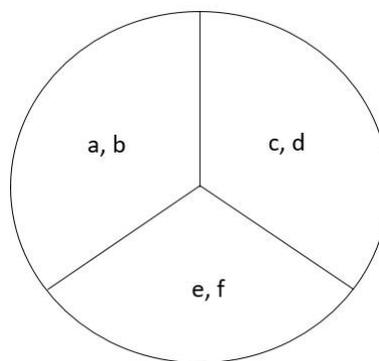
Welcome and thank you for taking part in this experiment. Everyone in the room has exactly the same instructions.

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.

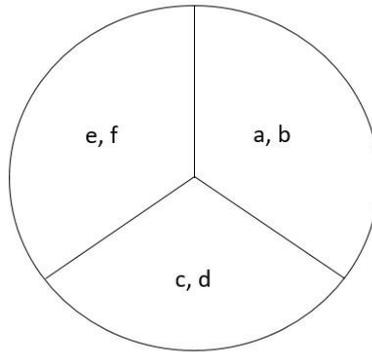


In the pie task, you and the other person will be presented with a three-slice pie, and asked to choose one slice.

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 different 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.

Your earnings

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.

Questions 1 of 4

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.

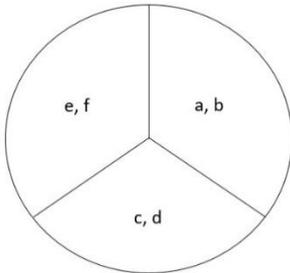
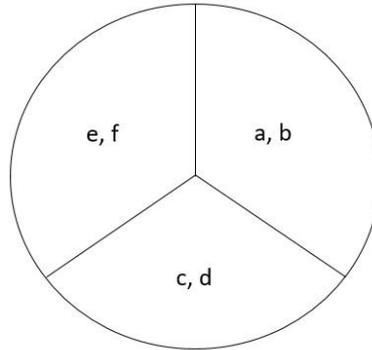
Questions 2 of 4

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?

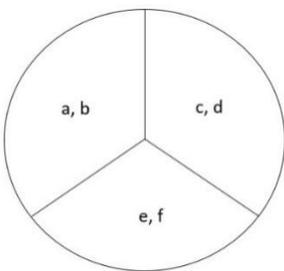
- A: You earn c and the other person earns d.
- B: You earn d and the other person earns c.
- C: You earn a and the other person earns b.
- D: You and the other person both earn nothing.

Questions 3 of 4

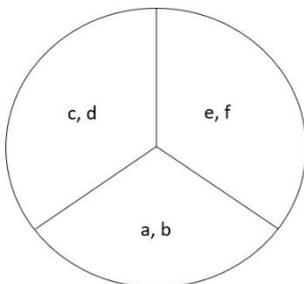
You see a pie task shown below in a pie task. What will the other person's pie task look like? ¹³



A:



B:



C:

¹³ We do not display the pie from the other player's perspective because this question is focused solely on checking the understanding of random rotation. Adding another layer of complexity such the mirrored payoffs in this question might overload our subjects. However, we recognize that this way of presentation may potentially confuse subjects. To find out if this is the case, in Experiment 2, we display the pie from the other player's perspective in this question. This modification, however, does not influence our findings.

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*** (0.04)	0.022 (0.81)	0.051 (0.08)
Game 2	0.028 (0.03)	-0.041 (0.05)	0.038 (0.06)
Game 3	0.052 (0.04)	0.023 (0.09)	0.082 (0.09)
Game 4	-0.031 (0.04)	0.208** (0.10)	0.042 (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 (0.04)	-0.056 (0.08)	0.010 (0.10)
Game 2	0.022 (0.05)	0.003 (0.08)	0.004 (0.08)
Game 3	0.084 (0.05)	0.003 (0.08)	0.037 (0.11)
Game 4	-0.042 (0.05)	-0.013 (0.09)	0.010 (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%

Appendix C: Experimental instructions for the additional experiment (translated from Chinese)

Experimental instructions

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 experiment is divided in four parts. Your earnings of the experiment will be the sum of the earnings from all the four parts.

Part 1

At the beginning of this part, you will be matched with another person in the room. You and the other person will not be told each other's identity. Each pair of participants is made up of a *Participant 1* and a *Participant 2*. In a corner of the computer screen, you can see whether you have been randomly selected by the computer to be a Participant 1 or a Participant 2.

The Allocation task

The task of Participant 1 is to make decisions in four allocation tasks. In each allocation task, each Participant 1 chooses between two allocations of earnings between himself or herself and the Participant 2 he or she is paired with. Participant 2 waits for the decisions made by the Participants 1 he or she is paired with.

Note that in the rest of the experiment, each Participant 1 will never interact again with the Participant 2 he or she is paired with in this part of the experiment. Furthermore, the decisions made in this part of the experiment will never be communicated to any other participant.

Your earnings

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.

Part 2

You will earn **1 extra RMB for each correct answer** to the questions below:

1. A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?
4. If three elves can wrap three toys in an hour, how many elves are needed to wrap six toys in 2 hours?
5. Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are there in the class?
6. In an athletics team, tall members are three times more likely to win a medal than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes?

Part 3

This part of the experiment has two tasks. At the end of the experiment, the computer will randomly choose one of the two tasks, and your earnings in that randomly chosen task will be realised.

Task 1

In this task, you are given **RMB 10**.

You are asked to decide how much, **between 0 and 10**, you wish to invest in a lottery.

For any amount invested in the lottery, there is a **50% chance to lose the amount you invested**, and a **50% chance to earn your initial investment plus two times the amount you invested**. This will be determined at random by the computer.

How much would you like to invest in this lottery (if anything)?

Task 2

In this task, you are given **RMB 10**.

You are asked to decide how much, **between 0 and 10**, you wish to invest in a lottery (if any).

For any amount invested in the lottery, there is a **50% chance to lose the amount you invested**, and a **50% chance to earn your initial investment plus half of the amount you invested**. This will be determined at random by the computer.

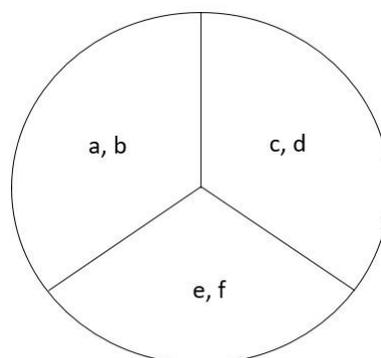
How much would you like to invest in this lottery (if anything)?

Part 4

The Pie task

At the beginning of this part, you will be matched with another person in the room. You and the other person will not be told each other's identity, but it will be someone different from whoever you have been matched with in Part 1. 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.



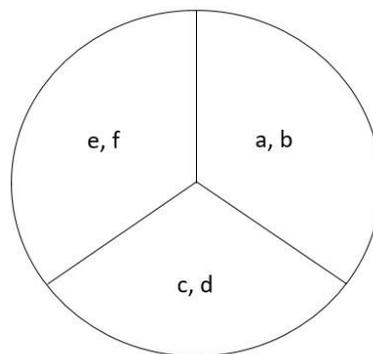
In the pie task, you and the other person will be presented with a three-slice pie, and asked to choose one slice.

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 different 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.

Your earnings

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.

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.

Questions 1 of 4

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.

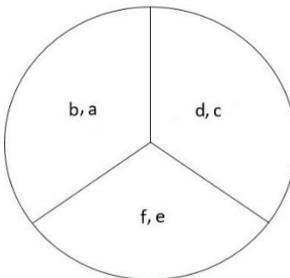
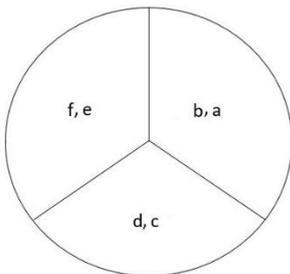
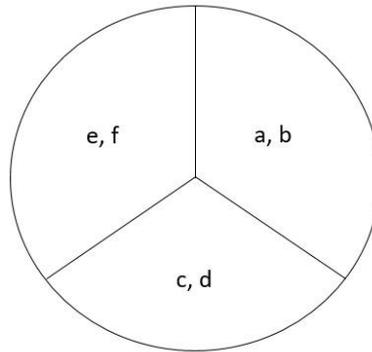
Questions 2 of 4

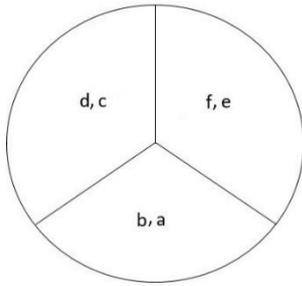
In the Pie task shown above, if you chose slice CD, and the other person chose slice AB, how much will you and the other person earn in that task?

- 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?





C:

D: All three above are possible.

Questions 4 of 4

Your earnings of the experiment will be:

A: The sum of the earnings from all the four parts plus 50 RMB participation fee.

B: The earnings from Part 1.

C: The earnings from Part 2.

D: The earnings from Part 3.

E: The earnings from Part 4.

Correct answers:

A D D A

Appendix D:

Table D1: Proportion of each slice's choices by game in the additional experiment

	S1 (LK)	S2 (LK)	S3 (TR)
Game 1	0.47	0.36	0.17
Game 2	0.53	0.44	0.03
Game 3	0.44	0.39	0.17
Game 4	0.25	0.39	0.36
Average	0.42	0.40	0.18