Experimental Economics
https://doi.org/10.1007/s10683-019-09632-z

ORIGINAL PAPER

Concentration and variability of forecasts in artificial investment games: an online experiment on WeChat

Xiu Chen[1] · Fuhai Hong[2] · Xiaojian Zhao[3,4]

Received: 6 November 2017 / Revised: 26 September 2019 / Accepted: 1 November 2019

Abstract

This paper is the first to use the WeChat platform, one of the largest social networks, to conduct an online experiment of artificial investment games. We investigate how people's forecasts about the financial market and investment decisions are shaped by whether they can observe others' forecasts and whether they engage in public or private investment decisions. We find that with forecast sharing, subjects' forecasts converge but in different directions across groups; consequently, forecast sharing does not lead to better forecasts nor more individually rational investment decisions. Whether or not subjects engage in public investment decisions does not significantly affect forecasts or investment.

Keywords Forecast · Investment · Online experiment · WeChat

JEL Classification C90 · D83 · D84 · G11

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10683-019-09632-z) contains supplementary material, which is available to authorized users.

✉ Xiu Chen
cxiu506@gmail.com

Fuhai Hong
fuhaihong@ln.edu.hk

Xiaojian Zhao
xjzhao81@gmail.com

1 Southern University of Science and Technology, Shenzhen, China

2 Department of Economics, Lingnan University, Tuen Mun, Hong Kong

3 Monash University, Clayton, Australia

4 Chinese University of Hong Kong (Shenzhen), Shenzhen, China

Published online: 25 November 2019

1 Introduction

Information and communication technologies have drastically improved information accessibility and transparency at the grassroot level. With these technologies, people are free to communicate and share information without barriers. For many real-world economic or political decisions such as shopping and voting, people typically share information and opinions before taking actions. While the actions taken are quite often unobservable, the advancement of technologies has greatly facilitated the pre-play communications. Take the financial market for example. An individual's investment decisions are usually private and hard to observe, but investors' attitudes are often exchanged through online or offline gossips before actual moves are made. Financial economists (e.g. Antweiler and Frank 2004; Chen et al. 2014) have started to use textual analysis to study the emerging role of social media in financial markets, and found that shared opinions as soft information can predict performance in financial markets. In this paper, we report findings from an online experiment where subjects made unobservable investment decisions but were allowed to share their forecasts on the financial market.

Our experiment investigates whether communication causes convergence in individuals' forecasts about the financial market and how this influences individuals' subsequent investment decisions. Economists have observed the phenomena of groupthink and group polarization, suggesting that individuals within a group tend to suppress divergent viewpoints and reach a (possibly more extreme) consensus (e.g. Bénabou 2013; see also the literature review in Sect. 2). Convergence of behavior or beliefs may have important implications for the financial market. While herd behavior, i.e. investors' imitation of others' actions, has been found to exacerbate market volatility and fragilize financial systems (Bikhchandani and Sharma 2000), some observers also express concerns that social contagion of beliefs might result in sizable valuation errors in financial markets (Shiller 1992) or collective delusions leading to investment frenzies or crashes (Bénabou 2013). In particular, with the slowdown of economic growth and wobbly confidence of investors in the context of Chinese economy, we are especially interested in whether communication on social media will lead to more convergent beliefs about the market and aim to provide experimental evidence on this question.

If with forecast sharing, a group indeed forms more converged forecasts, the next questions to ask are where the forecasts converge to and how the converged forecasts influence investment. A natural conjecture is that communication helps people generate collective wisdom and improve their investment performance. However, the results in the literature regarding whether group discussion generates collective wisdom are mixed. Hogg et al. (1990) demonstrate the heterogeneity of group polarization, in the sense that the direction of post-discussion individual choices shifts largely, depending on the norms of the particular group. If communication does not generate collective wisdom in forecasts, then it does not necessarily result in better investment performance. Moreover, liquidity in financial markets has a public good feature; people may thus strategically express overoptimistic opinions in communications in order to encourage market participation from others so as to increase liquidity. In this regard, a costless pre-play communication could result in over-optimism in forecasts about the financial market, leading to more investments.

In a nutshell, on an experimental basis, we study the social contagion of forecasts in artificial investment games and aim to address the following questions:

- Within a group, can pre-play communication produce convergent forecasts about the financial market? If so, where do the forecasts converge?
- How would individual investment decisions be influenced by the within-group forecast sharing?

We conducted an online experiment using the platform of WeChat, the world's biggest standalone messaging application in terms of user numbers and China's most popular online social network. The experiment included a group-based artificial investment game with 1385 subjects who were WeChat users. The payoff of the game relies on participants' investment decisions and a return multiplier determined by the opening price of the Shanghai Composite Index (henceforth SCI) on Monday, November 30, 2015. The experiment started at 7:00 PM on Friday, November 27, 2015, after the stock market was closed, and ended at 9:00 AM on the next Monday, November 30, 2015, before the stock market was opened. In the game, we asked each subject to forecast the opening price of the SCI and to decide whether to invest.

We adopted a $2 \times 2 \times 2$ between-subject design. First, to examine the effect of pre-play communication, in the *forecast sharing* groups, the subjects can observe other players' forecasts of SCI, whereas in the *baseline* groups, the subjects cannot. Second, we introduced two types of investment projects in our experiment, with a subject's payoff dependent on other group members' investments in one of them (i.e., the *interdependent project*), and independent of other group members' investments in the other (i.e., the *independent project*), to detect the potential over-(under-) reporting of forecasts driven by the strategic motives to induce other players (not) to invest as discussed above. Lastly, we design two worlds with different mappings between the SCI and the investment project's return multiplier for a robustness check.

We opted to take the lead in performing an online experiment on WeChat instead of using standard laboratories mainly because of two reasons: (1) a more representative sample can be obtained for our study. In contrast to standard laboratory experiments, in which most subjects are recruited from university students who typically have little experience in financial investment,[1] subjects recruited from WeChat users are more representative of investors in the Chinese financial market, given that WeChat is the most popular messaging application in China and that the financial market in China is dominated by retail investors.[2] (2) While all the subjects were

[1] It was shown that student subjects perform differently from subjects in the field in an experimental study of herd behavior in financial markets (Alevy et al. 2007).

[2] By 2017, WeChat had more than 963 million monthly active users; see the Wikipedia entry of WeChat. Retail investors conducted 85 percent of the stock market's transactions in China in 2015; see also the article "New Horizon Opening For China's Stock Market" by Thomson Reuters, available at http://share.thomsonreuters.com/general/China/Special-Report-New-Horizons-Opening-For-China's-Stock-Market.pdf.

given the same latest price of the stock market when the experiment started, it was important to maintain all the other conditions on which subjects based their decisions (including the related on-going media news) the same for all subjects. Due to the capacity of a laboratory, a laboratory experiment usually needs to be conducted in several successive sessions, which would create a time difference in subjects' decision making and thus contaminate the conditions. Conducting an online experiment solves this problem by allowing a large number of subjects to make choices online within the same time period.

We observe a strong, positive correlation between the subjects' forecasts on the stock market and their investment decisions, suggesting that subjects' reported (cheap-talk) forecasts in the pre-play communication are associated with their underlying beliefs, which drive the subsequent investment behavior. More importantly, our experimental results provide evidence of both concentration of within-group forecasts and variability of the concentration across groups: compared to baseline groups, subjects in forecast sharing groups tend to concentrate on some forecasts, but the concentrated forecast varies significantly across groups. These results suggest that while pre-play communication does lead to more converged beliefs, the converged beliefs do not necessarily converge to the true state. Consequently, forecast sharing does not help subjects make more individually rational investment decisions. We also find that, in our experiment, investment spillovers do not significantly influence forecasts or investment decisions.

The rest of this paper is organized as follows. The next section reviews the related literature. Section 3 describes our experimental design and procedure. Section 4 presents our experimental findings. The last section concludes.

2 Related literature

We study social learning by introducing pre-play communication to investment games. Studies on social learning have divergent opinions: on one hand, Mojzisch et al. (2010) empirically identify distinct biases in group discussions; on the other hand, Charness and Sutter (2012), in their review paper, suggest that groups are more cognitively sophisticated, and thus make better decisions than individuals. However, this line of literature is mainly based on investigations of group collective decisions. Our study looks into individual decisions with prior within-group communication.

In the literature of group polarization, following Brown (1986, p. 200), Glaeser and Sunstein's (2009) model suggests that decisions made after information sharing appear to be worse than independent decision making when individual learning is non-Bayesian. With a fully rational model, Sobel (2014) shows that, in the absence of any restrictions on the information structures, there is no systematic relationship between individual and group decisions under information aggregation. Roux and Sobel (2015) find that group decisions can be more extreme than individual decisions in monotone decision problems (i.e. problems where the action rule is increasing with respect to signals). Related to group polarization, Bénabou (2013) and Bénabou (2015) introduce the notion of *groupthink* and account for the formation of individually rational collective denial and willful blindness.[3] Our research questions are also related to herd behavior and peer effects. After Banerjee (1992) theoretically rationalizes herd behavior, empirical works confirm the existence of herd behavior and peer effects in both the laboratory (Cipriani and Guarino 2005) and the field (e.g. Grinblatt et al. 1995; Bursztyn et al. 2014). Among the online studies, Salganik et al.'s (2006) investigation on social influence in an artificial cultural market is the most closely related to our paper.[4] While the notion of herd behavior in the literature is typically based on observable actions, in reality it is often the case that actions are not observable while beliefs are shared, as discussed in the Introduction. Our experiment captures this feature by studying pre-play communication in investment games where investment is unobservable. We aim to explore the cause and effect of concentrated forecasts in pre-play communication.

Our experimental results provide evidence of group polarization and show that group polarization is indeed variable, giving rise to less reliable decisions. By utilizing WeChat, the most popular social network application in China, we explore whether within-group forecast sharing improves collective wisdom. The potential of online social networks to facilitate information diffusion and civic engagement in China has been documented in Zheng and Wu (2005). However, McGrath et al. (2012) question whether citizen engagement in the political process in such a speedy manner results in well-thought-out choices or rapid promises that could generate constant societal frustration. Our experimental study suggests that while communication leads to concentration of beliefs within a group, it does not consistently lead to a convergence on the correct belief, nor does it improve the individual investment performance.

Our experiment also includes spillover of investment in an artificial investment game, which smacks of public good games. A wide range of theoretical and experimental studies have found that costless pre-play communication is effective in facilitating cooperation. Dawes et al. (1977) and Isaac and Walker (1988) are among the first to provide experimental evidence of communication improving cooperation. Bochet et al. (2006) extend the literature by comparing the efficacy of communication with that of punishment and find that communication is more efficient for sustaining cooperation and increasing public goods provisions. Agastya et al. (2007) also find that a prior stage of non-binding communication improves the efficiency of the equilibria in a voluntary contribution game. Chaudhuri and Paichayontvijit (2006) investigate different communication schemes and find that the communication of common knowledge is the best scheme to improve cooperation. Relatedly, this paper explores the effect of communication in public goods provision with the presence of uncertainty. We empirically investigate how pre-play communication

[3] More recently, Bénabou and Tirole (2016) provide an overview on the topic of motivated beliefs and reasoning at both the individual and collective levels.

[4] Related studies include Muchnik et al. (2013), who perform an experiment on a social news aggregation website and find that prior ratings create herding effects and significant bias in individual rating behavior.

affects the confidence on the return of public goods investment and thus the investment behavior.

2.1 Online experiments

Until recently, online experiments in economics and finance have been scarce compared to the abundance of laboratory and field experiments. Chen and Konstan (2015) survey online field experiments and analyze the technologies and design choices for online economic or computer science studies. Field experiments on online platforms include Kramer et al. (2014) on Facebook, and Horton et al. (2011) and Berinsky et al. (2012) on Amazon Mechanical Turk (MTurk), which use "modified web interfaces," as discussed in Chen and Konstan (2015). While our study also involves modified web interfaces, the way we conduct our experiment is closer to that of studies using online resources to build an "online laboratory" (Hergueux and Jacquemet 2014; Salganik et al. 2006): Online field experiments use a "modified web interface" to "evaluate the effects of changes in user interface design" (Chen and Konstan 2015), while an "online laboratory" aims to "reach more diverse samples" and conduct laboratory experiments "directly over the Internet" (Hergueux and Jacquemet 2014). We build our own temporary laboratory on WeChat, which we use as a platform to recruit subjects, run experiments and, and transfer payments.

Horton et al. (2011) and Berinsky et al. (2012) find that online experiments can reach the same internal validity and external validity as laboratory experiments. For internal validity, they test problems including inattentiveness and identification, and conclude that these problems are insignificant and can be mitigated. For external validity, they find online subjects more representative than offline subjects. They replicate published experimental works using MTurk, which yields consistent results compared to offline experimental frameworks.

Besides the reasons for conducting an online experiment to investigate our research questions as discussed in Sect. 1, there are additional advantages of online experiments. It usually costs less to conduct an online experiment than a laboratory experiment owing to the reduced costs of laboratory maintenance and payments to subjects (as subjects do not have to physically show up). Table A.1 in Online Appendix A reviews payments of the studies that performed online experiments on MTurk, which shows that the average *hourly* payment varied but was typically not higher than US$15 while the duration of an experiment was often below 10 minutes. Therefore, the average cost per subject is low. Online Appendix B.1 further discusses how we take advantage of certain features of WeChat to avoid or mitigate some potential problems of online experiments, while Online Appendix B.2 discusses some implementation issues on conducting online experiments with WeChat.

Table 1 Options for the Forecasting Task

Option	Set I	Set II
A	3620.01 or above	3560.01 or above
B	3560.01 to 3620	3480.01 to 3560
C	3480.01 to 3560	3380.01 to 3480
D	3380.01 to 3480	3300.01 to 3380
E	3300.01 to 3380	3240.01 to 3300
F	3300 or below	3240 or below

3 Experimental design and procedure

We design artificial investment games to investigate whether forecasts of the opening price of the Shanghai Composite Index (SCI) on a future date converge when the forecasts are shared, and how the forecast sharing influences investment behavior. The payoff of a subject in the investment game depends on, among other factors, (1) the realized opening price of the SCI on that date and (2) the subjects' investment decisions after making forecasts.

3.1 Experimental design

In the investment game, a subject first chooses a range that she thinks to be the most likely for the opening price of the SCI on November 30, 2015 (the Forecasting Task). Subjects are told the SCI closing price of the date when the experiment started (November 27, 2015), which was 3436.40. We provide subjects with several price ranges to choose. Table 1 shows the two sets of options provided in the different treatments (as described later in this section):

The subject then decides whether to invest 250 Renminbi Yuan (RMB), out of a 500 RMB endowment, in an artificial project, which will be described in detail below (the Investment Task).

Subjects are divided into groups, and groups are further divided into treatments. Our $2 \times 2 \times 2$ treatments vary in three dimensions.

The first is whether forecasts are shared among group members.

- Baseline: There is no information provided regarding other group members' forecasts;
- Forecast Sharing: The subjects can observe the distribution of the forecasts of other group members.

Specifically, in treatments with *forecast sharing*, a bar is placed on each option to indicate the percentage of group members who have selected that option (see the left bottom figure in Online Appendix C, for which English translation is provided in

Online Appendix D).[5] Subjects in all the treatments are allowed to return at any time before the deadline to change their forecasts and investment decisions. This means that under the treatments with *forecast sharing*, subjects can return to view the latest distribution of forecasts.

While in general incentivizing subjects' forecast decisions is helpful to induce subjects to truthfully report their beliefs, we intentionally do not incentivize the Forecasting Task for the following reasons. First, this feature allows us to study subjects' behavior in a cheap-talk pre-play communication in the Forecast Sharing treatments. In particular, we will study whether subjects would strategically over- or under-report their beliefs to influence others' investment decisions in the presence of investment spillovers (as explained in the second treatment dimension below). Second, incentivizing forecasts in our game may cause a hedge problem: The correlation between earnings from the investment and earnings from beliefs, if incentivized, may motivate risk averse subjects to use stated beliefs as a hedge against adverse outcomes of decisions in the investment game and thus bias the elicited belief (Blanco et al. 2010). Third, even if in theory an incentivized truth-telling belief elicitation mechanism exists for our context, it may be overly complicated in practice given subjects' limited attention to an online experiment. Finally, Gächter and Renner (2010) show that non-incentivized beliefs are highly associated with incentivized beliefs.[6] In our analysis, we will also look into the correlation between the forecasts and the incentivized investment decisions.

In the second dimension, we manipulate the payoff function of the investment game by introducing spillover within groups. Investment provides liquidity to the market, which has a public good like feature. In one set of treatments, we follow Bénabou (2013) and include a spillover of other group members' investments in the payoff function, so as to reflect the real-world investment situation that each individual is embedded in a collective interaction where his final payoff is determined by both his own action and the actions of others. The spillover may give the subject an incentive to over-/under-report her forecast when costless pre-play communication is allowed in order to induce her group members to invest/not to invest. To identify the potential over-/under-reporting of forecasts driven by such a strategic motive, in another set of treatments, a subject's payoff depends solely on the individual's investment decision and the SCI. Therefore, there are two types of investment projects:

- Interdependent (or public) projects:

  $$PAYOFF = 500 - INVEST + m \times (INVEST + avg.GINVEST)$$

[5] This form of communication, although restricted, is easier to measure than open communication, and is richer than a binary signal, which is less interesting for our study of concentration and variability of forecasts.

[6] Examples of studies that investigate people's forecasts on the financial market without providing incentives include Oechssler et al. (2011).

Table 2 SCI and corresponding return multiplier m

World I		World II	
SCI	m (%)	SCI	m (%)
3620.01 or above	160	3560.01 or above	160
3560.01 to 3620	120	3480.01 to 3560	120
3480.01 to 3560	80	3380.01 to 3480	80
3380.01 to 3480	40	3300.01 to 3380	40
3300.01 to 3380	0	3240.01 to 3300	0
3300 or below	– 40	3240 or below	– 40

Table 3 Group allocation

	Baseline	Forecast sharing
World I		
Interdependent Project	1	6
Independent Project	1	6
World II		
Interdependent Project	1	6
Independent Project	1	6

where $INVEST = 250$ if the subject chooses to invest and 0 otherwise. The parameter m, which we call the *return multiplier*, is determined by the SCI opening price, as explained in detail below. The term $avg.GINVEST$ is the average investment amount of the group. Therefore, in the interdependent project, there is a spillover of group members' investment, which is positive when $m > 0$.

- Independent (or private) projects:

$$PAYOFF = 500 - INVEST + m \times INVEST$$

In independent projects, the subject's payoff is independently determined by his own investment decision and the SCI opening price.

The last dimension is designed for a robustness check, in which we have two worlds, World I and World II, with different probability distributions of the return multiplier m. In both worlds, the return multipliers are determined by the SCI opening price. The two worlds differ in their mappings from the SCI opening price to m, as shown in Table 2. Note that the SCI closing price on the day when the experiment started, i.e., 3436.40, corresponds to $m = 40\%$ in World I and $m = 80\%$ in World II. We are then able to check the robustness of subjects' behavioral patterns under different economic conditions.

In independent projects, the dominant strategy for an individual is to invest if and only if the SCI opening price is above 3560 in World I or above 3480 in World II. In interdependent projects, while each player should invest if and only if the SCI opening price is above 3480 in World I or above 3380 in World II to achieve the first best outcome, the dominant strategy, i.e. the individually rational decision is still to invest if and only if the SCI opening price is above 3560 in World I or above 3480 in World II.

In total, we have 28 groups. Table 3 shows the number of groups in each of the 8 treatments.

It is meaningless to have more than one group in a baseline treatment: Being in the same group or in different groups should not affect subjects' decisions as they receive no information regarding their group members' choices. The allocation scheme in Table 3, with one group in each of the baseline treatments, allows us to have more subjects in each group. Subjects were randomly assigned to the 28 groups. For the purpose of our statistical analysis (explained in Sect. 4), twice as many subjects were assigned to each baseline group as those assigned to each forecast sharing group. This allocation results in approximately 43 subjects in each of the 24 forecast sharing groups and approximately 86 subjects in each of the 4 baseline groups.

We aim to test the following hypotheses with the experimental design:

Hypothesis 1 (a) The forecast sharing groups display a higher concentration of forecasts compared to the baseline groups. (b) The forecast sharing groups make more individually rational investment decisions compared to the baseline groups.

Hypothesis 1 aims to investigate the effect of forecast sharing. In the psychology literature, Janis (1982) defines "groupthink" as the tendency of consensus seeking. Bénabou (2013) and Bénabou (2015) incorporate the notion into the economics literature and show that groupthink may contribute to the formation of collective delusions in groups. Based on the literature, by Hypothesis 1a, we aim to test whether subjects tend to suppress divergent forecasts and reach a conformity with other group members in communication. Such a tendency would increase the concentration level of forecasts in the forecast sharing groups. Moreover, if within-group information sharing leads to collective wisdom, the forecast sharing groups would perform better than the baseline groups in investment, leading to Hypothesis 1b.

In Hypothesis 2, we look into the comparison between the interdependent project and the independent project. Notice that when $m > 0$, there is a positive investment spillover in the interdependent projects. If the positive spillover of the interdependent projects incentivizes subjects to over-report their forecasts to induce other group members to invest, then subjects in the forecast sharing groups with interdependent projects will report higher forecasts than those in independent projects and in the baseline groups.[7] Furthermore, the over-reported forecasts may induce more investment in the interdependent projects. This leads to the following hypothesis:

[7] When $m < 0$, the spillover in the interdependent projects is negative, where subjects have incentives to under-report their forecasts to induce other group members not to invest. However, the opening price of SCI on November 30, 2015 has to fall below 3300 for World I or 3240 for World II for $m < 0$ to be satisfied, which is very unlikely given the closing price of 3436.4 on November 27, 2015. Even if a subject believes that the opening price would fall into this low range, he could not under-report his forecast given our experimental design, as this is already the choice with the lowest range that we provide (see Table 1).

Hypothesis 2 In the interdependent projects, the forecast sharing groups report higher forecasts and invest more than the baseline groups.

With the different return multipliers between the two worlds, investment return in World II is higher than in World I under a given realization of the SCI, which may lead to more investments in World II. We compare between the two worlds in Hypothesis 3:

Hypothesis 3 Compared to World I, subjects in World II invest more.

3.2 Experimental procedure

The experiment started at 7:00 PM on Friday, November 27, 2015 and ended at 9:00 AM on the next Monday, November 30, 2015.[8] We recruited subjects by circulating a link of the experiment webpage to the WeChat user population via several channels. The main channel by which we advertised our experiment was to post the advertisement on popular official accounts.[9] With the help of Shenzhen Quantum Net Technology Co., LTD,[10] over 10 million subscribers of the official accounts it operated received the advertisement along with the link to the experiment webpage. All WeChat users receiving the link to the experiment could easily circulate it within their own social networks on WeChat.

Online Appendix C presents the cellphone screenshots of our experiment webpage for the treatment of Interdependent Project-Forecast Sharing-World II (Inter-FS-II). The instructions were written in Chinese as the targeted population was WeChat users who were mostly Chinese. The English translation of the experimental instructions can be found in Online Appendix D.

A brief description of the investment game was presented on the first page, in which subjects were informed that they were randomly assigned to different groups. On the second page, subjects were informed that 6 of them would be randomly chosen after the experiment was closed and paid the amount of money which is equal to their final payoff in the game.[11] The ex post average payment was about 3 RMB

[8] It is worth noting that the SCI dropped 5.48% on November 27, 2015. But there was no news or policy involving a significant impact on the subjects' beliefs released during the experimental period.

[9] Official accounts, as opposed to private accounts, are interfaces an operator uses to gather subscribers, circulate notifications or redirect readers to a website.

[10] This is a company that provides operational services for WeChat official accounts. It operates up to a thousand official accounts on nearly all kinds of topics including fashion, lifestyle, tourism, sports and entertainment. As of 2016, it had hundreds of millions of subscribers.

[11] We adopt the approach of randomly paying only a subset of participants in order to lower the transaction costs due to the large number of participants. Recent examples of adopting the same approach include Exadaktylos et al. (2013), Attema et al. (2016) and Ehm et al. (2018). Charness et al. (2016) review related studies and conclude that the loss of motivation generated by paying only a subset of participants instead of paying all of them is small. Moreover, while the probability of being chosen is low (6 out of 1385), the highest possible payment is high (over 1000 RMB). Angner (2016, p175) points out that using lotteries is an effective way to incentivize subjects, because according to the prospect theory, people tend to overweigh a small probability of winning a large award.

per subject (equivalent to US$0.48) with the average experimental duration being 5–10 min, meaning an hourly wage of approximately US$3-6. Compared to the payment level of the online experiments conducted on MTurk (reviewed in Table A.1 in Online Appendix A) and considering the differences in minimum wages across countries, the payment level in our experiment suffices to provide incentives for the subjects.[12]

The subjects were then provided with the detailed instructions of the game. Two example payoff calculations were given in the link below the instructions. At the top of the third page, we presented a figure displaying the last 5 days of the SCI. Below that, we provided an online calculator to help subjects calculate the payoff. The calculator allowed subjects to adjust the level of the SCI, to change the investment decision, and, for subjects in the interdependent project, to vary the percentage of investors within the group, so as to see how the payoff was determined accordingly. Then, we asked the subjects two major questions: Question 1 on the forecast of the SCI opening price (i.e., the Forecasting Task) and Question 2 on whether to invest (i.e., the Investment Task), both of which were compulsory.

After the subjects completed the two tasks, they were asked to complete a demographic questionnaire on the fourth page. At the end of the experiment, on the fifth page, we reminded the subjects that they could return at anytime before the deadline, i.e., 9:00 AM on November 30, 2015, to view and change their selections.

The SCI, which determined the return multiplier m, was realized 20 minutes after the experiment was closed. Then, 6 participants were randomly selected. We calculated their payoffs according to m and their investment decisions, used their recorded WeChat User IDs to contact them and paid them via WeChat transfer. Online Appendix C.2 provides several original screenshots of online payments and Online Appendix D.3 provides English translations.

A total of 1569 WeChat-user subjects participated in the experiment, of whom 1385 completed the game. Table 4 summarizes the demographic statistics. The composition of our subjects is in general balanced in gender and is more diverse than that of a laboratory experiment in terms of geographic regions, education levels and annual household income. Table 4 also shows the demographic distribution of subjects broken down by each of the three treatment dimensions.

4 Experimental results

According to the opening price of SCI on November 30, 2015, i.e., 3433.86, the correct forecasts in our experimental setting are Option D in World I and Option C in World II, which correspond to the return multipliers $m = 40\%$ in World I and $m = 80\%$ in World II. Therefore, the individually rational decision in both worlds and under both types of projects is to not invest.

[12] The minimum hourly wages in 2015 were US$7.25 in the U.S. (federal nationwide), US$0.31 in India and ranged from US$1.2 to US$2.8 in China (varying across regions). See a list of minimum wages by country on Wikipedia, available at https://en.wikipedia.org/wiki/List_of_minimum_wages_by_country.

Table 4 Demographical summary statistics

	Mean						
	All	Dimension I		Dimension II		Dimension III	
		Baseline	Forecast sharing	Interdependent project	Independent project	World I	World II
Female	.48	.46	.49	.48	.48	.49	.47
Age	24.66	24.72	24.95	25.25	24.11	24.26	25.07
	(15.28)	(15.31)	(15.12)	(14.94)	(15.59)	(15.35)	(15.22)
Residence							
Eastern China	.55	.52	.56	.57	.53	.55	.55
Central China	.21	.22	.21	.20	.22	.21	.21
Western China	.14	.16	.13	.13	.14	.14	.13
Hong Kong, Macau, Taiwan or others	.10	.10	.10	.10	.11	.10	.11
Education							
Junior high school or below	.19	.23	.18	.19	.19	.18	.20
Senior high school	.25	.27	.25	.27	.24	.26	.25
Professional school	.13	.12	.14	.12	.15	.12	.15
Undergraduate	.19	.17	.20	.20	.18	.19	.19
Graduate or above	.23	.21	.23	.22	.24	.25	.21
Occupation							
Student	.14	.15	.14	.14	.14	.16	.12
Public officer	.05	.04	.05	.05	.05	.05	.05
Public institution employee	.11	.11	.11	.11	.11	.12	.11
State-owned enterprises employee	.10	.09	.10	.11	.10	.09	.11
Private sector employee	.18	.20	.18	.17	.20	.17	.20
Self-employer	.14	.15	.13	.14	.13	.14	.14
Unemployed	.10	.11	.10	.10	.11	.11	.10

Table 4 (continued)

	Mean						
	All	Dimension I		Dimension II		Dimension III	
		Baseline	Forecast sharing	Interdependent project	Independent project	World I	World II
Retired	.02	.02	.02	.02	.03	.03	.02
Other	.12	.12	.12	.13	.11	.12	.12
Annual household income							
30,000 or below	.28	.32	.27	.28	.29	.28	.28
30,000–80,000	.30	.31	.30	.32	.29	.29	.32
80,000–300,000	.26	.25	.27	.26	.26	.27	.26
300,000–500,000	.06	.06	.06	.06	.06	.06	.06
500,000–800,000	.02	.03	.02	.02	.03	.03	.02
800,000–2,000,000	.02	.02	.02	.02	.02	.02	.02
2,000,000 or above	.05	.01	.06	.04	.05	.01	.00
Forecast	2.74	2.77	2.73	2.79	2.69	2.95	2.53
	(1.28)	(1.34)	(1.26)	(1.26)	(1.31)	(1.27)	(1.26)
Investment	.77	.75	.78	.78	.76	.76	.78
	(.20)	(.19)	(.22)	(.20)	(.21)	(.21)	(.19)
Observation	1385	377	1008	665	720	699	686

(1) Standard deviation in parentheses

(2) The *t test* (for age) and *Chi-squared test* (for other demographic variables) show that demographic distributions are balanced from comparisons along each of the three dimensions

(3) The mean of forecasts is calculated based on the following transformations: A = 1, B = 2, … , F = 6

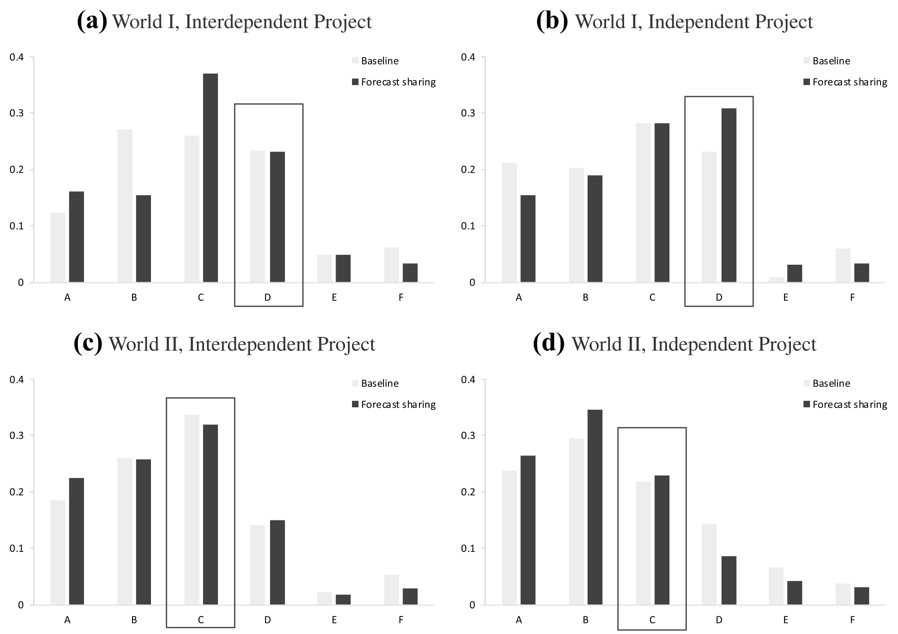

Fig. 1 (Average) Selection shares in the Forecasting Task. *Notes:* (1) Light bars are the selection shares of Option A to F for the baseline groups. Dark bars are the selection shares averaged over the six forecast sharing groups. (2) The bars highlighted with frames are the (average) selection shares of the correct forecast, i.e. Option D in World I and Option C in World II

We first define the subjects' selection shares of options in Question 1 (the Forecasting Task):

$$s_{ij} = \frac{c_{ij}}{\sum_{k=1}^{T} c_{kj}}$$

where c_{ij} is the number of times Option i is selected, or Option i's *selection count*, in group j, and T is the number of options. Figure 1 depicts the selection share of each option for the baseline groups and the average selection share of each option for the forecast sharing groups under each treatment, while Fig. 2 depicts the subjects' selection distribution for each group in the Forecasting Task. Figure 1 shows that many of the options have similar selection shares between the baseline treatments and the forecast sharing treatments, especially for the correct options, i.e., Option D in World I and Option C in World II (with the only exception in Fig. 1b). However, Fig. 2 shows that the baseline groups preserve a more even distribution for Option B, C and D in World I and Option A, B and C in World II than the forecast sharing groups. In addition, the share of the correct forecast is not necessarily larger in forecast sharing groups than in the corresponding baseline group. Figure 3 depicts subjects' choices in the Investment Task, showing the frequencies of those who chose not to invest, i.e. the individually rational decision, in various treatments.

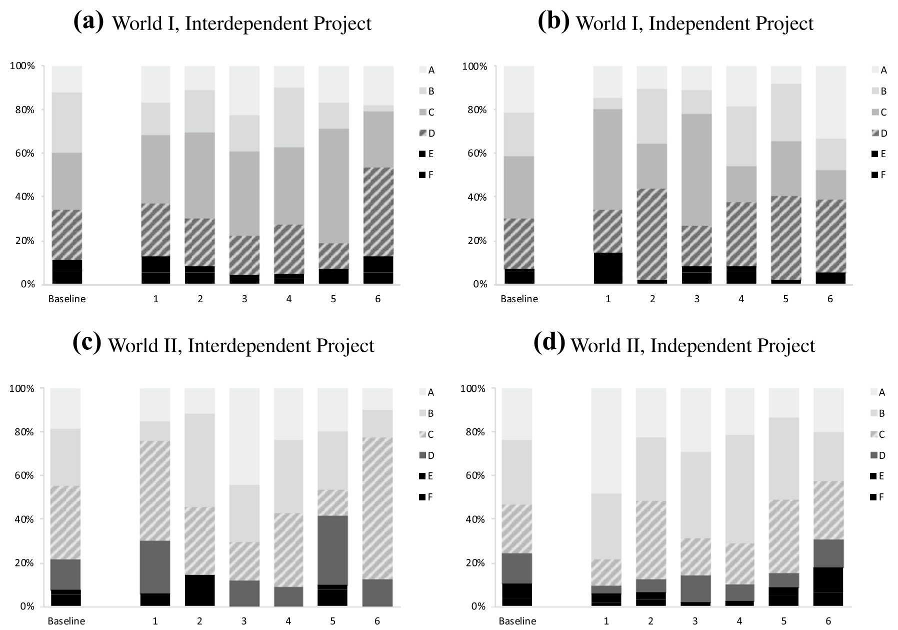

Fig. 2 Selection distributions in the Forecasting Task. *Notes:* (1) The numbers 1, 2, … 6 indicate the six forecast sharing groups respectively. (2) The shaded areas represent the selection share of the correct forecasts, i.e. Option D in World I and Option C in World II

4.1 Concentration and variability of forecasts

In the analysis below, we use the Herfindahl-Hirschman index (HHI) to examine the extent to which subjects' forecasts are concentrated within a group. HHI is commonly used to measure market concentration, and is also widely applied to study various social issues, such as measuring the variability of individuals' choice modes (see, e.g., Susilo and Axhausen 2014). In our study, we use HHI to measure the level of forecast concentration. HHI H_j for group j is defined as:

$$H_j = \sum_{i=1}^{T} s_{ij}^2$$

where s_{ij} is the selection share of Option i in group j as defined above. According to this definition, the higher the forecast concentration is for group j, the closer H_j is to 1. Consider an extreme case where all participants in group l select the same option k. Then option k has share $s_{kl} = 1$ while $s_{gl} = 0$ for all $g \neq k$, and thus H_l equals 1. If forecasts are more dispersed, H_j is lower. The lower bound for H_j depends on T, the number of options: For our forecasting question, H_j ranges from 0.17 to 1.

The subject allocation scheme (described in Sect. 3) allows us to adopt an approach that is closely related to Salganik et al. (2006, hereafter SDW) to

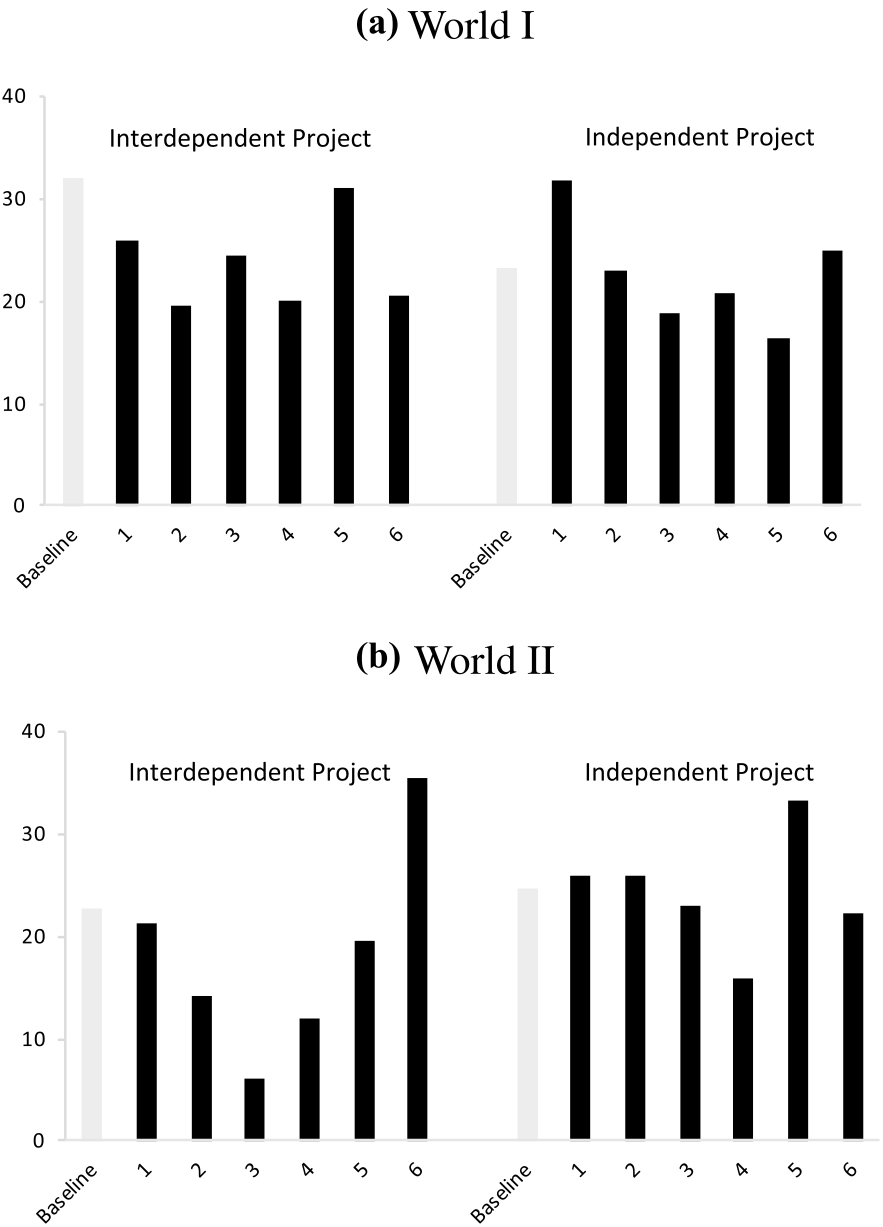

Fig. 3 Frequencies of "Not Invest". *Note:* The numbers 1, 2, … 6 indicate the six forecast sharing groups respectively

statistically compare the HHIs between the baseline and forecast sharing groups.[13] This statistical procedure falls in a broad class of nonparametric resampling methods and saves experimental costs by allowing us to collect a much smaller sample in the baseline treatments than in the forecast sharing treatments. The statistical procedure for each of the four comparisons between forecast sharing groups and the baseline group (for Independent Projects in World I or II, or Interdependent Projects in World I or II) is detailed as follows:

Step 1 We calculate the HHI for the 6 forecast sharing groups, resulting in 6 HHIs for the forecast sharing treatment;
Step 2 We randomly and evenly split the one group in the corresponding baseline treatment into two groups so that the size of the two groups is the same as that of the forecast sharing groups, and then calculate the HHI for one of the two groups. We repeat the step 6 times, resulting in 6 HHIs for the baseline treatment;
Step 3 We conduct the two-sided Mann–Whitney U test on the HHIs of the forecast sharing treatment and the baseline treatment (using the results in Step 1 and 2) and record the *p value*;
Step 4 We repeat Step 2 to 3 1000 times (as in SDW) and calculate the frequency of the cases where HHI is not significantly higher in the forecast sharing treatment than in the baseline treatment (where the significant difference is determined by Z-statistic < 0 and *p value* < 0.05).

Figure 4 reports the average HHI for each treatment using the above procedure, while Table 5 reports the frequencies from Step 4.

The statistical results suggest that the forecasts in the forecast sharing groups are significantly more concentrated within the group than those in the baseline groups, in both worlds and for both interdependent and independent investment projects. We thus have our first result:

Result 1a *Forecasts in the forecast sharing groups are significantly more concentrated than those in the baseline groups.*

Since forecasts are more concentrated with forecast sharing, a natural question is thus: Does forecast sharing help subjects make better forecasts and better individual investment decisions? Figure 5 presents the frequency of subjects' selection of the correct forecast for each group. This figure, together with Fig. 2, shows that the frequency of making correct forecasts fluctuates across groups in the forecast sharing treatments. The Fisher's exact tests show that in all the four types of projects (Inter- or Independent Projects in World I or II), there is no statistically significant difference in the frequency of making correct forecasts between the forecast sharing groups and the baseline group (*p values* > 0.1). This result suggests that while forecast sharing leads to converged forecasts among group members, subjects with forecast sharing do not necessarily make better forecasts. In addition, Fisher's exact tests

[13] Besides SDW, a very similar procedure has been employed in Bapna and Umyarov (2015).

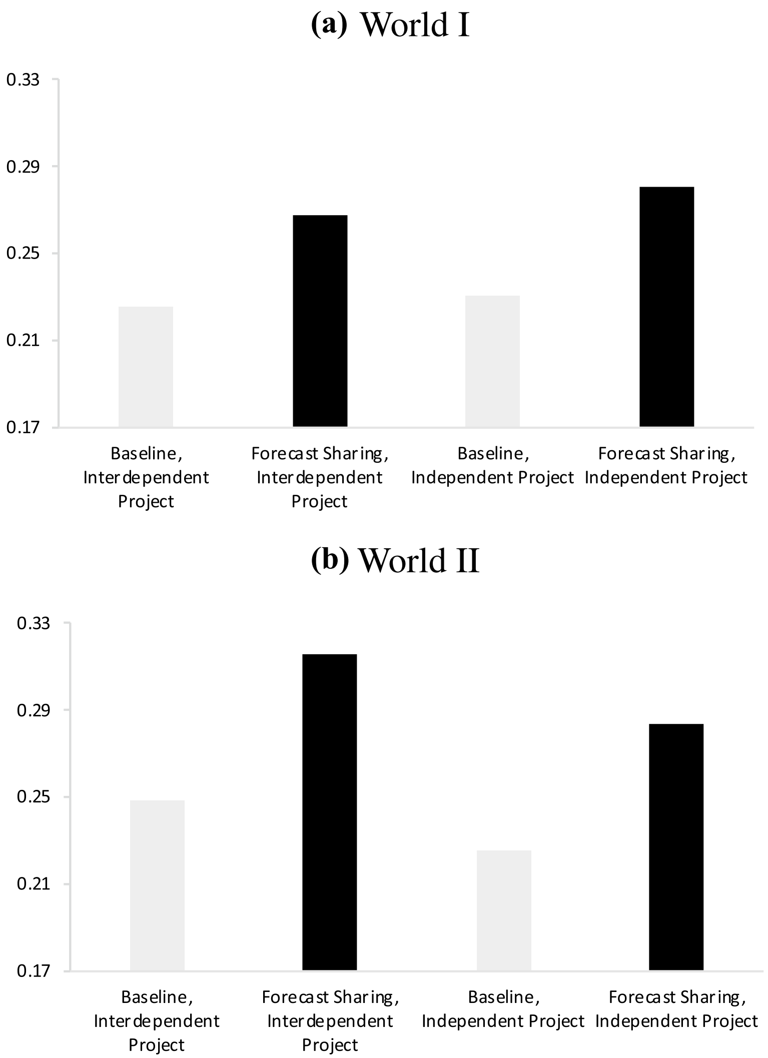

Fig. 4 Average HHI. *Note:* The bar for the baseline group is the mean of the 1000 simulations (each including 6 simulated HHIs)

Table 5 Comparisons of HHI

	Interdependent Project	Independent Project
World I		
Frequency	< 0.05	< 0.01
World II		
Frequency	< 0.05	< 0.01

This table reports the frequencies of not having a significantly higher HHI in the forecast sharing treatment than in the baseline treatment

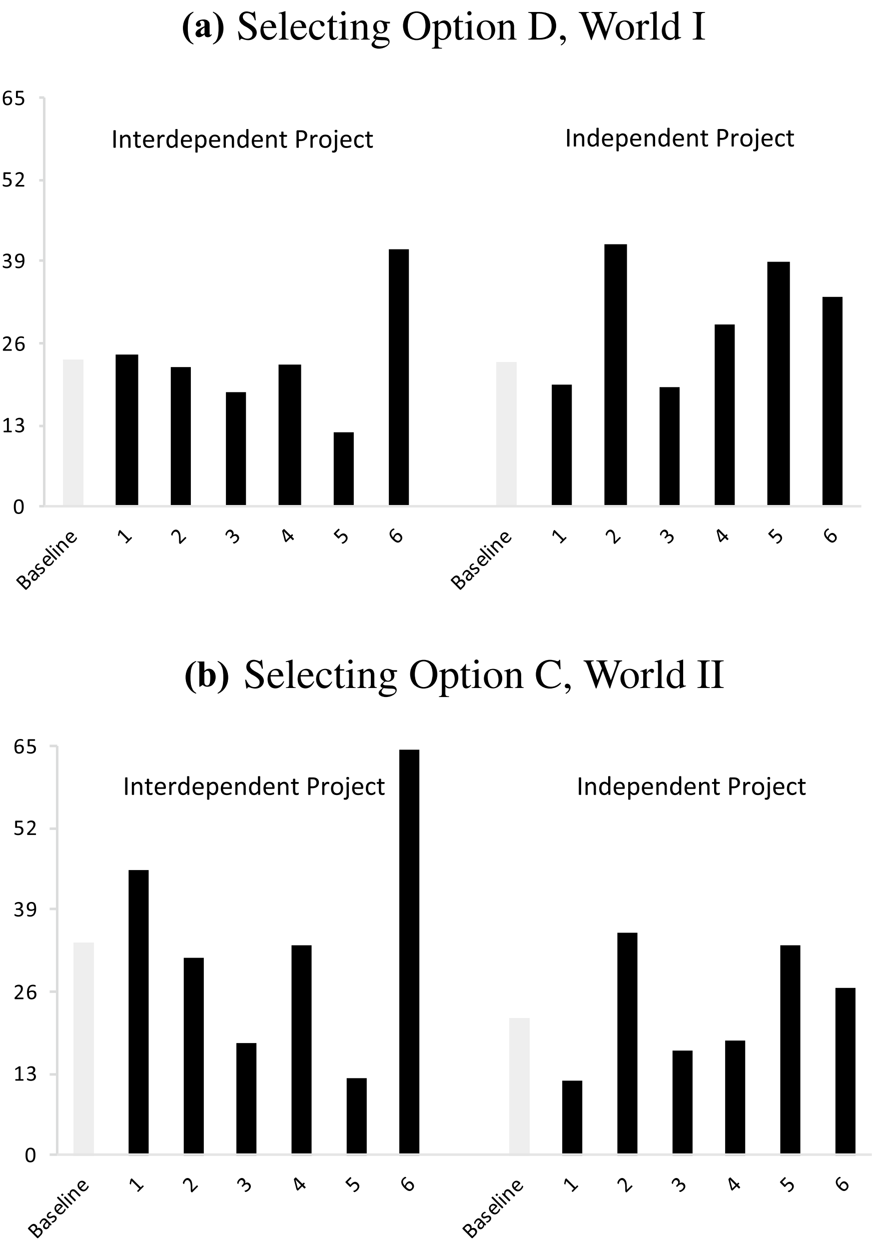

Fig. 5 Frequencies of making correct forecasts. *Note:* The numbers 1, 2, … 6 indicate the six forecast sharing groups respectively

show that for each type of projects, the frequencies of choosing the extreme forecast options (A or E) are not significantly different between the forecast sharing groups and the baseline group (p *values* > 0.1).

Figure 3 shows fluctuating frequencies of choosing not to invest across groups, i.e. the proportion of subjects making the individually rational decision diverges across the forecast sharing groups. We conducted the Fisher's exact test for each of the four types of projects, and found that in all the four types of projects, there is no statistically significant difference in terms of the frequency of investment between the forecast sharing groups and the baseline group (*p values* > 0.1).

Result 1b *Forecast sharing neither leads to better forecasts nor significantly alters the subjects' investment decisions.*

While Result 1a lends support to Hypothesis 1a, Result 1b does not support Hypothesis 1b. This result is consistent with findings in a recent study by Chen et al. (2018). They find that how the information of others' choices affects performance in answering knowledge-related questions depends on the difficulty of the question relative to the knowledge of the subjects: Improvement occurs for relatively easy questions, while for hard questions, providing the information could even be harmful for performance. In our context, with the complexity of the forecasting task, the information on others' choices does not help in improving individual performance.

To further investigate how divergent the forecasts are under forecast sharing, we adopt the index of dissimilarity to measure the variation in forecasts across groups. The index of dissimilarity was first used by Cutler et al. (1999) to measure the residential segregation of two groups of people distributed across areas. Gentzkow and Shapiro (2011) use the index to capture the extent to which liberals and conservatives are exposed to different facts and opinions. We use the dissimilarity index to measure the extent to which the options are selected disproportionately by subjects in different groups; a higher dissimilarity index indicates more variable choices of subjects and higher imbalance of forecasts across groups. Formally, the index of *option dissimilarity,* of option i is defined as

$$option\ dissimilarity_i = \frac{1}{2}\sum_{j=1}^{G}\left|\frac{c_{ij}}{c_i} - \frac{d_{ij}}{d_i}\right|$$

where c_{ij} is Option i's selection count in group j, c_i is Option i's selection count in all groups, d_{ij} is the selection count of all the options other than i in group j, and d_i is the selection count of all the options other than i in all groups. G is the number of groups in the treatment. The index ranges from 0, when the selection of Option i is as equally balanced among the groups as the selection of the other options, to 1, when option i is selected by everyone in some groups but not at all in the other groups.

The *treatment dissimilarity* is the average of the T option dissimilarities:

$$treatment\ dissimilarity = \frac{1}{T}\sum_{k=1}^{T} option\ dissimilarity_k$$

where T is the number of options.

In order to examine the effects of forecast sharing on dissimilarity indices, we conduct the following statistical procedure for both the option dissimilarity and the treatment dissimilarity, which is similar to the one of SDW and the procedure we have conducted for HHI.

Step 1 We calculate the dissimilarity index for every 2 of the 6 forecast sharing groups, resulting in 15 indices for the forecast sharing treatment;
Step 2 We randomly and evenly split the one group in the baseline treatment into two groups so that their group size is the same as that in the forecast sharing treatment, and then calculate the dissimilarity index for the two groups. We repeat the step 15 times, resulting in 15 indices for the baseline treatment;
Step 3 We conduct the two-sided Mann–Whitney *U* test on the dissimilarity indices of the forecast sharing treatment and the baseline treatment (using the results in Step 1 and 2) and record the *p value*;
Step 4 We repeat Step 2 to 3 1000 times and calculate the frequency of the cases where the value is not significantly higher in the forecast sharing treatment than in the baseline treatment, where significant difference is determined by $Z - statistic < 0$ and $p\, value < 0.05$.

Figures 6 and 7 display the average option dissimilarity of the correct forecasts and treatment dissimilarity respectively, for the baseline and the forecast sharing groups under the four types of projects, generated from the above statistical procedure. The results of the statistical tests in Step 4 are reported in Tables 6 and 7.

Figure 6 shows that the option dissimilarities of the correct forecasts in the forecast sharing treatments are higher than those in the baseline treatments, while the statistical results reported in Table 6 suggest that the differences are not statistically significant in most cases. Therefore, the bottom line we can draw is that pre-play communication does not make the selection of the correct forecast more stable across groups.

Figure 7 shows a higher treatment dissimilarity in the forecast sharing groups than groups in the baseline, and Table 7 shows that the differences are statistically significant. That is, while subjects who communicate tend to concentrate on some option as indicated by the higher HHI, subjects in different groups concentrate on different options. The variability of forecasts further substantiates the result that pre-play communication does not necessarily lead to better forecasts. The following result summarizes the analysis on dissimilarity indices:

Result 1c *The forecasts in forecast sharing treatments are of significantly higher variability across groups than those in baseline treatments.*

4.2 Effects of payoff interdependence and mapping between SCI and return multipliers

Hypothesis 2 suggests that when there is a positive spillover in investment, subjects may report higher forecasts and invest more. We then investigate the effects of payoff

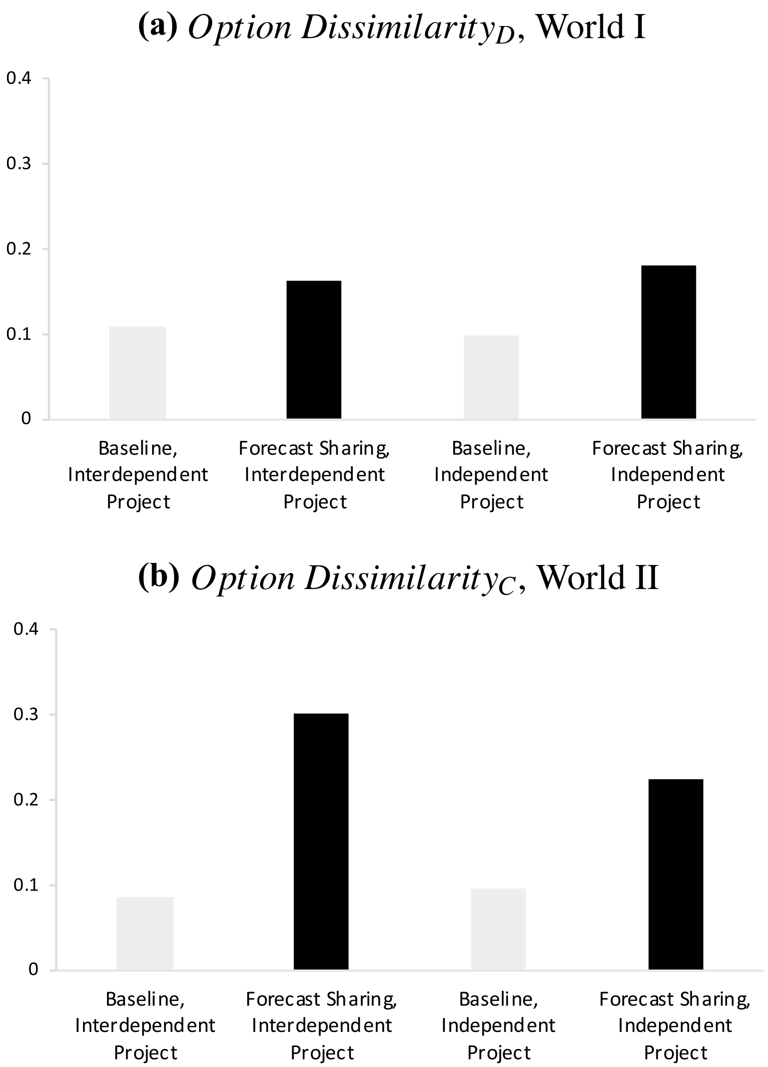

Fig. 6 Option dissimilarity of the correct forecasts. *Note:* The bar for the forecast sharing groups indicates the mean of the 15 option dissimilarity indices, while the bar for the baseline group is the mean of the 1000 simulations (each including 15 simulated option dissimilarity indices) generated from the statistical procedure above

interdependence on forecasts and investment. We let forecast option A = 6, B = 5, …, E = 2, F = 1.[14] The Mann–Whitney *U* test shows that for each of the four scenarios (baseline in World I or in World II, forecast sharing in World I or in World II), there is no statistically significant difference in terms of the level of forecasts or the

[14] Using an alternative coding such as letting forecast equal 1 if option A, B or C is chosen and 0 otherwise does not change our statistical results.

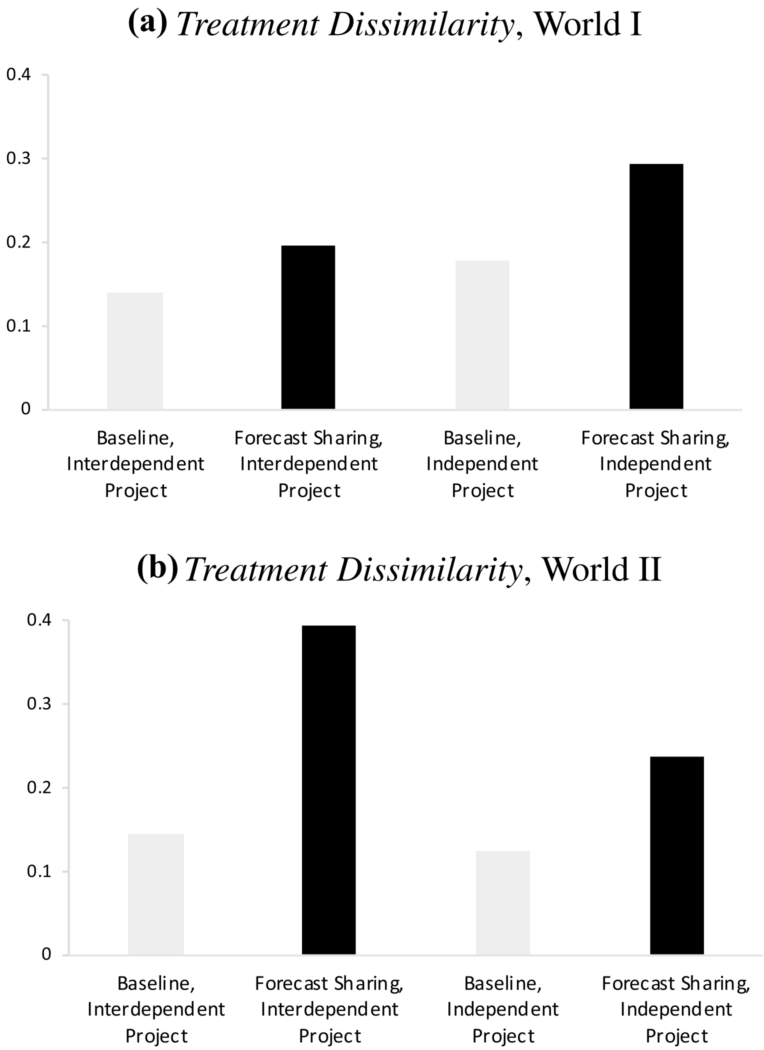

Fig. 7 Treatment dissimilarity. *Note:* The bar for the forecast sharing groups is the mean of the 15 treatment dissimilarity indices, while the bar for the baseline group is the mean of the 1000 simulations (each including 15 simulated treatment dissimilarity indices) generated from the statistical procedure above)

frequency of investment between independent projects and interdependent projects (*p values* > 0.1).[15] This result suggests that subjects do not strategically manipulate

[15] Under forecast sharing treatments, we have six groups for each scenario and thus forecasts (investment) may be correlated within a group; therefore, we use average forecasts (investment frequencies) at the group level in running the Mann–Whitney *U* tests. Under baseline treatments, however, there is one group in each scenario; the tests are thus conducted at the subject level. The unique group in each of the baseline treatments also allows us to use the Fisher's exact test to test investment for the baseline scenarios. The above description of statistical tests we conducted also apply for the tests between World I and World II reported below. The Fisher's exact test shows that there is no significant difference in terms of investment frequency between independent and interdependent projects (*p values* > 0.1).

Table 6 Comparisons of option dissimilarity of the correct forecasts

	Interdependent project	Independent project
World I		
Frequency	> 0.1	> 0.1
World II		
Frequency	< 0.05	> 0.1

This table reports the frequency of not having a significantly higher option dissimilarity of the correct forecast in the forecast sharing treatment than in the baseline treatment

Table 7 Comparisons of treatment dissimilarity

	Interdependent project	Independent project
World I		
Frequency	< 0.1	< 0.05
World II		
Frequency	< 0.01	< 0.01

This table reports the frequency of not having a significantly higher treatment dissimilarity in the forecast sharing treatment than in the baseline treatment

their forecasts, when forecasts are shared, to induce peers' investment in the presence of investment spillovers. One possible reason that may account for this observation is that with the relatively large group size (43 subjects in a group on average), one subject's influence on the overall distribution of forecasts is small, and thus subjects have weak incentive to manipulate the forecasts.[16,17]

Result 2 *There is no significant difference in terms of forecasts or investment frequency between independent and interdependent projects.*

World I and World II differ in the mapping between SCI and return multipliers. Given each realization of SCI, World II always yields a higher return multiplier, m, than World I. Hypothesis 3 suggests that subjects in World II would invest more often than in World I. However, the Mann–Whitney U test shows that, there is no

[16] We thank an anonymous reviewer for constructive comments on this issue.

[17] We also observe that for baseline groups in World I, subjects invest less often in interdependent projects (68%) than in independent projects (77%). The Fisher's exact test shows that the difference is statistically insignificant. More pessimistic forecasts in the interdependent projects (at least partially) account for the difference in investment frequency: In the interdependent projects, 39% of the subjects chose Option A or B in their forecasts, while in the independent projects, 41% of the subjects chose Option A or B in their forecasts, in the baseline World I.

statistically significant difference in terms of investment frequency between World I and World II in the four scenarios (baseline or forecast sharing treatment with independent or interdependent payoffs; *p values* > 0.1).[18]

Result 3 *There is no significant difference in terms of investment frequency between World I and World II.*

We also observe from Figs. 3 and 5 that in the baseline World I, while the percentages of making correct forecasts are very close between independent and interdependent projects, the investment frequency is lower under interdependent projects than under independent projects. Moreover, in the baseline World II, for interdependent projects, the percentage of correct forecasts (33%) is higher than the percentage of making the individually rational decision, "Not Invest" (23%); for independent projects, however, the percentage of correct forecasts (22%) is lower than the percentage of making the individually rational decision (24%). These results might be surprising because given beliefs, subjects with pro-social preferences may invest more in the presence of investment spillovers. Two explanations may account for these observations. First, a correct forecast is not a necessary condition for an *ex post* individually rational decision, even for a purely self-interested subject. In World I (II), any self-interested subject with forecast of SCI below 3560 (3480) would choose not to invest. Second, given an individual's belief which is a distribution over the options in the forecasting task, the forecast option that one believes to be the most likely may not coincide with the option that maximizes the expected utility of the subject (assuming that the subject will make the individually rational investment decision according to the m under this option). Therefore, choosing the correct forecast option is even not a sufficient condition for making the individually rational investment decision. In addition, the Fisher's exact test shows that there is no significant difference in terms of the frequency of choosing the correct forecast or of investment between independent and interdependent projects in the baseline World II (*p values* > 0.1).

4.3 Regression results

In this subsection, we report results from regressions of forecasts and investment. Table 8 reports results from ordinal logistic regressions of subjects' forecasts on dummy variables FS and $INTER$, with $FS = 1$ standing for forecast sharing and $INTER = 1$ for the interdependent projects, and their interaction term $FS \cdot INTER$, for World I (Columns (1)–(2)) and World II (Columns (3)–(4)) respectively, controlling for group fixed effects. Columns (1) and (3) also control for demographical

[18] The Fisher's exact test also shows that there is no significant difference in terms of investment frequency between World I and II for the baseline scenarios (*p values* > 0.1).

Table 8 Regression results on forecasts

Forecast	World I		World II	
	(1)	(2)	(3)	(4)
FS	0.1809	0.1561	− 0.3943	− 0.4522
	(0.4832)	(0.4642)	(0.4386)	(0.3884)
INTER	− 0.3161	− 0.2553	− 0.1079	− 0.2041
	(0.3159)	(0.2940)	(0.2924)	(0.2756)
$FS \cdot INTER$	− 1.0101	− 0.8173	0.0174	0.1538
	(0.5985)	(0.5972)	(0.5207)	(0.4678)
Test of linear restrictions [with sum of point estimates presented]				
$INTER + FS \cdot INTER = 0$	− 1.3262*	− 1.0726*	− 0.0905	− 0.0503
$FS + FS \cdot INTER = 0$	− 0.8292*	− 0.6612	− 0.3769	− 0.2984
Pseudo R^2	0.0251	0.0065	0.0531	0.0261
Demographic variables controlled for	Yes	No	Yes	No
Observation	649	699	630	686

(1) This table presents ordinal logistic regressions of forecasts on dummies denoting forecast sharing treatments and interdependent projects, and their interaction term

(2) Group fixed effects are controlled for

(3) The first two columns use data in World I. The last two columns use data in World II

(4) Heteroskedasticity robust standard errors are presented in parentheses. *, ** and *** represent significance at 0.05, 0.01, and 0.001 levels respectively

(5) Results of the demographic correlates are presented in Online Appendix E

variables.[19] We still use the coding of forecast option A = 6, B = 5, …, E = 2, F = 1 so that a more optimistic forecast corresponds to a higher value. Table 9 reports results from logit regressions of investment decisions on treatment variables *FS*, *INTER* and $FS \cdot INTER$, for World I (Columns (1)–(3)) and World II (Columns (4)–(6)) respectively, controlling for demographical variables and group fixed effects. Column (2) and (5) include *Forecast*, an ordinal variable with forecast option A = 6, B = 5, …, E = 2, F = 1. Column (3) and (6), instead, include dummy variables on the subjects' forecasts, $f_E - f_A$ (in increasing order).

In Table 8, the coefficients of *FS* and $FS + FS \cdot INTER$ represent the effects of forecast sharing on subjects' forecasts given payoff independence and payoff interdependence respectively. We find that the coefficients of *FS* are insignificant while the coefficient of $FS + FS \cdot INTER$ is negative and significant only in World

[19] We included the demographic variables reported in Table 4: a dummy variable on gender, dummy variables on residential areas (with living in *Eastern China* being the default) and on occupation (with *Student* being the default), as well as variables indicating age, education (*EDU*) and household income (*INCOME*). EDU = 1, 2, 3, 4, or 5 means the highest education level being junior high school or below, senior high school, professional school, university or graduate school respectively; INCOME = 1, 2, 3, 4, 5, 6 or 7 means annual household income being 30,000 or below, 30,000–80,000, 80,000–300,000, 300,000–500,000, 500,000–800,000, 800,000–2,000,000 or 2,000,000 or above respectively. The same applies to the other regression tables when demographical variables are controlled for.

Table 9 Regression results on investment

Investment	World I			World II		
	(1)	(2)	(3)	(4)	(5)	(6)
FS	− 0.0787	− 0.1904	− 0.1633	0.4513	0.6399	0.6817
	(0.4615)	(0.4742)	(0.4704)	(0.4592)	(0.4991)	(0.5067)
INTER	− 0.4741	− 0.4510	− 0.4947	0.2828	0.2779	0.2897
	(0.3685)	(0.3859)	(0.3921)	(0.3739)	(0.3856)	(0.3935)
$FS \cdot INTER$	0.6308	1.1235	1.1277	− 0.8857	− 1.0368	− 1.0042
	(0.6604)	(0.6928)	(0.6934)	(0.6528)	(0.6952)	(0.7069)
Forecast		0.6083***			0.5774***	
		(0.0910)			(0.0919)	
f_E			1.0078			0.1036
			(0.6462)			(0.6814)
f_D			1.7030***			1.2800*
			(0.4957)			(0.5623)
f_C			2.2819***			1.4219**
			(0.4985)			(0.5403)
f_B			3.0107***			2.4557***
			(0.5544)			(0.5706)
f_A			3.0668***			2.5877***
			(0.5748)			(0.5892)
Test of linear restrictions [with sum of point estimates presented]						
$INTER + FS \cdot INTER$	0.1567	0.6725	0.6330	− 0.6029	− 0.7589	− 0.7145
$FS + FS \cdot INTER$	0.5521	0.9331	0.9644	− 0.4344	− 0.3970	− 0.3225
Pseudo R^2	0.0403	0.1169	0.1204	0.0776	0.1498	0.1562
Observation	649	649	649	630	630	630

(1) This table presents logistic regressions of investment on dummies denoting forecast sharing and interdependent projects, and their interaction term. In column (2) and (5), an ordinal variable denoting forecast is included (with $Forecast = 6, 5, \ldots, 1$ corresponding to option A to option F respectively). In column (3) and (6), dummy variables denoting forecast are included (with $f_A \ldots f_E$ denoting the choice being option A to E respectively)

(2) Demographic variables and group effects are controlled for. Demographic correlates are presented in Online Appendix E

(3) The first three columns use data in World I. The last three columns use data in World II

(4) Heteroskedasticity robust standard errors are presented inside parentheses. *, **, and *** represent significance at 0.05, 0.01 and 0.001 levels, respectively

I when demographic variables are controlled for (and insignificant in all the other three cases), meaning that for interdependent projects in World I, subjects seem to report more pessimistic forecasts when sharing forecasts. The coefficients of *INTER* and $INTER + FS \cdot INTER$ represent the effects of payoff interdependence on subjects' forecasts under the baseline and under forecast sharing respectively. We find that the coefficients of *INTER* are insignificant while the coefficients of $INTER + FS \cdot INTER$ are negative and significant in World I (and insignificant in

World II). Meanwhile, the coefficients of the interaction term, $FS \cdot INTER$, are statistically insignificant in all the regressions. These results suggest that forecast sharing or payoff interdependence do not have unambiguous effects on the optimism of forecasts. This could be explained by the variability of forecasts under forecast sharing reported in Result 1c. However, a bottom line we can draw is that given spillovers of investment, the subjects did not over-report forecasts out of strategic motives when forecasts are shared.

In regressions on investment reported in Table 9, the coefficients of *INTER*, *FS* and the interaction term $FS \cdot INTER$ are all statistically insignificant. Moreover, we cannot reject the null hypotheses that the sum of the coefficients of *INTER* and $FS \cdot INTER$ equals zero and that the sum of the coefficients of *FS* and $FS \cdot INTER$ equals zero. Overall, the regression results reported in Table 9 suggest that neither forecast sharing nor investment spillovers has unambiguous effects on investment. If we repeat the regressions in Table 9 using only the data of those who report forecasts corresponding to positive m (A, B, C or D), the results are qualitatively the same.

We then look into the relationship between the unincentivized forecasts and the incentivized investment decisions. Columns (2), (3) (5), and (6) in Table 9 reveal a substantial, positive correlation between forecasts and investment decisions. In Columns (2) and (5), the coefficients of *Forecast* are positive and statistically significant. In Columns (3) and (6), four out of five coefficients of the forecast variables are positively significant; and the coefficients increase in the forecasts.[20] These results indicate that there is a significant increase in the likelihood of investment as forecasts increase. Therefore, while the Forecasting Task is not incentivized, we find a significant connection between the reported forecast and the investment decision, which confirms that the (cheap-talk) forecasts reported in the costless communication are associated with individuals' underlying beliefs.

Result 4 *Forecasts reported are significantly and positively associated with investment decisions.*

Online Appendix E reports the deomographic correlates of regressions in Tables 8 and 9. It shows that there is no substantial individual heterogeneous effect on forecasts or investment: Most of the demographic correlates are statistically insignificant. However, we find that in World I, private sector employees report significantly lower forecasts than students, while in World II, the reported forecasts significantly decrease as education level increases; meanwhile, investment in World II is more likely for female, older or lower income subjects. However, no

[20] Moreover, the tests of linear restrictions show that the coefficients for f_i are significantly different from each other. Pairwise tests among the coefficients for f_i yield similar results: The difference is significant in all of the comparisons ($p\,values < 0.05$), except for the comparisons of f_A versus f_B and f_D versus f_E in World I, and f_A versus f_B and f_C versus f_D in World II.

Table 10 Path dependence in forecast sharing groups

Dependent variable: I_{mode}	(1)	(2)
ID	0.0149*	0.0142*
	(0.0068)	(0.0062)
WORLDI	0.5509	0.0492
	(0.6023)	(0.5805)
INTER	− 0.1616	− 0.0470
	(0.5505)	(0.5165)
Demographic variables controlled for	Yes	No
Pseudo R^2	0.0617	0.0415
Observation	867	931

(1) This table reports results of logistic regressions using data from the forecast sharing groups.

$I_{mode} = 1$ if the subject chose the option that was most commonly selected by group members entering earlier than her. *ID* indicates the entering order such that $ID = j$ for the jth subject entering a group. $WORLDI = 1$ for World I and 0 otherwise

(2) Group fixed effects are controlled for

(3) Heteroskedasticity robust standard errors are presented in parentheses

(4) *, **, and *** represent significance at 0.05, 0.01, and 0.001 levels respectively

demographical variable is consistently associated with forecasts or investment in both worlds.[21]

4.3.1 Path dependence of forecasts

The concentration and variability of forecasts under forecast sharing reported in Sect. 4.1 suggests possible path dependence under forecast sharing. To examine this possibility, we investigate in forecast sharing treatments, whether and how the forecasts made by group members who entered the game earlier than a subject influenced the subject's choice in the forecasting task. We created a dummy variable, I_{mode}, which equals one if the subject chose the option that was most commonly chosen by the group members who entered earlier than the subject, and zero otherwise. Using the data from the forecast sharing groups, we then regressed this dummy variable on a variable indicating the entering order of the subject, *ID* (with $ID = j$ for the jth subject entering the group) and treatment dummy variables, *WORLDI* and *INTER* indicating World I and interdependence of payoffs respectively. Table 10 reports the logistic regression results, where column (1) controls for demographic variables and column (2) does not. Table 10 shows that the coefficients of I_{mode} are

[21] We also ran regressions similar to those reported in Column (1) and (3) of Table 8, adding interaction terms between the demographical variables and treatment variables, *FS*, *INTER* and $FS \cdot INTER$, and found that none of these interaction terms is significantly associated with forecasts in both worlds.

positively significant, suggesting that the later a subject enters the game, the more likely s/he is to choose the most popular choice of the group members entering earlier than him/her, suggesting a pattern of path dependence. We also ran similar regressions with data from the baseline groups, where the coefficients of I_{mode} are statistically insignificant.

5 Conclusion

This paper explores how individual forecasts and investment decisions are influenced by costless communication among individuals. In particular, we focus on the effects of forecast sharing on individuals' forecasts about the financial market as well as the effects on their subsequent investment decisions. First, our online experimental evidence shows that the unincentivized forecasts reported in the costless preplay communication are positively associated with subjects' investment decisions. Moreover, the reported forecasts tend to converge when forecasts are shared. However, communication does not necessarily lead to better individual performance in the financial market: The shared forecasts are not closer to the true state, nor do people make more individually rational investment decisions. Instead, we show that the forecasts while converging within a group, become significantly more variable across groups when shared. Moreover, in the investment games, the public good feature of investment does not affect either forecasts on or investment decisions in the market.

Our paper is the first to explore the use of WeChat as a platform for economic experiments. We have detailed in the paper the advantages and procedure of running online experiments using WeChat. This platform is particularly advantageous for researchers to study dynamic issues, such as social learning and contagious manias in financial markets followed by a crash, which are difficult to study using traditional experimental methods in the laboratory with a limited number of subjects. Our experiment conducted via WeChat thus showcases a promising experimental approach to study massive-scale economic and financial systems.

Acknowledgements We thank the editor Charles Noussair and three anonymous reviewers for their constructive comments, which helped us improve the manuscript. We are grateful to Jiong Li and Shenzhen Quantum Net Technology Co., LTD for sharing WeChat user pools and providing web-based experimental programs for our study. We acknowledge the excellent research assistance from Ran Kang and Junjie Ren, and the support provided by Qianhai Institute for Innovative Research (QIIR). We also thank Liang Guo, Arthur Schram and conference participants in the China Meeting of the Econometric Society (Chengdu), the International Workshop on Experimental Economics (Xiamen), the Asian Meeting of the Econometric Society (Hong Kong), the Experimental Finance 2017 (Nice), the 13th Artificial Economics Conference (Tianjin), the Behavioral and Experimental Conference (Dalian) and the 2019 Asian-Pacific Meeting of ESA (Abu Dhabi), as well as seminar participants at Chinese University of Hong Kong, Dongbei University of Finance and Economics, Hong Kong University of Science and Technology, Nankai University, Sun Yat-sen University and Xiamen University for their valuable comments.

References

Agastya, M., Menezes, F., & Sengupta, K. (2007). Cheap talk, efficiency and egalitarian cost sharing in joint projects. *Games and Economic Behavior*, *60*(1), 1–19.

Alevy, J. E., Haigh, M. S., & List, J. A. (2007). Information cascades: Evidence from a field experiment with financial market professionals. *Journal of Finance*, *62*(1), 151–180.

Angner, E. (2016). *A course in behavioral economics*. New York: Palgrave Macmillan.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance*, *59*(3), 1259–1294.

Attema, A. E., Bleichrodt, H., Gao, Y., Huang, Z., & Wakker, P. P. (2016). Measuring discounting without measuring utility. *American Economic Review*, *106*(6), 1476–1494.

Banerjee, A. V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics*, *107*(3), 797–817.

Bapna, R., & Umyarov, A. (2015). Do your online friends make you pay? A randomized field experiment on peer influence in online social networks. *Management Science*, *61*(8), 1741–2011.

Bénabou, R. (2013). Groupthink: Collective delusions in organizations and markets. *Review of Economic Studies*, *80*(2), 429–462.

Bénabou, R. (2015). The economics of motivated beliefs. *Revue d'économie politique*, *125*(5), 665–685.

Bénabou, R., & Tirole, J. (2016). Mindful economics: The production, consumption and value of beliefs. *Journal of Economic Perspectives*, *30*(3), 141–164.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis*, *20*(3), 351–368.

Bikhchandani, S., & Sharma, S. (2000). Herd behavior in financial markets. *IMF Staff Papers*, *47*(3), 279–310.

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H. (2010). Belief elicitation in experiments: Is there a hedging problem? *Experimental Economics*, *13*(4), 413–438.

Bochet, O., Page, T., & Putterman, L. (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization*, *60*(1), 11–26.

Brown, R. (1986). *Group polarization*. New York: The Free Press.

Bursztyn, L., Ederer, F., Ferman, B., & Yuchtman, N. (2014). Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica*, *82*(4), 1273–1301.

Charness, G., Gneezy, U., & Halladay, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior and Organization*, *131*(Part A), 141–150.

Charness, G., & Sutter, M. (2012). Groups make better self-interested decisions. *Journal of Economic Perspectives*, *26*(3), 157–176.

Chaudhuri, A., & Paichayontvijit, T. (2006). Conditional cooperation and voluntary contributions to a public good. *Economics Bulletin*, *3*(8), 1–14.

Chen, G., Lien, J., & Zheng, J. (2018). *The value of the knowledge of others*. Paper presented at 2nd DUFE Behavioral and Experimental Economics Workshop, Dalian, China.

Chen, H., De, P., Hu, Y., & Hwang, B. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, *27*(5), 1367–1403.

Chen, Y., & Konstan, J. (2015). Online field experiments: A selective survey of methods. *Journal of the Economic Science Association*, *1*(1), 29–42.

Cipriani, M., & Guarino, A. (2005). Herd behavior in a laboratory financial market. *American Economic Review*, *95*(5), 1427–1443.

Cutler, D. M., Glaeser, E. L., & Vigdor, J. L. (1999). The rise and decline of the American ghetto. *Journal of Political Economy*, *107*(3), 455–506.

Dawes, R., McTavish, J., & Shaklee, H. (1977). Behavior, communication and assumptions about other people's behavior in a common dilemma situation. *Journal of Personality and Social Psychology*, *35*, 1–11.

Ehm, C., Laudenbach, C., & Weber, M. (2018). Focusing on volatility information instead of portfolio weights as an aid to investor decisions. *Experimental Economics*, *21*(2), 447–480.

Exadaktylos, F., Espin, A. M., & Branas-Garza, P. (2013). Experimental subjects are not different. *Scientific Report*, *3*, 1213.

Gächter, S., & Renner, E. (2010). The effect of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, *13*(3), 365–377.

Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *Quarterly Journal of Economics*, *126*(4), 1799–1839.
Glaeser, E., & Sunstein, C. (2009). Extremism and social learning. *Journal of Legal Analysis*, *1*(1), 263–324.
Grinblatt, M., Titman, S., & Wermers, R. (1995). Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. *American Economic Review*, *85*(5), 1088–1105.
Hergueux, J., & Jacquemet, N. (2014). Social preferences in the online laboratory: A randomized experiment. *Experimental Economics*, *18*(2), 251–283.
Hogg, M. A., Turner, J. C., & Davidson, B. (1990). Polarized norms and social frames of reference: A test of the self-categorization theory of group polarization. *Basic and Applied Social Psychology*, *11*(1), 77–100.
Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, *14*(3), 399–425.
Isaac, R., & Walker, J. (1988). Communication and free riding behavior: The voluntary contributions mechanism. *Economic Inquiry*, *26*(4), 585–608.
Janis, I. L. (1982). *Victims of groupthink*. Boston: Houghton Mifflin.
Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790.
McGrath, K., Elbanna, A., Hercheui, M., Panagiotopoulos, P., & Saad, E. (2012). Exploring the democratic potential of online social networking: The scope and limitations of E-participation. *Communications of the Association for Information Systems*, *30*: Article No. 16.
Mojzisch, A., Grouneva, L., & Schulz-Hardt, S. (2010). Biased evaluation of information during discussion: Disentangling the effects of preference consistency, social validation, and ownership of information. *European Journal of Social Psychology*, *40*(6), 946–956.
Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, *341*(6146), 647–651.
Oechssler, J., Schmidt, C., & Schnedler, W. (2011). On the ingredients for bubble formation: Informed traders and communication. *Journal of Economic Dynamics and Control*, *35*(11), 1831–1851.
Roux, N., & Sobel, J. (2015). Group polarization in a model of information aggregation. *American Economic Journal: Microeconomics*, *7*(4), 202–232.
Salganik, J. M., Dodds, S. P., & Watts, J. D. (2006). Experimental study of inequality and variability in artificial cultural market. *Science*, *311*(5762), 854–856.
Shiller, R. J. (1992). *Market volatility*. Cambridge: MIT Press.
Sobel, J. (2014). On the relationship between group and individual decisions. *Theoretical Economics*, *9*(1), 163–185.
Susilo, Y., & Axhausen, K. W. (2014). Repetitions in individual daily activity–travel–location patterns: A study using the Herfindahl–Hirschman Index. *Transportation*, *41*(5), 995–1011.
Zheng, Y., & Wu, G. (2005). Information technology, public space, and collective action in China. *Comparative Political Studies*, *38*(5), 507–536.