Identifying the reasons for coordination failure in a laboratory experiment*

Philipp Külpmann† Davit Khantadze‡

September 27, 2022

We investigate the effect of the absence of common knowledge on the outcomes of coordination games in a laboratory experiment. Using cognitive types, we can explain coordination failure in pure coordination games while differentiating between coordination failure due to first- and higher-order beliefs.

In our experiment, around 76% of the players chose the payoff-dominant equilibrium strategy despite the absence of common knowledge. However, 9.33% of the players had first-order beliefs that led to coordination failure, and another 9.33% exhibited coordination failure due to higher-order beliefs.

JEL codes:	C72, C92, D83
Keywords:	Higher-order beliefs, coordination failure, cognitive abilities, experimental economics, game theory

*We would like to thank Andreas Blume, Zahra Gambarova, Wieland Müller, Mahnaz Nazneen, Eugenio Proto, Andis Sofianos, Stefan Trautmann and especially Christoph Kuzmics and Daniel Sgroi for valuable comments and suggestions. Philipp acknowledges support received through grants from the German Research Foundation (DFG, grant number: KU 3071/1) and the German Academic Exchange Service (DAAD). Furthermore, he would like to thank the Department of Economics at the University of Warwick, where this research was carried out, for its hospitality.

†Vienna Center for Experimental Economics, University of Vienna, `philipp.kuelpmann@univie.ac.at`

‡ISET, Tbilisi State University, `d.khantadze@gmail.com`

1. Introduction

When modeling human interactions, most researchers working in the applied and theoretical fields of economics have assumed that players have common knowledge about the structure of the game. This means that every player knows the structure and that every player knows that everyone else knows the structure, and so on. Furthermore, players have common knowledge about the distribution of unknown factors in the game, such as the other players' types or preferences, and they know that everyone else knows, ... These assumptions of common knowledge are usually made for reasons of tractability: The absence of common knowledge leads to complex belief hierarchies, so-called *higher-order beliefs*. In this paper, we relax the assumption of common knowledge about player types. In the following, we call a belief about the other player's type a first-order belief, and a belief about the belief of the other player about your type (i.e., a belief about the other player's first-order belief) a second-order belief, and so on ad infinitum. When common knowledge is assumed, all levels of beliefs are determined by the structure of the game. However, if common knowledge is not assumed, the structure of beliefs is not clear, and each level of belief can potentially influence the result of the game.

Let's say you want to meet someone in an unknown city, and there are two salient meeting points: the well-known central station and the less-well-known station A. Both you and your partner know that station A is more convenient for each of you, but this is not common knowledge. Thus, if you think your partner does not know how convenient station A is, you would go to the central station. This might result in you meeting at the less-convenient central station or even failing to meet each other entirely. We call this a coordination failure due to a first-order belief.[1] However, even if you are sure that your partner knows how convenient station A is, you might think that your partner thinks you do not know about A. In this case, you should go to the central station, because they would go there. This would be an example of coordination failure due to second-order beliefs. Unfortunately, this does not stop at the level of second-order beliefs. In this example, just like in the game discussed in this paper, coordination could break down due to pessimistic beliefs at any of the infinitely many levels of beliefs.

Based on this example, we can already see that it is not clear what will

[1] Based on the literature, we will call anything except coordination due to the best equilibrium coordination a failure, including coordinating on the basis of a worse outcome (i.e., in this example, the central station).

happen in a coordination game without common knowledge. Thus, we are trying to answer the following questions:

1. **Are the players able to coordinate in the absence of common knowledge?**
2. Can coordination fail because the players underestimate the skill of the other players? Or, in other words, **is there coordination failure due to first-order beliefs?**
3. Can coordination fail because players think "too much" about what others might think? Or, in other words, **does coordination failure occur due to higher-order beliefs?**

The main challenge that is faced when attempting to answer these questions is that directly asking about beliefs might change them (see, e.g., Gächter and Renner (2010)). In our case, asking about beliefs might make subjects start thinking about higher-order beliefs, even if they had none before. Therefore, we need to take a choice-based approach to identify higher-order beliefs. To do so, we modify the setup of Blume and Gneezy (2010), who used a slightly difficult coordination game (the so-called five-sector disc) with a better (i.e., risk- and payoff-dominant) option which is harder to detect and a worse option that everyone can detect. We modified their experimental setup to distinguish the effects of the first- and higher-order beliefs that players form about the cognitive abilities of their opponents (i.e., if players believe that their partners have these cognitive abilities).

Our results show that 76% of the subjects who could identify the better option also used it to coordinate on the bassis of better equilibrium. However, 9.33% of these subjects did not believe that the player they were matched with could detect this better option, causing a coordination failure due to first-order beliefs. Another 9.33% did believe that the other player could detect the better option, but they still chose the worse option, and a coordination failure occurred due to higher-order beliefs.

Thus, we are the first to identify higher-order beliefs using a choice-based approach in a coordination game and to directly link these to coordination failure. Unlike the existing literature (as discussed in more detail in Section 2), we identified beliefs without making any assumptions about the model of higher-order beliefs or limiting rationality. Furthermore, we are investigating the role of beliefs in a coordination game without trying to introduce beliefs "artificially" into the design. Thus, there is no incentive for players to "outsmart" the other players, and we even see that having pessimistic higher-order beliefs hurts both the individual players and the overall welfare of the team (see Section 5).

Section 2 includes a discussion on the related literature and the connections of this literature to this work. Then, in Section 3, we explain the experimental design, predictions derived using the underlying model, and the hypotheses derived from the three questions mentioned above. In Section 4, we present the choice data from the experiment and test the three previously derived hypotheses. The effects of coordination failure due to first- and higher-order beliefs on the expected utility of individuals and the overall welfare of the team is discussed in Section 5. Finally, we briefly conclude in Section 6 and discuss some implications of our results.

2. Related Literature

The literature most closely related to this paper is the literature on models of higher-order beliefs. Within this body of literature, the most popular example and the one most closely related to our work is that initiated by Carlsson and Van Damme (1993) and later extended by Morris, Shin, and Yildiz (2016). These authors use higher-order beliefs (in their model of global games) to identify the risk-dominant equilibrium as a unique and rationalizable outcome of the coordination game. This uniqueness result spawned an extensive number of papers on, among other things, bank runs, arms races, and speculative attacks. Several studies tested the different implications of global games in the laboratory. For example, Heinemann, Nagel, and Ockenfels (2004) shows that differences in behavior are only minor in a game with either private or common information and that the result is more similar to the payoff-dominant equilibrium than to the global games solution in both information settings. Furthermore, Szkup and Trevino (2020) show that the behavior approximates the efficient rather than the risk-dominant equilibrium of the coordination game when the private information is highly precise. However, Weinstein and Yildiz (2007) show that the unique result of global games is fragile based on the exact specifications of the higher-order belief model. Other "nearby" higher-order belief models yield very different "unique" predictions. They show that any rationalizable outcome of the original game can be obtained as the unique rationalizable strategy profile of some higher-order belief model. Thus, the question of what kind of higher-order belief model do players use seems to be an empirical question. We take the first step toward answering this question in this paper.

Applying the original global games model of higher-order beliefs to the game played in our experiment would give us a unique prediction: According to the theory, each player should choose the worse option regardless of whether they can detect a better option or not. Thus, coordination failure occurs due to the higher-order beliefs of each player. The main difference between the previous experimental investigations on global games and this work is that these studies have implemented the model of higher-order beliefs developed by Carlsson and Van Damme (1993) as a feature of the game. Thus, these researchers did not test the underlying model of higher-order beliefs; instead, they tested the prediction while enforcing a certain (commonly known) structure of beliefs. In this study, we are interested in elucidating the underlying structure of beliefs if there is no common belief.

However, the most similar paper to ours is Kneeland (2015). In this work, she experimentally explores the players' level of rationality, a requirement for higher-order beliefs. Her experiment shows that 94% of all players are rational with decreasing numbers for second- (71%), third- (44%), and fourth-order (22%) rationality. Thus, there is no common knowledge of rationality.

While the goals of the latter work and our work are ultimately highly similar (i.e., to describe a fundamental underlying assumption of game theory), the papers differ not only in the described property (rationality vs. beliefs about a type) but also in the approach taken. It is important to consider, however, that Kneeland describes the levels of rationality when subjects are pushed and incentivized to consider rationality, while we are more interested in identifying the problems caused by higher-order beliefs, even if they are not induced by the experimental design. In our experiment, a player has nothing to gain from having higher-order beliefs, and, as you can see in Section 5, the losses experienced by both individuals due to their beliefs, as well as the effect of these losses on the entire welfare of the team, are substantial.

Higher-order beliefs have not only been used in the theoretical and experimental microeconomics literature. Angeletos and Lian (2018) investigated the effect of the absence of common knowledge on general equilibrium effects in a macroeconomic model. Chwe (2013) published a book illustrating the importance of common knowledge in several settings in the social sciences. And finally, Bergemann and Morris (2005) developed and described a robust mechanism design.

3. Experimental design

Designing research questions about higher-order beliefs in the absence of common knowledge presents two major challenges: The first challenge to be overcome is the fact that we want to identify higher-order beliefs without inducing them. It is well known that asking questions about beliefs changes both the beliefs and behavior (see, for example, Gächter and Renner (2010)). Therefore, we need to take a choice-based approach to higher-order belief identification.

On the other hand, we need to naturally evaluate the absence of common knowledge without introducing secondary effects or using deception.[2]

We solve both problems by utilizing Blume and Gneezy's (2000) 5-sector disc. This is a disc with five equally large sectors on it, two black and three white, as depicted in Figure 1.[3] The disc has the same sectors on the front and back sides of the disc, and it can be flipped and rotated.

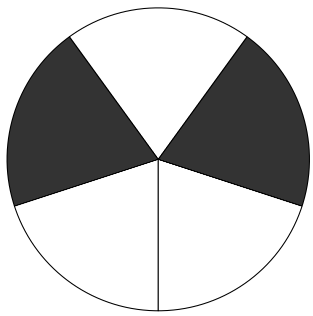


Figure 1: 5-sector-disc

3.1. Treatments

Our subjects played three treatments in a random order without receiving feedback using this disc.[4]

In the **Self Treatment**, each subject gets the disc twice. The disc is randomly turned and rotated each time, and the subject gets £5 if they pick the same sector twice.

[2]Examples of secondary effects include inducing Knightian uncertainty or provoking thoughts that the experimenter is trying to trick the individual.

[3]We have also used a second version of this disc, with a significantly harder-to-find distinct sector with adjacent black sectors. However, the fraction of players who could identify the distinct sector on this disc is too small, as explained in Section 3.2.

[4]The first and last treatment are variations of Blume and Gneezy's (2010) original treatments, whereas the second treatment allows us to identify the first-order beliefs and distinguish them from higher-order beliefs.

This treatment uses the most important properties of the disc:

As the disc can be flipped, the subjects face *symmetry constraints* and cannot distinguish all five sectors. For example, let us assume a subject chooses the left black sector when given the disc for the first time. Now the disc is taken away and given back to the subject again, as depicted in Figure 1. If the disc were flipped, the sector would now be the right black sector, but if it were not flipped, it would still be the left black sector. Therefore, the subject cannot be sure that they will choose the correct sector again; they only have a 50% chance (given that the disc was flipped with 50%) to win by choosing either of the two black sectors.

These symmetry constraints cannot be overcome; therefore, only certain "attainable" equilibria are possible, as defined originally in Crawford and Haller (1990), and further developed by Blume (2000) and Alós-Ferrer and Kuzmics (2013).

However, a clever subject might realize that the disc has a single distinct white sector: the sector adjacent to both black sectors (Figure 1). If they choose this sector, it does not matter if the disc was rotated or flipped; they will always be able to choose this uniquely identifiable sector again and win with 100%.

For the subjects, there are then principally three distinguishable sets of sectors: the black sectors (B), the uniquely identifiable white sector (D), and the two other white sectors (W').

However, the key assumption made in the experiment (and also made in Blume and Gneezy (2000) and Blume and Gneezy (2010) and highly supported by their and our findings) is that not all subjects realize that a uniquely identifiable sector leads to *two different cognitive types*: the high type, who can identify the distinct sector, and the low type, who cannot.[5]

This low type faces an additional symmetry constraint and has only two distinguishable sets of sectors to choose from: one of the two black sectors (B) or one of the three white sectors (W) Note that the low type does not know of the existence of another type or a distinct sector.

In the **Prediction Treatment**, one subject (she) is told that another subject (he) plays the *Self Treatment* (with a possibly differently turned and rotated disc). She has to pick one sector, and every time he picks the sector she picked, she gets £2.5.[6]

[5]In this paper, we are going to use female pronouns for high-type players and male pronouns for low-type players. When two high- or low-type players are identified, the first player will be referred to as female and the second as male.

[6]Adding a treatment in which subjects have to predict what another subject does in the Prediction Treatment would, in theory, allow us to check for second-order beliefs explicitly (or, when repeating

Finally, in the **Coordination Treatment**, two players simultaneously pick a sector on a (randomly turned and rotated) disc, and both receive £5 if they both pick the same sector.

All subjects played all three treatments in a random order without receiving feedback after hearing and reading the instructions and completing an extensive quiz:[7]

The experiment was conducted at the DR@W Laboratory at the University of Warwick using the experimental software "z-Tree" developed by Fischbacher (2007). In the experiment, 130 subjects were recruited and received payments between £3 and £18. The entire experiment took around 45 minutes.

3.2. Predictions

To formalize the three research questions stated in the introduction, we need to discuss the predictions for each of the three treatments: In the Self Treatment, a high-cognition player's decision problem has a unique optimal solution, i.e., to pick the distinct sector twice. This gives her a probability of winning of 1. A low-cognition player, however, is not aware of the existence of the distinct sector. Therefore, his unique optimal choice is to pick B at both stages, giving him a probability of winning of $\frac{1}{2}$.

Prediction 1 (Self Treatment)**.** *In the Self Treatment, subjects will pick either D twice or B twice.*

We will refer to the players according to their choices in the first treatment as high- (picked D twice) or low-cognition (picked B twice) type players.

In the Prediction Treatment, the only payoff-relevant information is whether the other player chooses B or D in the Self Treatment. As discussed above, this is solely determined by the type of the observed player.

A low-cognition player will always choose B, as he is unaware that a better choice is available. A high-cognition player's choice depends on her belief β that the other player is or is not a high-cognition type as well. Playing D gives her a chance of $\beta 1$ of winning, whereas playing B gives her only a chance of $(1-\beta)\frac{1}{2}$ of winning. Thus, she will pick D if her belief β is at least $\frac{1}{3}$ and B otherwise.

---

this, any higher-order belief). However, we do not believe this will work with the 5-sector disc, as it probably requires the subject to pay too much attention and invest too much mental effort, which most subjects might not be willing or able to do.

[7] The instructions, data, and a description of the quiz can be viewed in the Online Appendix: `http://hob.kuelpmann.org`.

Prediction 2 (Prediction Treatment)**.** *In the Prediction Treatment, low cognition subjects will pick B, and high-cognition subjects will pick either B or D.*

In the third treatment, we predict low types to play B, as this is the payoff- and risk-dominant equilibrium action from their point of view. High types, however, should play either D or B, depending on their beliefs.

Prediction 3 (Coordination Treatment)**.** *In the Coordination Treatment, we predict that low types will play B and high types will either play B or D.*

In this game, the structure of the belief hierarchy allowing coordination on the basis of D is that each belief has to exceed a certain threshold, regardless of the level in the belief hierarchy.[8]

Let us say that a player has a belief of 1 that she will play against another high-type, and she is 100% certain that this player thinks she is of the high type, but also thinks that the other player thinks that she thinks he is of the low type (i.e., a second-order belief of 0). In this case, she would assume that he would think she would play B. Therefore, he would play B, which makes her best response to play B. This procedure can be extended and applied to any level of the (theoretically) infinite levels of belief: Whenever there is a belief below $\frac{1}{3}$ on any level, the coordination breaks down as the entire hierarchy unravels.

Remark (The model)**.** *The game played in the Coordination Treatment is a pure coordination game with two types, the absence of common knowledge about the type distribution, and symmetry constraints. While you can find the formal model in Appendix A, the exact definition is not necessary to understand the design or results of this work. The most important takeaway from the model is that either one or two equilibria can be found for high-cognition players, namely, (B, B) and, depending on the belief hierarchy, (D, D).*

The validity of these predictions in our experiment will be shown in Section 4.

Remark (Coordination failure)**.** *As we want to investigate coordination failure, we need to have a sufficiently large fraction of players that are of the high type. If this fraction is not sufficiently large, playing B in the Coordination Treatment is the first-best equilibrium for every player. Therefore, we have adjusted the difficulty of the 5-sector disc such that more than* 50% *of the subjects were able to find the distinct sector in our experiment.*[9]

[8] To be more precise, the common p belief that both players are of the high type has to be at least $\frac{1}{3}$. For the derivation of this threshold, see Appendix A.

[9] The difficulty can be adjusted by using a different version of this disc; for more information, see the Online Appendix (`http://hob.kuelpmann.org`).

Under the assumption that high type players have "common knowledge" of the type distribution, $\theta = \frac{1}{3}$ would be sufficient to support a distinct sector equilibrium being the first-best equilibrium.

Remark (Other-regarding preferences and risk-aversion)**.** *This design is a pure-coordination game with only two possible payoffs (either you win or you do not). It is not influenced by other-regarding preferences or risk aversion.*

In the following, we will use shorthand descriptions for the players' strategies, such as "DD B D" means that a player selected the distinct sector twice in the Self Treatment, one of the black sectors in the Prediction Treatment, and the distinct sector in the Coordination Treatment.

3.3. Hypotheses

Now, using the predictions from above, each hypothesis has a corresponding strategy type. "Is coordination possible in the absence of common knowledge?" or, in the words of our model, "Can high-cognition players coordinate on the first-best equilibrium despite the absence of common knowledge?" is represented by a subject playing $DDDD$: She is of the high type, thinks her opponent is of the high type, and plays the strategy corresponding to the first-best equilibrium in the Coordination Treatment.

Hypothesis 1 (Coordination is possible)**.** *A fraction of players which is significantly different from 0 play $DDDD$.*

More precisely, we are testing the null hypothesis, i.e., that the fraction of players we observe playing $DDDD$ can be explained by noise. If we reject the null hypothesis, we will also report the confidence intervals for a fraction of players playing this strategy for the confidence levels 95% and 99%.

The next two hypotheses are based on Blume and Gneezy's (2010) hypothesis that "beliefs matter": Hypothesis 2 formalizes the question "Does coordination fail because some high-cognition players underestimate the fraction of high-cognition players?" or, in other words, "Does coordination failure occur due to first-order beliefs?"

Hypothesis 2 (First-order beliefs matter)**.** *A fraction of players which is significantly different from 0 plays DD B B.*

DD B B means that the subject is of the high type but has a probability larger than $\frac{2}{3}$ of thinking that the other player is of the low type. Therefore, she plays B in both the Prediction and the Coordination Treatment.

Now, an examination of the third type of subject can reveal whether either first-order beliefs or higher-order beliefs are a reason for coordination failure:

Hypothesis 3 (Higher-order beliefs matter)**.** *A fraction of players which is significantly different from 0 play* $DDDB$*.*

If a subject plays $DDDB$, she believes that she is of the high type and thinks her opponent is of the high type, but there is a coordination problem due to higher-order beliefs.

In addition to these tests, our design also allows us to perform another robustness check, namely, an attainable strategy is available that is highly similar to the one we use to identify first- and higher-order beliefs: $DDBD$. This strategy, however, can not be explained by our model. Thus, we perform the following additional robustness check.

Hypothesis (Robustness check)**.** *DD B D is played significantly less often than* $DDBB$ *and* $DDDB$*.*

Referring to the three predictions described above, we consider only four types of strategies as the result of deliberate play: $BBBB$, $DDBB$, $DDDB$ and $DDDD$ We categorize everything else as "noise" or do not recognize it as a result of deliberate play. This is an extremely conservative way to categorize players, as this "noise" not only includes subjects who did not understand the experiment or behave randomly, "Eureka"-learning beliefs (i.e., learning during the experiment that a distinct sector exists) and (non-trivial) beliefs held by low-cognition players, and subjects who always pick a white sector. Based on this categorization, only 3.68% of all strategies are considered as deliberate play, and only 1.12% support one of our hypotheses (see Table 1 for an overview). Everything else is categorized as "noise". While we will also formally test the hypotheses mentioned above, this already shows that misidentifying a subject's random behavior as conforming our hypotheses is very unlikely.

Type	Hypothesis/Description	# of strategies	Chance of random selection
$DDDD$	1: Coordination is possible	1	0.16%
$DDBB$	2: First-order beliefs matter	4	0.64%
$DDDB$	3: Higher-order beliefs matter	2	0.32%
$BBBB$	Low-cognition players	16	2.56%
"Noise"	-	602	96.32%
$WWWW$	(part of "Noise")	80	12.80%

Table 1: Overview of all strategy types

Remark (Types of strategies and number of strategies)**.** *Each type of strategy does not necessarily consist of only one strategy. There are four ways of playing BB in the Self Treatment: I could play: the "left" black sector first, then the "right" black sector, first "left", then "right", "left" twice, or "right" twice.*[10]

Thus, the strategy type BBBB consists of $2^4 = 16$ *different strategies. On the other hand, the strategy type DDDD consists of only a single strategy (because only one D sector exists). Thus, overall, we have* 5^4 *different strategies.*

4. Results

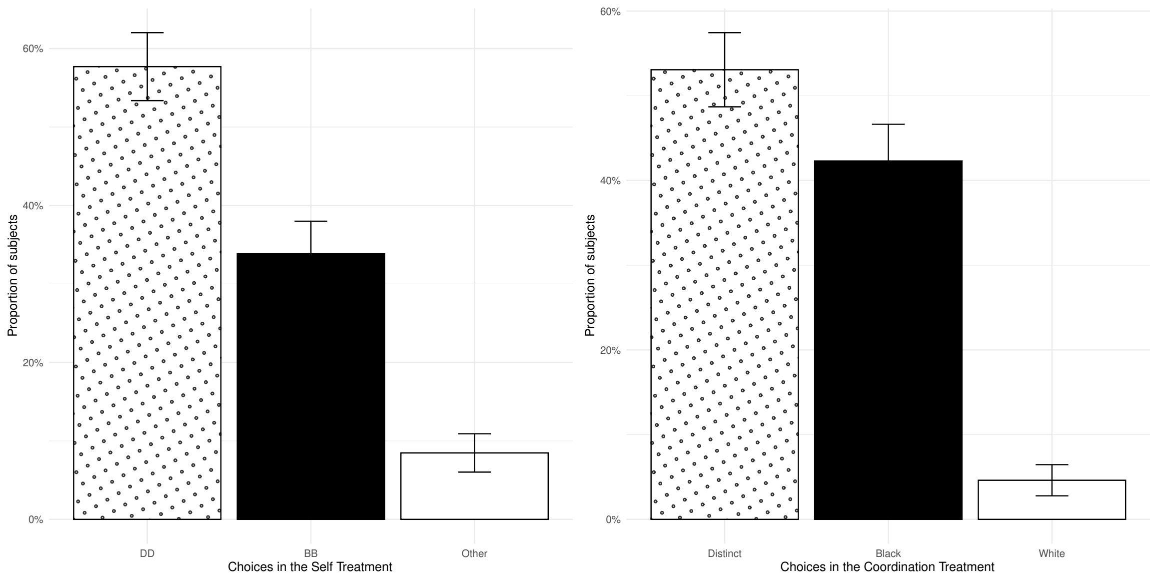
4.1. Predictions

According to **Prediction 1**, we should observe subjects choosing either BB or DD in the Self Treatment. In Figure 2a, we can see that the combined fractions of players playing DD and BB make up around 92%. Any other combination is denoted as "Other", which only accounts for around 8%. The error bars in this and every other figure in this section denote the standard error of the given proportion. Furthermore, we can see that 58% of the subjects have been classified as high types, easily exceeding the threshold of $\frac{1}{3}$ required for playing D in the Prediction and Coordination Treatment.

As we can see in Figure 3a, 80% of the low-cognition subjects played B in the Prediction Treatment, and 97% of the high-cognition subjects played either B or D in the Prediction Treatment (Figure 3b). This confirms **Prediction 2**.

Finally, in Figure 2b, we can see that more than 95% of the subjects have chosen either B or D in the Coordination Treatment, confirming **Prediction 3**.

[10] You could replace "left" and "right" by every other arbitrary description of the two black sectors, e.g., the black sector mentioned first or last on the selection screen.



(a) Self Treatment (b) Coordination Treatment

Figure 2: Player choices in the Self and Coordination Treatment

Given the validity of our auxiliary assumptions, we can now test Hypotheses 1 through 3.

4.2. Hypotheses

Hypothesis 1 (Coordination is possible)**.** *A fraction of players which is significantly different from 0 plays DD D D.*

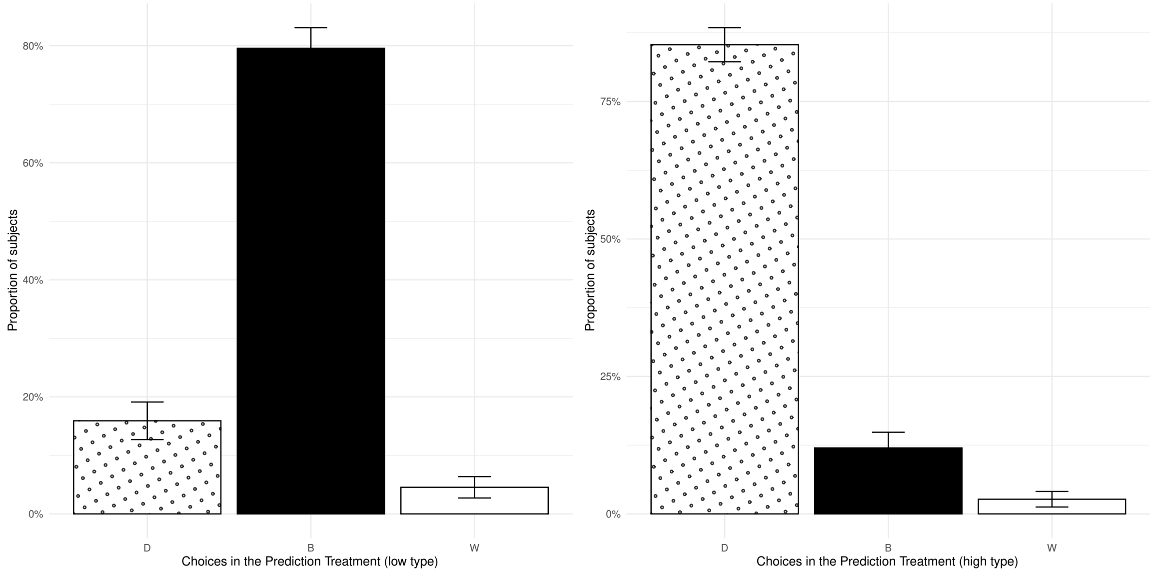
The choice data from our experiment support the first hypothesis. In Figure 4a, we can see that 76% of the high-cognition players have chosen the strategy DD D D. As this strategy represents only 0.16% of all available strategies, we can reject the null hypothesis, i.e., that coordination results from random play ($p < 0.0001$, using a one-sided exact binomial test).[11]

Blume and Gneezy (2010) claim that "beliefs matter", and we test in Hypothesis 2 whether some subjects' pessimistic beliefs about the other players' skills lead to coordination failure.

Hypothesis 2 (First-order beliefs matter)**.** *A fraction of players which is significantly different from 0 plays DD B B.*

Our data support this hypothesis. In Figure 4, we can see that around 9% of the high-cognition players have a first-order belief which leads to

[11]In hindsight, this result is not surprising. However, depending on the underlying model, whether this result would be obtained was not clear. The literature on global games indicate that this is the most popular model, which includes coordination games and the absence of common knowledge. However, the underlying model of the game played in the experiment predicted that this equilibrium would disappear entirely.

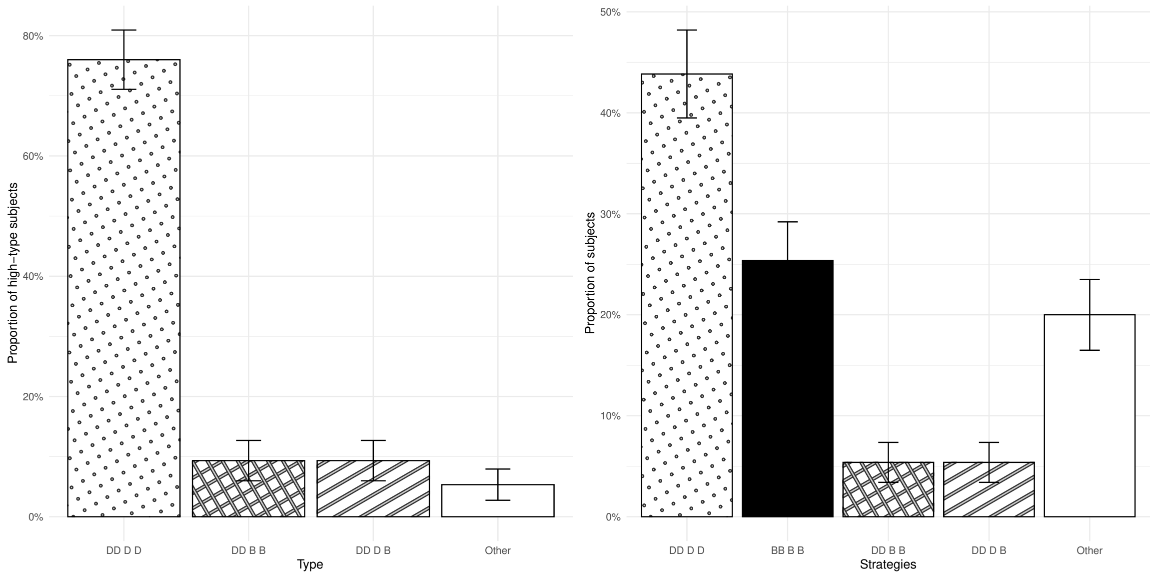


(a) Low-type players

(b) High-types players

Figure 3: Player choices in the Prediction Treatment

coordination failure. Thus, we can reject the null hypothesis, i.e., that this result is due to chance ($p < 0.0001$, using a one-sided exact binomial test, $\alpha = 1\%$).



(a) High-cognition players

(b) All players

Figure 4: Proportions of strategies used

However, do players use higher-order beliefs in this type of game? Testing Hypothesis 3 enables us to answer this question.

Hypothesis 3 (Higher-order beliefs matter)**.** *A fraction of players which is significantly different from 0 plays DD D B.*

From Figure 4b and Figure 4a, we can see that some high-cognition players think that their partner has a high probability of being a high type, but also think that coordination problems may occur. Again, we can reject the null hypothesis at the 1% level ($p < 0.0001$).

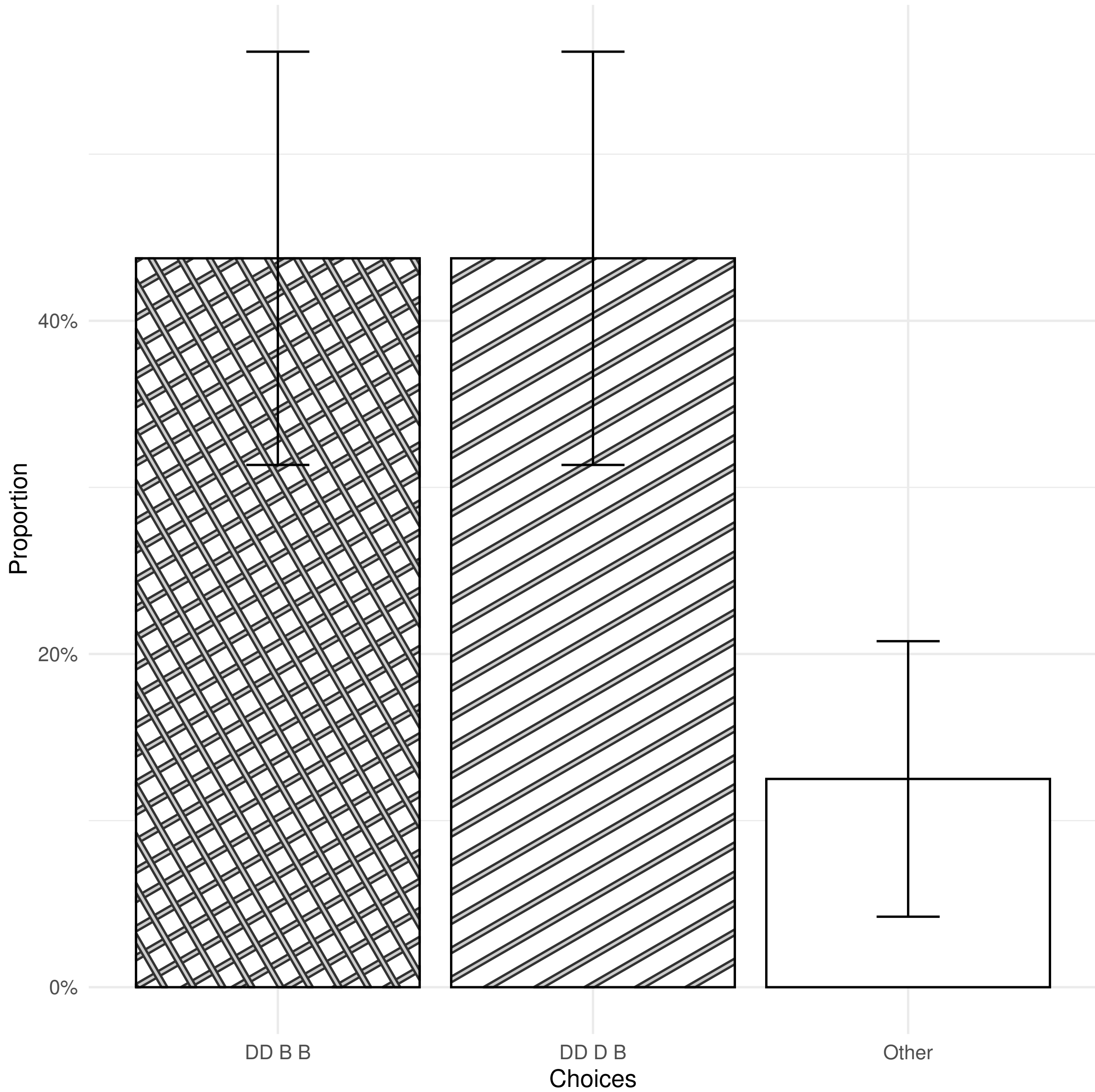


Figure 5: Robustness check

However, we are testing the null hypothesis that subjects pick sectors at random. As this might not be the most plausible null hypothesis, our design allows us to perform another robustness check: There is a strategy that rational players should not play: $DDBD$, which is about as likely to be picked randomly as $DDBB$ and $DDDB$, but which cannot be explained by our model.

Hypothesis (Robustness check)**.** *DDBD is played less often than DDBB and DDDB.*

Using a one-sided binomial test, we have to reject the null hypothesis, i.e., that *DDBD* is played at least as often as *DDBB* and as *DDDB*, if *DDBD* is observed at least 3.15 times (95% CI) or at least 2.24 times (99% CI).

As only two subjects have chosen DD B D, we have to reject this null hypothesis at both levels for both first- and higher-order beliefs.

These results show that both first and higher-order beliefs are reasons for coordination failure. However, coordination on the first best equilibrium does not break down entirely; instead, it remains at a high level.

Remark (Comparison with Blume and Gneezy (2010))**.** *These results agree with Blume and Gneezy's (2010) results, whereby around* 52% *(*58% *in our experiment) have been able to identify the distinct sector, and around* 23% *(*34%*) have chosen the black sector. We attribute the lower level of noise (*8% *vs.* 25%*) to the detailed instructions and the quiz we conducted before the experiment and the fact that the experiment was conducted using a computer.*

Remark (Order effects)**.** *To prevent an order effect (and, therefore, "Eureka" learning), we explained all three treatments simultaneously before asking the subjects to make choices. Then, choices were made in a random order. We did not find any order effect among the subjects that was relevant with respect to our hypotheses, but we found potential order effects among the subjects categorized as "Noise" players (when not controlling for multiple testing). An extensive discussion of order effects can be found in the Online Appendix.*

5. Welfare effects

We have seen that around 18% of the high-cognition subjects fail to coordinate on the best outcome due to their first- or higher-order beliefs. This has a large effect on the individual earnings. In Table 2, we can see a comparison between the expected winning/coordination rates and expected earnings based on the most important strategies played and the actual distribution of strategies played in the experiment. For example, if a subject would have switched from playing B to D in the Coordination Treatment, her chances of winning would be more than twice as high (21.15% $\rightarrow$ 53.08%).

What about the "welfare effects" of coordination failure due to first- and higher-order beliefs in our experiment? The expected coordination rate can describe the welfare in a pure coordination game. The expected coordination

Strategy	Self Treatment	Guessing Treatment	Coordination Treatment	Expected Earnings
Random	20%	20%	20%	£3
BB B B	20%	17.69%	21.15%	£4.44
DD B B	100%	17.69%	21.15%	£6.94
DD D B	100%	58.08%	21.15%	£8.96
DD D D	100%	58.08%	53.08%	£10.56

Table 2: Expected winning rates and monetary outcomes (rounded)

rate derived from the choice data from the experiment is 37.23% (53.08% for high types). To get an idea of the welfare losses due to first- and higher-order beliefs, we compare the expected coordination rate from the experiment with the expected coordination rate in the (hypothetical) situation without coordination failure due to first- and higher-order beliefs. In other words, we ask how the coordination rate would change if all subjects who played either $DDBB$ or $DDDB$ played D instead of B in the Coordination Treatment. Avoiding coordination failure due to first- and higher-order beliefs would increase the expected coordination rate to 45.84% (63.85% for high types), i.e., by 31.18% or 11.61 percentage points (20, 29% or 10.77).

6. Concluding remarks

This paper presents the first example of an investigation carried out to determine the role of first- and higher-order beliefs in coordination failure using a choice-based elicitation approach. We have shown that, even if the game does not provide rewards or evoke higher-order beliefs, both first-order (9.33%) and higher-order beliefs (9.33%) about the type of the other player can cause coordination failure. However, a significant fraction of the subjects (78%) were still able to coordinate with other subjects, despite the absence of common knowledge.

This indicates that not only first- but also higher-order beliefs might be an underestimated source of coordination failure, even in situations where the absence of common knowledge is not apparent. Our results suggest that it is a good reason to carefully evaluate the validity of the common knowledge assumption in game theoretic models and provide additional support for robust game theory mechanism design. Furthermore, our results enable us to take the first step toward conducting a more thorough empirical investigation of the structure of belief hierarchies.

Finally, the result that higher-order beliefs might cause coordination failure also has implications for preventive measures against traditional coordination problems, such as bank runs, speculative attacks, or arms races. This result might also explain differences in success observed in attempts to prevent dangerous practices, such as foot binding in China and female genital mutilation in Sub-Saharan Africa, which we discuss in more detail in the Online Appendix.

Appendix A The model

Using our auxiliary assumptions, we can write our third treatment (the coordination game) as a Bayesian coordination game in which both players try to choose the same sector on the 5-sector disc. In this game, we have two players, one each of the low type or the high type. The difference between these types is that they face different symmetry constraints.[12] Low-type players can only distinguish between white and black sectors, i.e., they have two attainable actions. High-type players can distinguish between black sectors, the distinct white sector, and the two other white sectors, i.e., they have three attainable actions. Furthermore, low-type players have no knowledge of the distinct sector or the existence of the high type.

If both players have chosen the same sector, they receive a payoff normalized to 1, and they receive 0 if they do not manage to choose the same sector. As we have only two possible outcomes, we can disregard the players' attitudes toward risk and define the players' payoffs as the expected chances of choosing the same sector.

In this case, two low-type players face the symmetric coordination problem as shown in Appendix A. If both players choose one of the white sectors, they have a $\frac{1}{3}$ probability of choosing the same white sector. Therefore, the expected payoff associated with both players for choosing a white sector is $\frac{1}{3}$ and $\frac{1}{2}$ for choosing a black sector.

	W	B
W	$\frac{1}{3}$	0
B	0	$\frac{1}{2}$

Table 3: Low types' payoffs

This game has two Nash equilibria in pure strategies and one Nash equilibrium in mixed strategies. However, (B, B) is this game's payoff dominant equilibrium and focal.

To simplify the exposition, we model the low-cognition players as automatons, choosing B with certainty. This behavior is corroborated by the choice data from our experiment in Section 4.1.

Let us now analyze the coordination game as faced by the high-cognition player. Her decision depends on her belief regarding the proportion of high-cognition players. We will denote the actual proportion of high-cognition

[12]For more information on symmetry constraints and attainable strategies, see Blume (2000) and Alós-Ferrer and Kuzmics (2013).

players in the experiment as p.

She has a probability p chance of playing against another high-cognition player. In this case, high cognition players play a game (as the row player) that is depicted by the left-hand matrix as shown in Appendix A. She has a complementary probability of being matched with a low-cognition player. The payoff matrix on the right-hand side depicts a game played by her with a low-cognition "player", i.e., against an automaton that always plays B.

p

	W′	B	D
W′	$\frac{1}{2}$	0	0
B	0	$\frac{1}{2}$	0
D	0	0	1

1-p

	B
W′	0
B	$\frac{1}{2}$
D	0

Table 4: Coordination game - high-cognition player

Her behavior, therefore, depends on p, which is not common knowledge.

If she believes that $p < \frac{1}{3}$, B is a strictly dominant strategy for every high-cognition player, and the unique Nash equilibrium of the game is (B, B).

However, even if she knows that $p > \frac{1}{3}$ but believes that the other player believes p is below $\frac{1}{3}$, and he will never play D, then she should choose B. Even if she thinks that both think $p > \frac{1}{3}$, if he thinks her type is low, he would play B; therefore, she should play B. This reasoning continues ad infinitum; therefore, her decision depends on her entire belief hierarchy about p.

To get a better idea of the influence of the structure of the higher-order beliefs, we compare the results of the two most popular (and most frequently used) models: Common knowledge and the structure underlying the global games literature.

If we assume common knowledge of $p = 58\%$ (the percentage of high-cognition subjects in the experiment) in the Bayesian game described above (Appendix A), we can see that three Nash equilibria in pure strategies occur: (W', W'), (B, B), (D, D). However, (D, D) is payoff-dominant and represents the focal equilibrium. Therefore, the prediction made when assuming that the *distribution of types is common knowledge* is that we should expect full cooperation when p is high enough.

On the other hand, the Bayesian game can be rewritten as an incomplete information game for the players of the high-type. In this case, p is the belief about the type of game that the player is in, and the game is an example of the classical symmetric binary action global game (based on Morris and Shin (2001) or Carlsson and Van Damme (1993)). Therefore, the surviving equilibrium is the worse (B, B) equilibrium.

References

Carlos Alós-Ferrer and Christoph Kuzmics. Hidden symmetries and focal points. *Journal of Economic Theory*, 148(1):226–258, 2013.

George-Marios Angeletos and Chen Lian. Forward guidance without common knowledge. *American Economic Review*, 108(9):2477–2512, 2018.

Dirk Bergemann and Stephen Morris. Robust mechanism design. *Econometrica*, pages 1771–1813, 2005.

Andreas Blume. Coordination and learning with a partial language. *Journal of Economic Theory*, 95(1):1–36, 2000.

Andreas Blume and Uri Gneezy. An experimental investigation of optimal learning in coordination games. *Journal of Economic Theory*, 90(1):161–172, 2000.

Andreas Blume and Uri Gneezy. Cognitive forward induction and coordination without common knowledge: An experimental study. *Games and Economic Behavior*, 68(2):488–511, 2010.

Hans Carlsson and Eric Van Damme. Global games and equilibrium selection. *Econometrica*, pages 989–1018, 1993.

Michael Suk-Young Chwe. *Rational ritual: Culture, coordination, and common knowledge*. Princeton University Press, 2013.

Vincent P Crawford and Hans Haller. Learning how to cooperate: Optimal play in repeated coordination games. *Econometrica*, pages 571–595, 1990.

Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, 2007.

Simon Gächter and Elke Renner. The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, 13(3):364–377, 2010.

Frank Heinemann, Rosemarie Nagel, and Peter Ockenfels. The theory of global games on test: Experimental analysis of coordination games with public and private information. *Econometrica*, 72(5):1583–1599, 2004.

Terri Kneeland. Identifying higher-order rationality. *Econometrica*, 83(5): 2065–2079, 2015.

Stephen Morris and Hyun Song Shin. Global games: Theory and applications. *Cowles Foundation discussion paper*, 2001.

Stephen Morris, Hyun Song Shin, and Muhamet Yildiz. Common belief foundations of global games. *Journal of Economic Theory*, 163:826–848, 2016.

Michal Szkup and Isabel Trevino. Sentiments, strategic uncertainty, and information structures in coordination games. *Games and Economic Behavior*, 124:534–553, 2020.

Jonathan Weinstein and Muhamet Yildiz. A structure theorem for rationalizability with application to robust predictions of refinements. *Econometrica*, 75(2):365–400, 2007.