Identifying the reasons for coordination failure in a laboratory experiment*

Online Appendix

Philipp Külpmann† Davit Khantadze‡

September 24, 2022

This is the Online Appendix for Külpmann and Khantadze (2016). Section 1 discusses order effects, in Section 2 you can find more detailed derivation of theory results. In Section 3 you can find the instructions (Section 3.1) and screenshots of the control questions (Section 3.2). Finally, in Section 4 you can find a description of the application in the fight against female genital mutilation.

For most recent versions of the paper and this online appendix, the code of the experiment and the statistical analysis and the data in a machine readable format, please see `http://hob.kuelpmann.org`.

1 Order effects

As this is a robustness check, we have not corrected for multiple hypotheses testing.

The two main effects of order effects are that (a) the order of treatments had no effect on our main results and (b) the only significant order effect is that, that the number of mistakes in the Self is getting smaller, if the treatment is conducted later.

The most important order effect, we expected in this experiment was "Eureka!" learning, i.e., "Having a player play against himself may trigger an insight that switches a

*This is the Online Appendix to Külpmann and Khantadze (2016).

†Vienna Center for Experimental Economics, `philipp.kuelpmann@univie.ac.at`
‡Department of Economics, University of Warwick, `d.khantadze@warwick.ac.uk`

player from low to high cognition ("Eureka!" learning). There may be an uncertainly principle at work here in that we cannot measure a player's cognition without altering it." (Blume and Gneezy (2010)).

Thus, we implemented a random ordering of treatments on the subject level, to be able to control for this.

First of all, let us have a look at (a), i.e., the possible effects of the order of treatments on our main results. For certain orders of treatments, "Eureka!" learning might have been misidentified for either first- or higher-order beliefs.

If "Eureka" learning occurs between the Prediction Treatment and the Coordination Treatment, we would misidentify this person as someone with higher-order believe problems (i.e., she would play DD D B). However, this is not a problem, as, of our 7 higher-order believe subjects, 4 played the Coordination Treatment before the Prediction Treatment (compared to 50.1% of all subjects).

Misidentifying can a subject as having a first-order belief problem (DD B B) due to an order effect can only happen if the Self Treatment is played last. Of the 7 subjects who we identify as having a first-order belief problem only 3 subjects played the Self Treatment last (compared to 40.1% of all subjects - this is higher than the expected 33% due to the unbalanced randomization).[1]

Therefore, we believe that order effects had no significant effect on the main results.

However, having a look at the general order effects might be interesting nonetheless. Therefore, we had a look at every possible order effect and tested every possible of the 27 combinations (3 treatments, 3 choices each and 3 combinations of orderings for the three treatments).

We report the (rounded) p-values of a two-sided Fisher's test in Table 1. In it, we test if there are significant differences in the frequency of a certain action (e.g. picking the distinct sector D), when it is played first against when it is played second (and every other combination).[2]

Let's have a closer look at every treatment:

[1] To see the complete data, the calculations and some additional robustness checks, refer to the Online Appendix and the R code in "OrderEffects.R" (`http://hob.kuelpmann.org`).

[2] In the Self Treatment, we have only distinguished between DD, BB and everything else, as everything else was rare enough (11 subjects overall). In the following we consider "Other" as subjects making a mistake of some kind.

Treatment	Self			Prediction			Coordination		
Order	**DD**	**BB**	**Other**	**D**	**B**	**W**	**D**	**B**	**W**
2nd and 3rd	0.40	0.13	0.08	1.00	0.65	0.44	0.25	0.16	0.39
1st and 3rd	0.67	0.18	0.0004	0.29	0.66	0.26	0.02	0.10	0.39
1st and 2nd	0.17	1.00	0.10	0.27	0.26	1.00	0.16	0.84	0.02

Table 1: Order effects of the different treatments

In the self treatment, one row stands out: The people who have chosen something "Other", i.e., the people who made a mistake. Here, we can see that there is a significant order effect, especially between the Self Treatment being played first and last. The direction is as expected: People are getting better at playing against themselves and, the effect is the strongest if we compare the self treatment being played first and last.

Apart from this, it seems that there are no order effects in the Self Treatment.[3]

In the Prediction treatment, we don't see any significant order effects in either direction.

In the Coordination treatment, we see three two effects: Comparing the 1st and 3rd round, we see significantly more subjects playing D and less B in the 3rd round than in the 1st round. However, the difference between the 1st and 2nd and the 2nd and 3rd is not significant.

Furthermore, a see a significant lower amount of subjects playing W if the Treatment is played first than second.

Thus, while we see some "Eureka" learning effect in the Self Treatment, the "Eureka" learning effect vanishes in almost every other treatment.[4]

Why did Blume and Gneezy (2010) encounter strong "Eureka!"-learning effects whereas we had (almost) no significant effect. We attribute this to the fact that the participants were instructed in all three treatments before they played the first game which most likely triggered the learning before the first decision, whereas in Blume and Gneezy (2010) the instructions for the second treatment were distributed after completion of the first treatment. Furthermore, we used more extensive instructions and a quiz to make sure the instructions where understood.

[3]That we don't see any significant change due to subjects moving from the "Other" group into one of the other two groups is not surprising, as only 11 out of 130 subjects have been clasified as "Other"in the Self Treatment.

[4]We do not want to claim the existence or non-existence of any order effects here, as we clearly not have enough data to show non-existence and a p-value of 0.02 in two of 27 tests is no sufficient basis for the existence. The only thing we claim is that there is an order effect in the Self Treatment (less mistakes later on) and, if there is an order effect anywhere else, it is not particularly strong.

2 Theory

2.1 Formal definition of belief hierarchies

Let $B_i^0 := T_j$ and $B_i^k = T_j \times \Delta(B_i^{k-1})$ with $\Delta(B)$ being the space of probability measures on B and $\Delta(X)$ being the space of probability measures on the Borel field of X, endowed with the weak topology. Using this notation, we can define a belief hierarchy as follows.

Definition 1 (Belief hierarchy)**.** A *k-th order belief* is defined as

$$b_i^k \in \Delta(B_i^k)$$

with $B_i^0 = T_j$ and $B_j^k = T_j \times \Delta(B_j^{k-1})$
Furthermore, let us set $b_i^0 := t_i$.
A belief hierarchy of a player i is then $b = \{b_i^0, b_i^1, \ldots.\}$

We therefore have a first order belief $b_i^1 \in \Delta(\{low, high\}) = [0,1]$ and higher-order beliefs $b_i^k \in [0,1]^k$.

Furthermore, we assume these beliefs to be coherent, i.e. that beliefs of different orders do not contradict one another, and that a low-cognition type does not know about higher cognitive types, i.e., $b_i^k = 0 \Rightarrow b_i^{k+1} = 0 \quad \forall k \geq 0$.[5]

This excludes, on the one hand, that a low-cognition player thinks that the other player is a high-cognition player and, on the other hand, that a player has a first-order belief that the other player is of a the high type and a higher-order belief that the player is of the low type.

3 Experiment

3.1 Instructions

Welcome to this experiment in economic decision making. It will take approximately 60 minutes. First of all, please check that the number on the card handed to you matches the number on the cubicle that you are seated in and that your mobile phones are turned off.

[5] I.e., higher-order beliefs of a player mapped onto the space of beliefs of a lower order are the same.

Before we start, we will explain the rules of this experiment. You will also find these rules on the paper provided, so you can read along and check again during the experiment. If you have any questions, please do not speak up but raise your hand and we will come to you and answer your question privately.

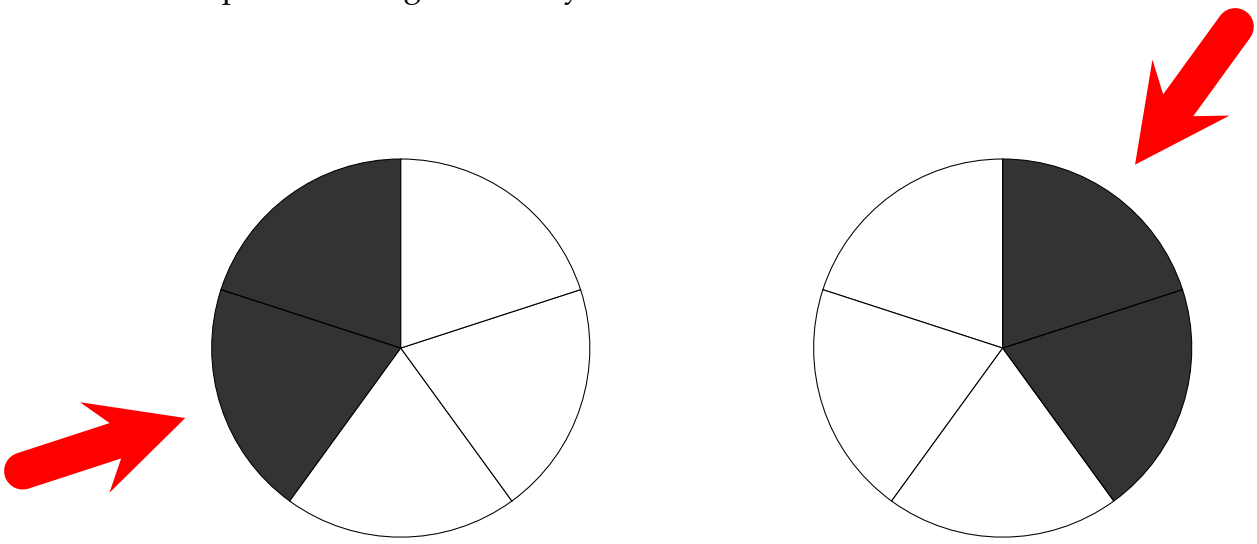
From now on, please do not talk, and listen carefully. In this experiment you will earn a minimum of £3, and potentially up to £18. How much money you earn will depend on your decisions and those of the other participants. Your reward will be paid out at the end of the experiment. None of the other participants will know how much money you made.

In this experiment you will be asked to make decisions related to a disc that has 5 sectors, similar to the disc provided to you. The disc has two identical sides. Your goal will be to pick the same sector twice (more on that later). During this experiment the disc will be flipped and/or rotated randomly.

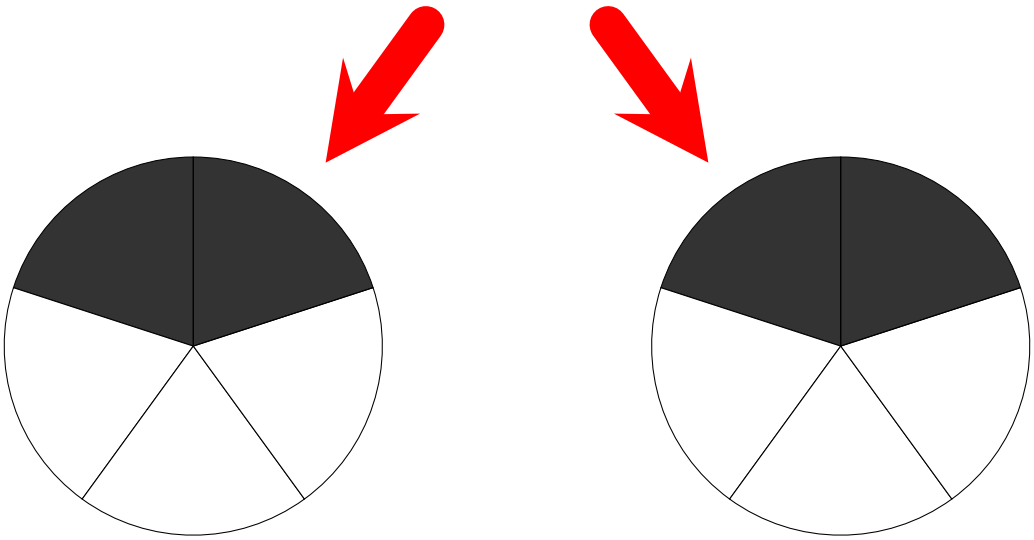
Pictures on page 2 illustrate rotation and flipping. Since you will not be told if the disc was flipped and/or rotated, it might even be the case that disc looks exactly the same though sectors have changed their positions.

The arrow tracks one specific sector that changes its position as the disc is rotated and/or flipped.

This is an example of rotating the disc by two sectors:



This is an example of flipping the disc:

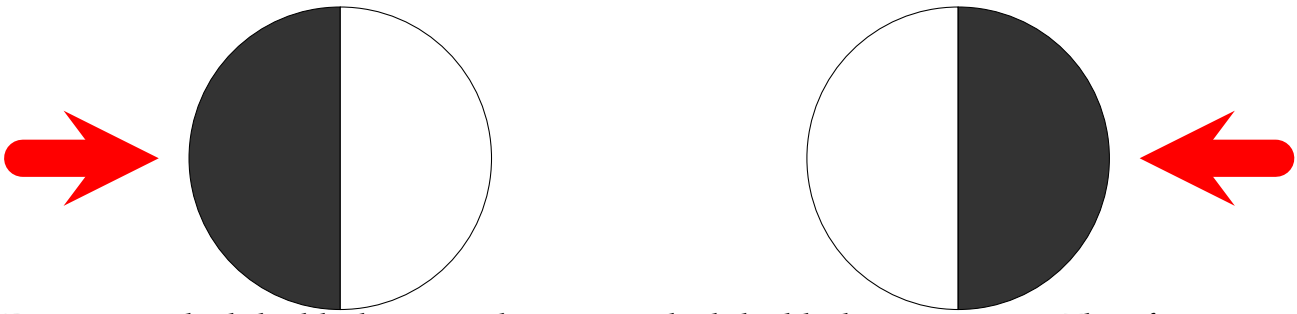


In the experiment the disc will be surrounded by the letters A, B, C, D, and E. **These labels are not part of the disc!** They are only included to allow you to choose a sector.

In the experiment you will make decisions in the following environments (the order will be chosen randomly):

(**Self Game**) You will be asked to pick a sector twice; first you choose a sector; then the disc might be flipped and/or rotated. After this you are shown the same disc and have to choose a sector again. You will not observe the flipping/rotation of the disc. If you manage to guess the same sector twice, your payoff will be £5. Otherwise, you will receive 0. Therefore, to earn more money you want to maximise your chances to pick the same sector twice.

Here is an example of the choices made in a **Self Game**, using a simpler disc with only 2 instead of 5 sectors:

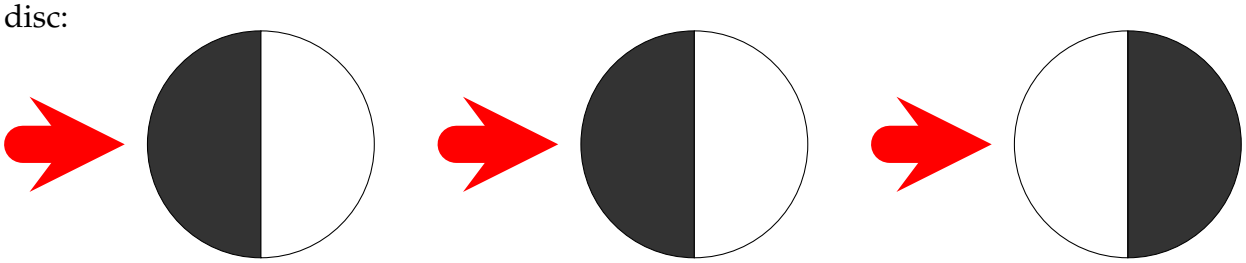


First you picked the black sector; then you picked the black sector again. Therefore, you pick the same sector twice and earn £5.

(**Prediction Game**) You are matched randomly with another person and you have to guess the choice of this person, while she plays the **Self Game**. First, you choose a sector on the disc; each time the other person picks the sector you chose, you will

receive £2.5. As the other player picks twice in the **Self Game**, you can earn £0, £2.5 or £5 in this situation, depending on your and the other person's choice. Therefore, to earn more money you want to guess what the other player is playing in the **Self Game** described above.

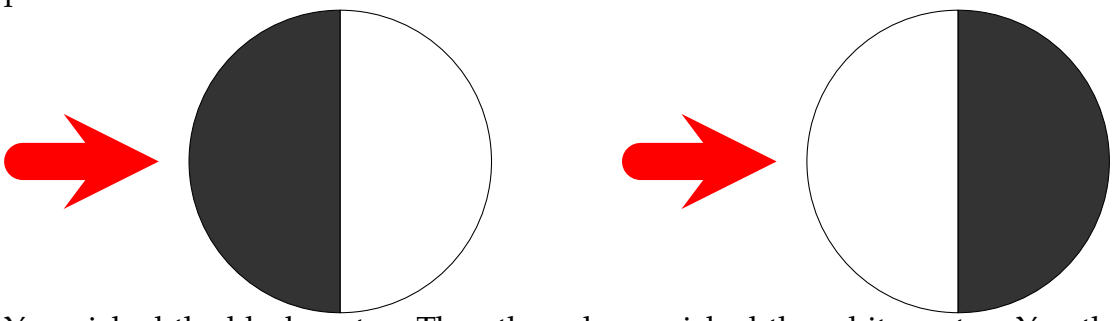
Here is an example of the choices made in a **Prediction Game**, again with the simpler disc:



First you picked the black sector. The other player then plays the **Self Game**. He first picks the black sector and therefore you earn £2.5. Then he picks the white sector and therefore you earn £0. Thus you earn £2.5 in total.

(**Coordination Game**) You are matched randomly with another person and both of you are asked to pick a sector on the disc simultaneously. Both of you know that you play the **Coordination Game**. You both see the same disc but possibly differently flipped and rotated. If both of you pick the same sector, then your payoff will be £5. Otherwise, you will receive £0. Therefore, to earn more money you want to guess the sector your partner is picking here, while he is trying to do the same.

Here is an example of the choices made in a **Coordination Game**, again with the simpler disc.



You picked the black sector. The other player picked the white sector. You therefore failed to coordinate and both of you earn £5 each.

The experiment consists of two periods. Each period consists of the three games as described above, using a 5-sector disc; the order of the games is random. At the end of the experiment one of the two periods will be randomly chosen. The earnings made in this period will be paid out in cash.

Again, please do not talk during this experiment! If you have questions just raise your hand.

Before the experiment there will be a quiz to check your understanding. Read hints carefully if you get stuck during the quiz.

3.2 Control questions

In this appendix you can find screenshots of the control questions which was conducted before the experiment. Participants who made a mistake in some part of the quiz were given a small hint and then were asked to repeat this part of the quiz.

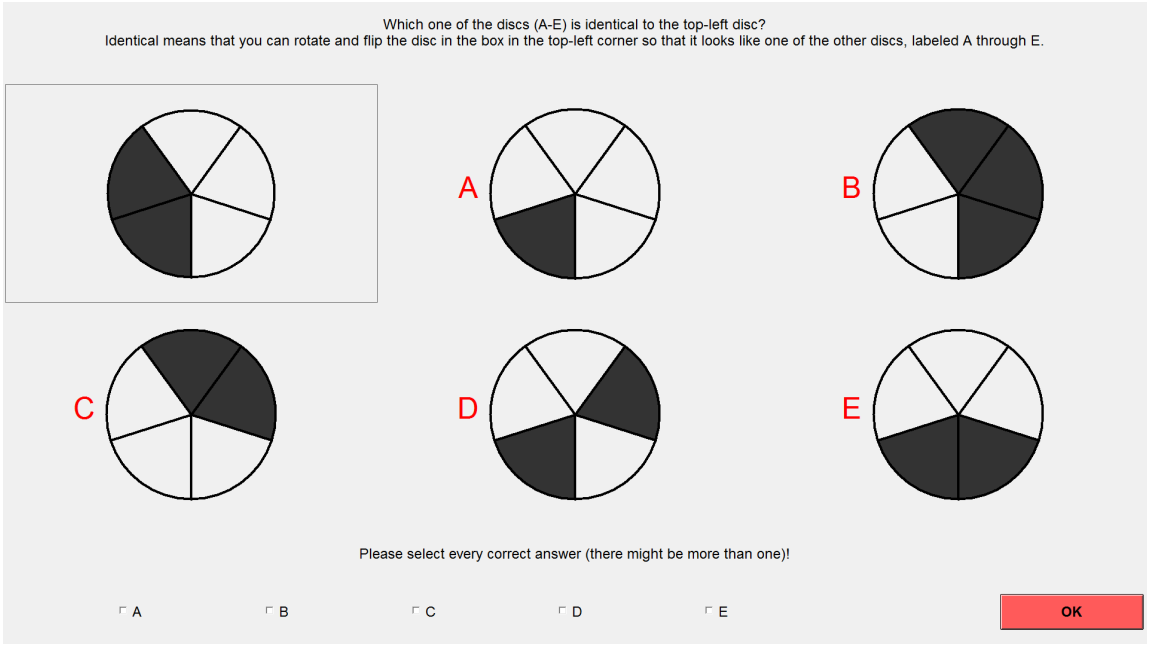


Figure 1: Quiz part 1

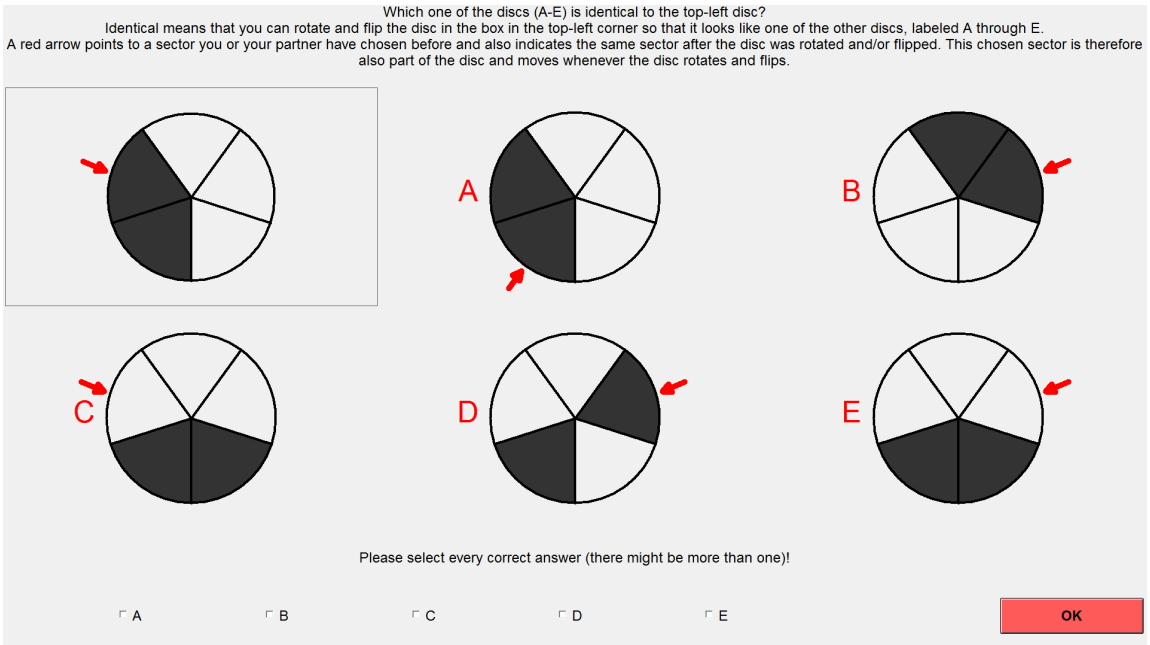


Figure 2: Quiz part 2

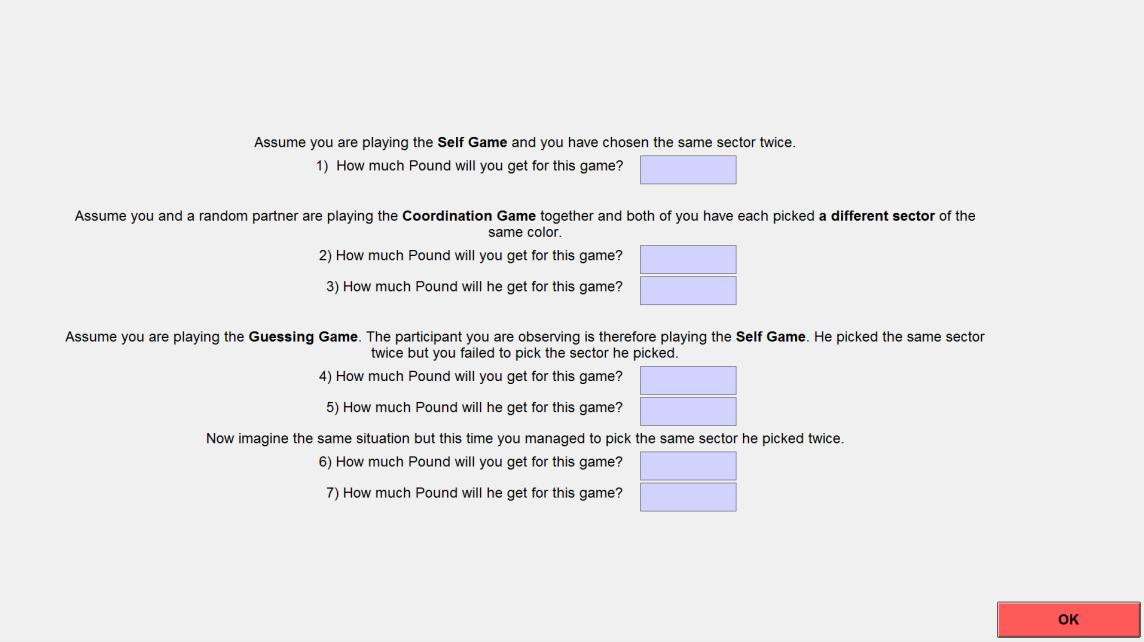


Figure 3: Quiz part 3

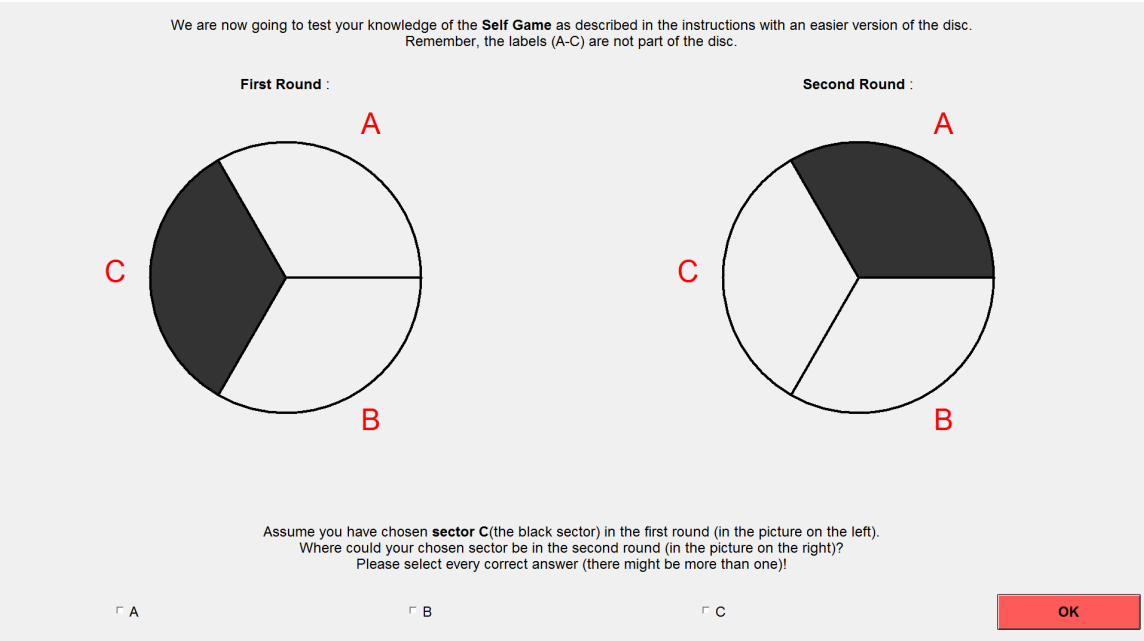


Figure 4: Quiz part 4

4 The role of beliefs in the fight against female genital mutilation

Now, let move from "welfare effects" in the lab to a real world application with very strong welfare effects. Female genital mutilation (FGM) is a wide spread problem in many parts of the world and it is estimated to effect up to 200 million women in 2016

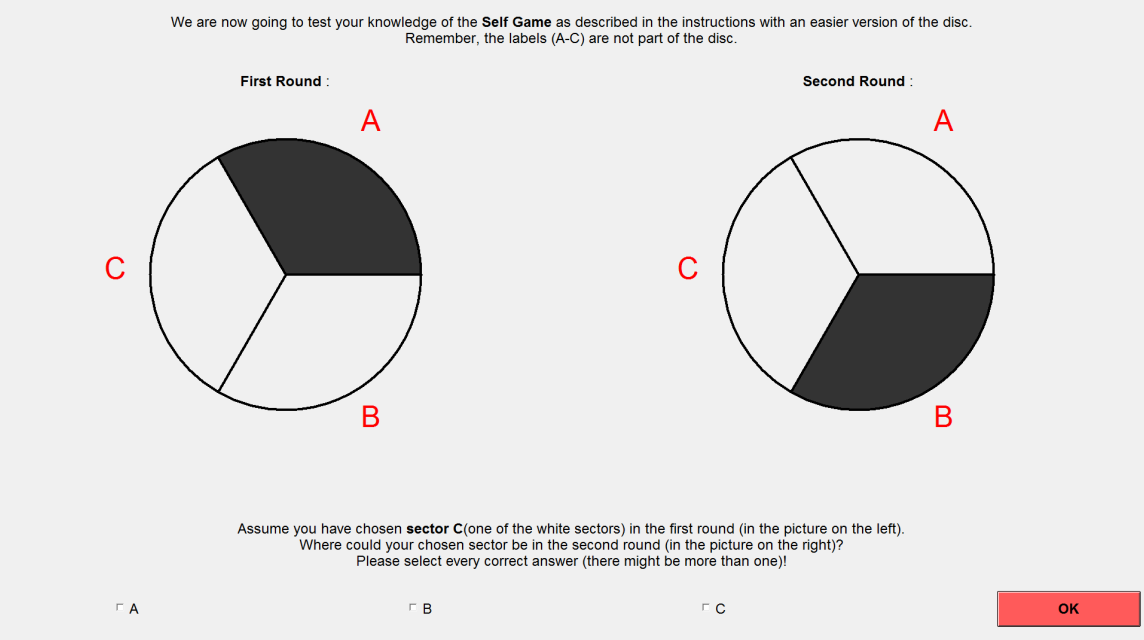


Figure 5: Quiz part 5

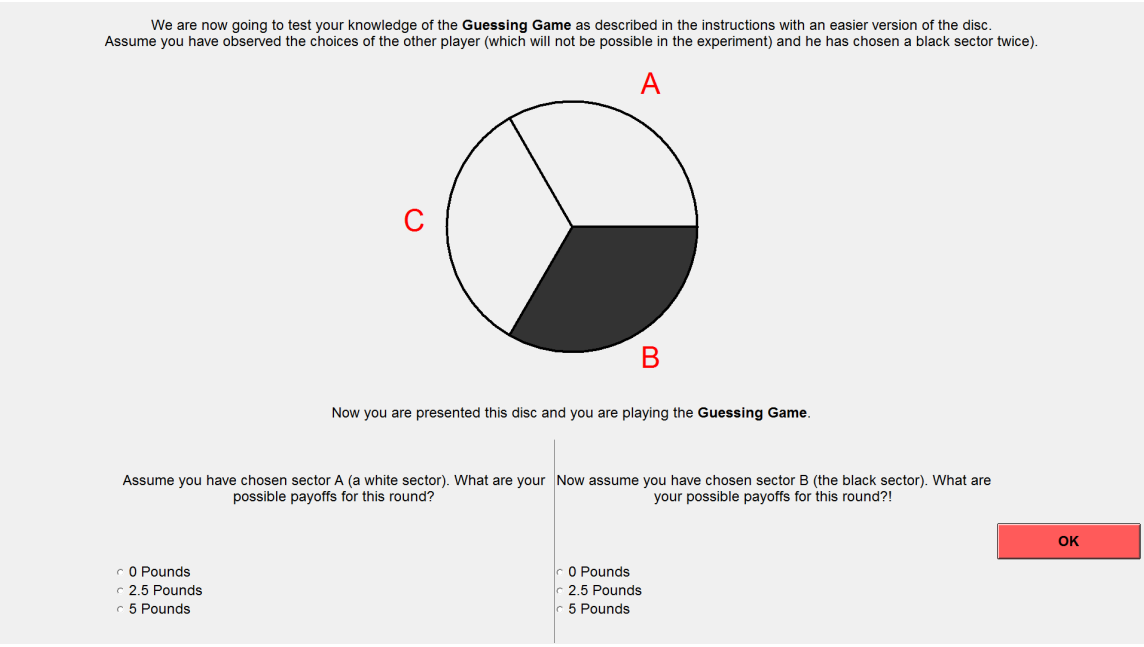


Figure 6: Quiz part 6

(UNICEF (2016)).[6]

From the literature mentioned above, it seems that the problem seems to be two-fold: Strong beliefs in the benefits of FGM (e.g., hygiene issues, pleasure of the man, ...) and, even among families who do not think it is beneficial, peer pressure.

In the last couple of years, different NGOs and governments have fought against FGM

[6]For more information on FGM see for example Gupta (2013) and Bicchieri (2005).

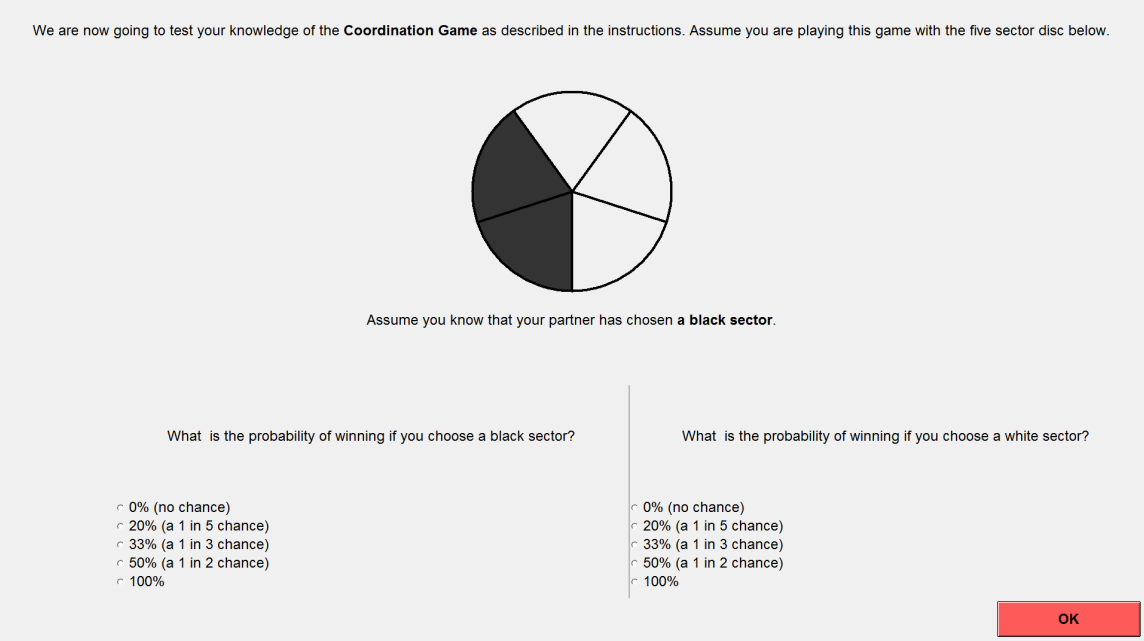


Figure 7: Quiz part 7

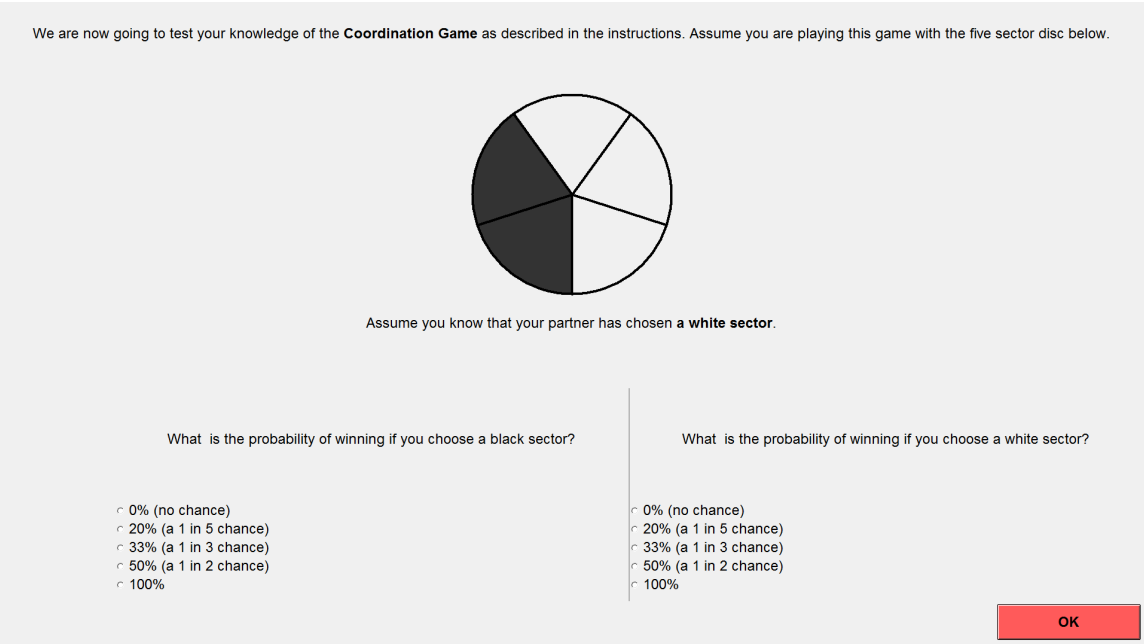


Figure 8: Quiz part 8

by educating families about the dangers of it. However, it was commonly observed that, while informing families changes their opinion about FGM, the effect in behavior (i.e., not infibulating their daughters) is much smaller than the change in believes (see for example, Bicchieri (2005)).

One possible reason for that might be higher-order beliefs: If I know FGM is bad, but enough other women believe in it, my daughter will not get married if she is not infibulated. Or, even worse, even if everyone knows FGM is bad, but everyone things,

that everyone else thinks that infibulation is good, I will still infibulate my daughter.

Here, like in our experiment, coordination on the better "No-FGM" equilibrium could therefore be hindered by pessimistic first- or higher-order beliefs. Even if everyone thinks FGM is bad but thinks that every other family expects it, every family would choose to infibulate her daughters.

The link from this (simplified) problem to our experiment is clear: In both, we have a bad ("FGM", BB) and a good equilibrium ("no-FGM" and DD) and the latter can only be played by "informed" or "educated" players (i.e., families who know about the dangers of FGM/players who see the distinct sector). Unfortunately, the lack of common knowledge makes the problems much more complicated and there is a chance that even informed players choose the bad equilibrium due to 1st or higher-order believes.

We have learned from our experiment that, even in simple coordination game in which most subjects were of the "educated" type (i.e., the ones who were able to find the distinct sector), some educated subjects switched to the bad equilibrium because of first- and higher-order believes.

Therefore, it is reasonable to assume that this result might generalize to more general populations and that believes might play a role in the fight against FGM.[7]

What can we now learn from these results for the fight against FGM?

First of all, making the education common knowledge, e.g., by gathering women from the area and explaining everything to everyone at the same time. [8]

Unfortunately, this is not always possible and it might be prohibitively costly. However, due to the structure of the problem (i.e., that the decision for the daughters and sons of the family are separable), we can replicate the Guessing Treament in our experiment. This would reduce the problem from a higher-order belief problem to a first-order belief problem, which is much easier to handle.

To do this, one has to convince every family to (A) not only not infibulate their daughters but also to marry their sons to not infibulated women. And (B) make every family

[7]It might be far-fetched to extrapolate the behavior of students at a British university to villagers in rural Africa or South-East Asia. However, we believe that our results are a good enough reason to either test these in a field experiment or, as an experiment itself, implement the suggestions below into existing programs, as they are very cheap.

[8]This is done by Tostan, an NGO, which emphasizes that education together with public discussion and public declaration is the best way to fight FGM (Tostan (2016)). Their claims are supported by their success rates, which are reported for example in World Bank Group (2012)). While there are probably also other factors at work (,e.g., group dynamics, ...), this also points towards the possible importance of higher-order beliefs.

sign a letter of intent that they will marry their sons also to uninfibulated women and will not infibulate their daughters and distribute these letters afterwards to every family.

Then, due to (A), the structure is very similar to our Guessing Treament: now, the decision only depends on the type (i.e., has the other family accepted that FGM is bad) and not on the action of other families (i.e., will they let their son marry my daughter, even if she isn't infibulated). Due to (B) we have made sure that the everyone knows the first-order believe of everyone else.

This inexpensive and simple addition might help to boost success rates of the widely used education approach in the fight against FGM more efficiently.[9]

References

Cristina Bicchieri. *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press, 2005.

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.

Geeta Rao Gupta. Female genital mutilation/cutting: a statistical overview and exploration of the dynamics of change. *Reproductive Health Matters*, 21(42):184–191, 2013.

Philipp Külpmann and Davit Khantadze. Identifying the reasons for coordination failure in a laboratory experiment. 2016.

Tostan. Female genital cutting. `http://www.tostan.org/female-genital-cutting`, 2016. Accessed: 2016-05-21.

UNICEF. Female genital mutilation/cutting: a global concern. `http://www.unicef.org/media/files/FGMC_2016_brochure_final_UNICEF_SPREAD.pdf`, 2016.

World Bank Group. *World Development Report 2012: Gender Equality and Development*. World Bank Publications, 2012.

[9] Of course, as mentioned above, some testing in the field is needed.