The False Beliefs of Women -

How Women Believe Their Male Counterparts to Be Better Than Themselves

Christoph Stumm

(Christoph.Stumm@rub.de)

January 2012

University of Passau

Abstract: By conducting a P-Beauty Contest while monitoring player's beliefs, the gender sensitivity of men and women is investigated. The evidence suggests that women believe men to be better in the field of iterated reasoning, even though the contest itself reveals this not to be the case. Men on the other hand seem to be free from gender prejudices and also show signs of diminished "male overconfidence".

1. Introduction

Over the last few years a huge amount of literature has emerged, describing various differences between men and women, which might explain the current gender gap in the labor market. One explanation for the observed differences is that both sexes tend to have different attitudes toward psychological attributes as described by Bertrand (2011). When examining gender differences, it becomes apparent that primarily women tend to be gender sensitive, i.e. they treat men differently from how they treat other women. Gneezy et. al. (2003) for example noted that a woman's performance in competition based experiments is increased when playing against other women, while it is typically decreased when playing against men. Another example comes from Ben-Ner et al. (2004), describing that women in a dictator game tend to give less to other women and more to men. Croson and Gneezy (2009) also noted that in an ultimatum game women do accept offers more often when coming from men. This of course raises the question of the importance of the gender composition of the environment. Bertrand (2011) explains the existence of the gender sensitivity partly with different psychological attributes. One of her arguments is the existence of the "Gender Identity", i.e. the typical gender role, or what could be called "expected actions from men/women". Reading her paper and looking further into the field of education also provides the notion that women expect men to be better particularly in the fields of mathematics and rational thinking, even though this is not always the case.

I therefore had the hypothesis that women believe men to be better in the field of iterated reasoning, even if they perform as well as men do. My hypothesis can also be put simply by saying: Women believe that men are more capable of applying iterated reasoning than other women. The hypothesis is based on the notion that iterated reasoning is often being associated with mathematical skills at first sight, even though Camerer (2003) notes that this is not the case at all.

I also had the hypothesis of men underestimating their opponents more than women (especially other women, due to the prejudice of women being weaker in terms of mathematical/logical reasoning) do, since extreme overconfidence among men is a common feature of male thinking. In this regard I wanted the beliefs of men to be tested as well. The question to be answered was: Do men overestimate or underestimate women, or do they believe that both genders are equally capable of iterated reasoning?

To examine the differences in the beliefs of iterated reasoning between men and women a

simple P-Beauty Contest-Experiment (often also called the "Moulianian Beauty Contest") with individual belief monitoring was conducted, hoping to give crucial insights about the different beliefs of the genders and their thinking processes.

2. The P-Beauty Contest

The P-Beauty Contest is often described as the perfect experiment to measure the level of iterated reasoning. The rules of the game are as follows: The game requires at least two players, but can theoretically also be played with an infinite amount of players. Before the game starts, the value of p, the multiplier, is given to the players. In this example the value of p shall be 0.6. Each player then has to choose a number between 0 and 100. After every player made his choice the mean of all submissions is multiplied with p (in this case 0.6). The player, who is closest to the mean multiplied with p (=0.6) wins the game.

In theory, where every player is capable of perfect iterated reasoning, each and everyone will pick 0 (Nash equilibrium) to be his/her value, since the potential payoff is maximized at this point. The thinking process behind these outcomes is quite intuitive: Let's assume that Player 1 might think to herself that every player in her group would pick the value 100 and thus implying that every opponent applies no iterated reasoning. Then the average of the group would be 100, which is multiplied with p=0.6, leaving the group with a final value of 60. In this case Player 1 might be tempted to choose 60 to be her submission value (level 1 of iterated reasoning), but only until she realizes that all the other participants might think the same and might also choose a submission value of 60. In this case the final value would be 36 (=60 multiplied with the p value of 0.6). Now again, Player1 might be tempted to submit the value of 36, which could be characterized as "level 2 of iterated reasoning", but only until she again realizes that the other participants might think the same and pick the number 36 as their submission value and thus lower the final value even further. This downward spiral continues until eventually all players will choose the value 0 to be their submission value. Once the value of 0 is reached no player has an incentive to deviate, making 0 the Nash equilibrium of this game.

In the case of p>1 the predicted outcome will be 100, since the dynamic process seen above will lead the players to submit the highest possible value and thus maximizing their payoff. Yet when the beauty contest is being played with individuals, who do not have any experience in game theory, the results are more subtle than predicted by economic theory. The first experiment done by Nagel (1995) found, that the typical player makes 1-2 steps in iterated

reasoning, when playing the game for the first time and having no insights into game theory what so ever.

To confirm my hypothesis it is crucial to know, whether the players really understood the mechanism of the game itself. One can legitimately argue that participants do not understand the game when their submission value is above the highest possible mathematical value. If, like in this case p=0.6, the highest submission value to submit would be 60, which would leave us with the certainty that participants with submission values higher than 60 did not even made one step in reasoning and therefore did not understand the game nor its fundamentals, which is necessary when analyzing player's beliefs. Therefore belief submissions of players, who obviously did not understand the game shall be neglected, since in order to make a credible statement on beliefs, the mechanism at work has to be understood first.

3. Experimental Procedure

The experiment was computerized with the software "z-Tree" (Fischbacher 2007) and conducted at the University of Passau.

Participants played in small groups of four in which at least one player had a different gender than the rest. The beauty contest consisted of 4 rounds, in which the p-value changed from 0.6 in the first two rounds, to 0.3 in round 3 and to 1.3 in the final round. The first round was designed as a "test round" and constructed primarily for the players to get accustomed to the principle of the game. As many researchers have pointed out, one cannot assume that players will understand the game and will play "the right way" when playing for the first time. In this sense, the second round offered a second chance for the contestants, who (hopefully) would have had understood the mechanism of the game by that point.

The p-value of round 3 was chosen to be very low, primarily to sensitize the players to the forces involved in this game. A very low p-value will drastically decrease "the winning value" and thus should lead to an enormous decrease in submission values, conveying the idea to the contestants, that the all determining factor in this game is the value of p itself. The final round had a p-value of > 1, just to show players that the mechanism at work can also pave the way to the opposite direction and making it "a race to the top" instead of "a race to the bottom". Right after the players received their game instructions (Appendix: Screenshot 1) and right before they submitted their value for the actual beauty contest (Screenshot 3), each player had to submit they beliefs about other player's submission values (Screenshot 2).

The monitoring of the player's beliefs was challenging, since players had to know against whom they were playing (male or female), without realizing that the game was about measuring gender effects. Thus, simply giving players the information about the gender of their opponent wasn't an option to begin with, since it might convey the intention of a gender experiment and hence may have altered player's choices or beliefs. Instead, a system was designed, in which all contestants could physically see against whom they were playing, which of course would also provide subliminally the gender of the opponent to each player. In order to form groups and show people who they were playing against, small pennons were installed at each work place. The pennons indicated the group[1] and the number within the group, making it possible for every player to pinpoint which number they were and against whom they were playing. The individual identification was also pointed out to each player at the beginning of the game (Screenshot 2). Thus, the risk of participants mistaking their group and number was deemed to a minimum.

At the beginning of the experiment it was also noted verbally, as well as in written form that it was allowed to (if necessary) turn around and take a closer look at the players they were playing against, which the contestants did. Further, consultations were strictly forbidden and did not occur during the game.

After the players submitted their value, they were only informed about whether they had won the game or not (Screenshot 4&5). Displaying the winner of the round to the players who lost, might have influenced the loser's beliefs in the next round. Yet, the participants were informed about the "winning value", i.e. the average group value multiplied by p, providing some fundamental figures to the players, without giving away easily the most important one: The average of the group.

It was lambasted, that asking the participants for their beliefs, might be similar to giving them hints of how to play the game, since the first step of the reasoning process is to ask yourself, what the other players might pick. It might therefore be likely, that players show a higher level of iterated reasoning than measured for example by Nagel (1995). Also, some of the participants had experience in game theory, which would further lower the average numbers submitted by a reasonable amount. Hints of the existence of both factors lowering the submission value in general, can be found in the appendix in the distribution graphics. An extraordinary amount of players chose very low numbers to begin with, a large proportion of them even chose numbers between 0 and 5 in the first round. Fortunately, having experienced players does not bias the experiment, since it is primarily about beliefs, not about performance.

[1] The groups were indicated by different colors. There were 5 different Groups: Blue, green, red, white and yellow.

4. Results

A total of 56 Player participated in the experiment, consisting of 30 women and 26 men.[2] I looked at the submissions the players made during the whole experiment and sorted out the submission values, as well as the belief values, when it was inevitable that the players had not understood the mechanism of the game (cleansed values). As mentioned above, this was the case, when a player submitted a value, which was above the highest possible mathematical value. The removing of these values can only be applied on the first three rounds, since there is no maximum submission value in round 4, making it impossible to see, whether the player had understood the game by that time. I also neglected the submission and belief value submissions in round 1 or 2 (even though the values were mathematically possible), when a player submitted a value above 30 in the third round, implying that he had not understood the game mechanism up to this point and picked values lower than 60 in the first rounds by pure coincidence.

On the other hand, if a player submitted a technically impossible value in the first round, but submitted a mathematically possible value in the second or third round, the submissions of the second/third round were taken into consideration, suggesting that the player had learned the mechanism of the game.

The number of values, which were taken into consideration for the belief analysis, are listed in table below.

	Round 1	**Round 2**	**Round 3**	**Round 4**
Submissions taken into consideration women	23	26	28	30
Beliefs taken into consideration women	69	78	84	90
Submissions taken into consideration men	19	22	22	26
Beliefs taken into consideration men	57	66	66	78
Submissions taken into consideration total	42	48	50	56
Beliefs taken into consideration total	126	144	150	168

Table 1 - Cleansed data set

[2] For the University of Passau the amount of men participating was unwontedly high. The high participation rate can be explained with the recruitment process in lectures, in which it was stated that: “In the last years we had a shortage of men. In this sense, men are also very welcome to join the experiment this year”.

The beliefs (based on the cleansed data pool) of men and women are displayed in figure 1. The notion is as following: The first letter depicts the gender of the believer, while the second letter depicts the person, of whom the belief is about. In this sense "Belief M – W" displays what men believe women to choose as their submission value. "Belief W – M" on the other hand displays the value women believe men to pick, etc. Figure 1 indicates that the hypothesis of women being more gender sensitive seems to be legitimate, while the hypothesis about the beliefs of men can be dismissed. Men being extremely overconfident or being gender sensitive cannot be verified, since no clear tendency is observable.[3]

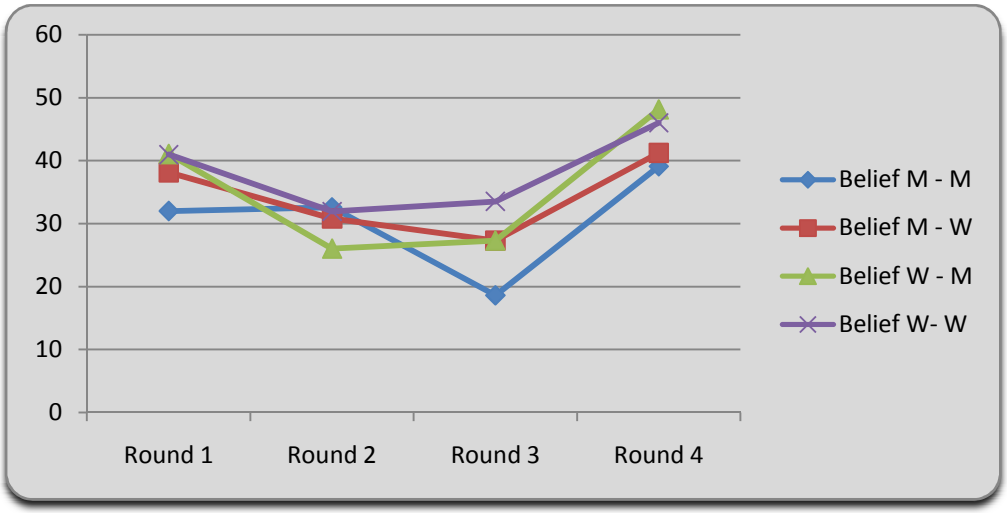


Figure 1 - Beliefs of men and women (mean)

From now on, I therefore concentrate on my first hypothesis: For a better overview, the beliefs of women are once again displayed in more detail in figure 2. Women, so it seems, believe men to submit lower values than other women and hence assuming men to be better when it comes to iterated reasoning. The first round builds an exception to this rule, but as previously mentioned also represents a test round, which hence does not necessarily neglect the trend observed.[4]

In Round 2 and 3 the beliefs about men are well beneath the beliefs about other women, while in round 4 this trend is reversed (Reminder: in round 4 a higher value is associated with a higher level of reasoning since p=1,3).

[3] Assuming that men are equally capable of iterated reasoning, overconfidence among men would manifest in higher belief values. This cannot be observed.

[4] Participants might have been "surprised" by the belief monitoring, since it was not announced in the game description. When they were asked for their beliefs again in the following rounds, the players might have had more time to think about their actual beliefs and might have stated a more accurate belief.

Even though the statics are cleansed, extreme values above 60 are wiped out of the data pool and no real difference between the mean and median can be expected, the median is depicted in figure 10 (Appendix). Both figures show an identical pattern, indicating a uniform distribution of beliefs.

Women, so it seems, tend indeed to believe that men are better in the field of iterated reasoning, even though they still underestimate male iterated thinking capabilities (with the exception of round 2).

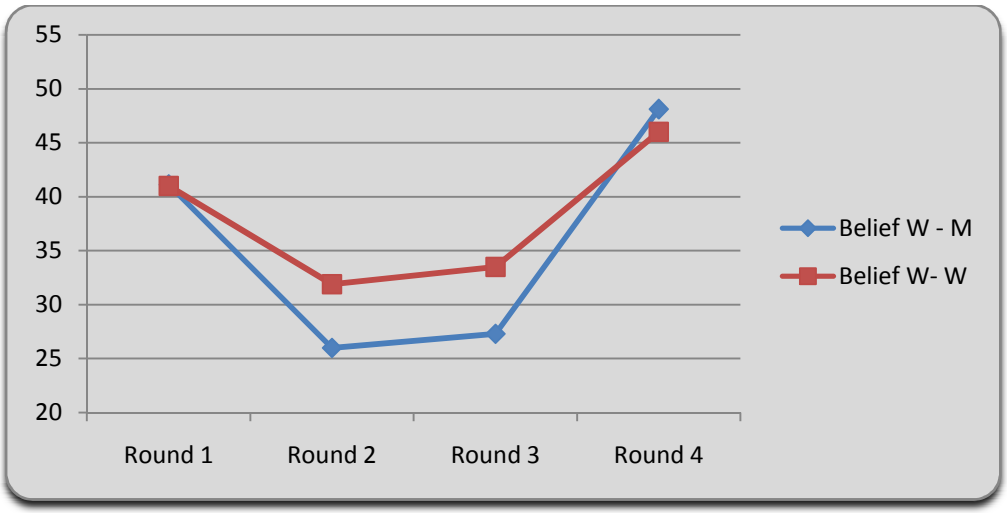


Figure 2 - Beliefs of women (mean)

But do men really do better than women? The answer is a clear "no". The results of the game are depicted in table 2. It might be added though, that the overall amount of wins per round is not 14 in every round, as one might suspect. This is due to the existence of more than one winner per round, in case more than one player submitted the same winning-value. Table 2 also depicts the “double wins” of each gender, giving a better, unbiased overview.

	Round 1	Round 2	Round 3	Round4
Men	4	7	5	7
Women	11	9	9	8
Double Win Men	-	-	-	-
Double Win Women	1	1	-	-
Double Win mixed (M&W)	-	1	-	1

Table 2 - Game results

Looking closer, table 2 gives the impression of women actually performing better than their male counterparts, even if one takes the higher amount of women participating in the experiment into consideration. One reasonable explanation for this figure might be that (compared to men) a considerable amount of women had prior knowledge of game theory for sure, as pointed out by a fellow student.[5] They hence had the advantage of knowing about the iterated reasoning process. Unfortunately I did not monitor for prior knowledge in game theory, which would have been helpful and could have explained some of the figures in detail. Camerer (2003) noted that people with prior knowledge win almost every time in the first round when playing against "newbies"[6]. This advantage, according to Camerer, vanishes after the first round. Camerer's remarks would therefore account perfectly for the sheer flood of female victories observable in the first round.

Looking at Figure 3 (the average values played by men and women), it seems to confirm the hypothesis that there is no big gap between male and female performances, even though female pick a value which is slightly lower (higher in the last round), than the values picked by men. I also marked the best response[7], to show which values on average would have won the game.

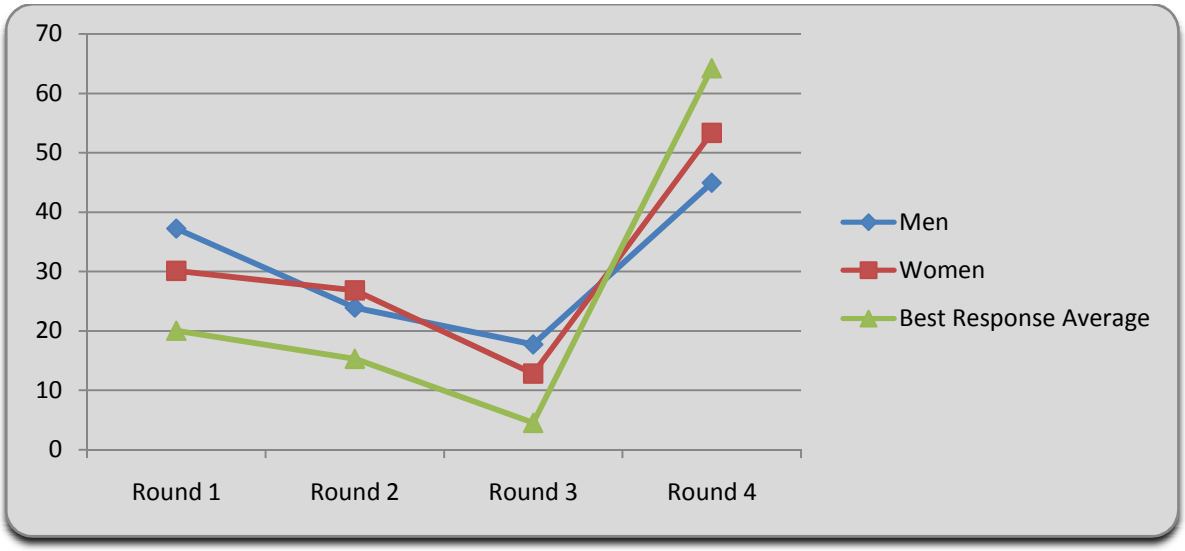


Figure 3 - Average values played by men and women (normal data set)

[5] According to the student a group of female students, who participated in the experiment had played a beauty contest before on a seminar held by the University of Passau.

[6] An enormous amount of female participants were picking values between 11 and 20 in round 1 (see Appendix), which were close to the average value to pick in order to win the game.

[7] The best response is achieved by looking at the average submission values of all men and women multiplied with p. cleansed

To explain these results without the "prior knowledge" theory, it was further suggested that there might be a gender difference in terms of belief and submission values, i.e. men may have “better” (more realistic) beliefs than women, but fail to discount properly. Looking at table 3 suggests that the theory of women discounting more accurately might indeed be right. Table 3 elucidates on the average difference between the values one should have submitted according to the own beliefs stated and the actual values submitted. Women on average deviate less from the best response (to their own beliefs) than men do. They "discount" more accurately than their male counterparts, with exception of round 4, where both genders are deviating to a large degree from their best response.

	Submissions of men deviate by X from best response to their belief	**Submissions of women deviate by X from best response to their belief**
Round 1	-3,6	-0,7
Round 2	-3,7	-1,2
Round 3	-2,4	-1,2
Round 4	7,7	7,9

Table 3 - Deviations from the "correct" values (According to the beliefs stated)

For a better understanding I depicted the beliefs of men and women compared to their actual submission values in Figure 4 and 5.All three depictions are giving crucial insight into the thinking process of the contestants, which is astonishingly close (in particular when looking at women), as to what is predicted by economic theory. The average submission values of men and women are clearly one step beneath (in round 4: above) the submitted beliefs. This is backing up the validity of the experiment, showing that people clearly form an opinion about the value of other players and “discount” it subsequently. The fact that both sexes are "overshooting", i.e. picking numbers which are slightly lower (round 4: higher) than the best response, could be explained by limited computation (humans cannot calculate as precisely as machines). Alternative explanations could be of psychological nature. Contestants in this sense could overshoot because they have slight doubts about their own beliefs (Such as: "Maybe I've underestimated my opponent?"[8]) and thus choosing a value slightly lower (round 4: higher) for security reasons.

[8] As a contestant told me afterwards, this was a thought she had right between the steps of submitting her beliefs and picking her submission values.

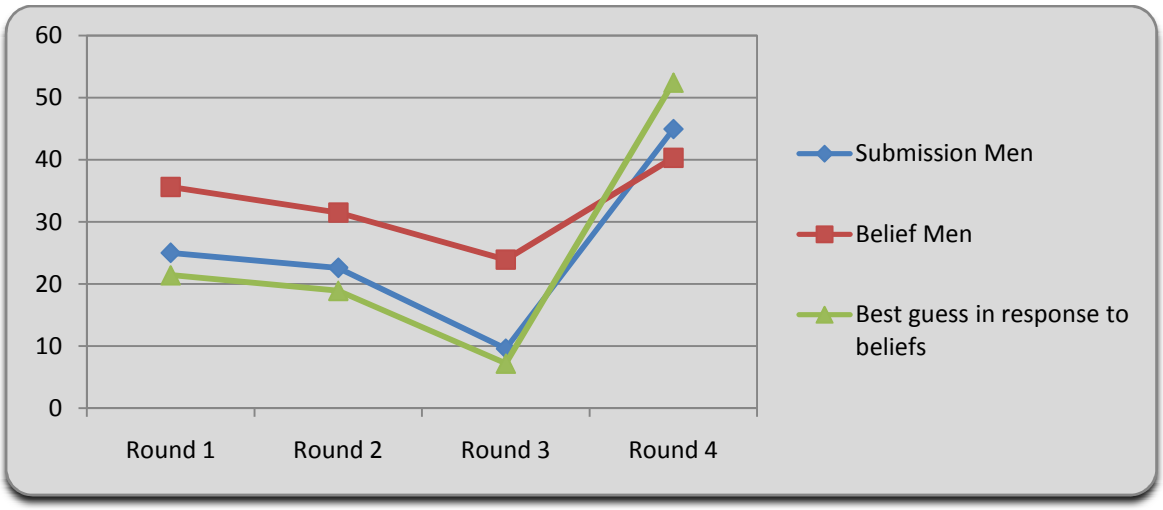


Figure 4 - Average beliefs and submission values of men (cleansed data set)

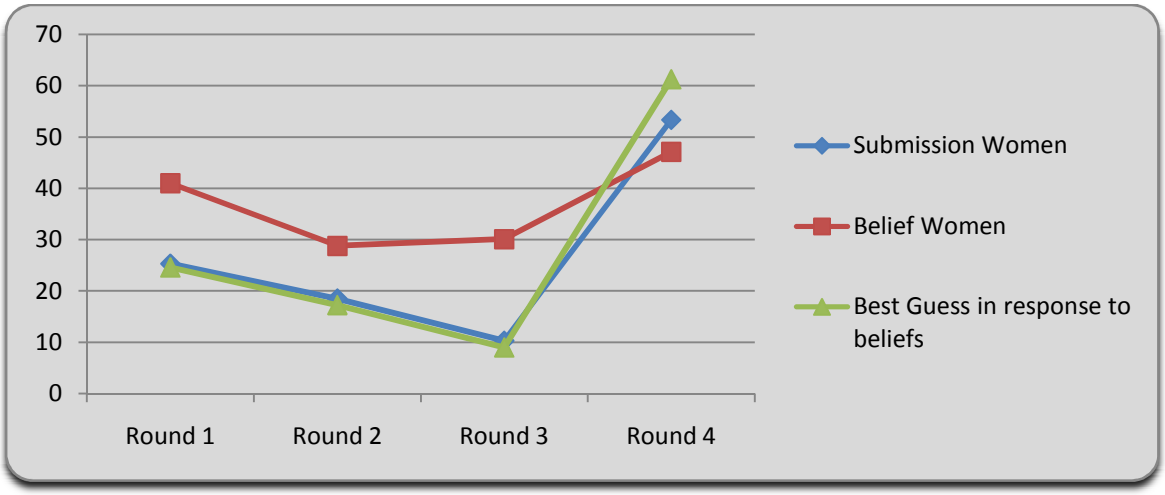


Figure 5 - Average beliefs and submission values of women (cleansed data set)

5. Conclusion

The hypothesis of women being gender sensitive seems not be without merit. Women expect men to pick lower (in round 4: higher) values than they expect other women to do and therefore believe men to be potentially better in terms of iterated reasoning, which is not the case, as can be seen, when looking at the results. Both genders seem to be equal in terms of performance to say the least. Even more: Women seem to perform slightly better, than their male counterparts. Reasons for these results could, among others, be a high amount of women, who had prior knowledge of game theory.

Casting a glance at the beliefs and the actual values played also reveals that both genders play extremely close to the best response to their stated beliefs. Women in particular are right on the spot, choosing on average a submission value, which fits their beliefs precisely.

Extreme overconfidence among men, in terms of men believing other players to not apply as much steps in iterated reasoning as women do, cannot be observed. This then again raises an important question about the male overconfidence being diminished, when the estimation of the opponent is essential for winning. Further experiments could therefore try to answer the question whether the extreme overconfidence among males, as observed in many experiments before, may only occur, when males “can afford it”.

6. References

Ben-Ner, Avner, Putterman, Louis, Kong, F. Magan, D, 2004. *Reciprocity in a Two-Part Dictator Game.* Journal of Economic Behavior & Organization, Vol. 53, 333-352

Betrand, Marianne, 2010. *New Perspectives on Gender*. Handbook of Labor Economics, Vol. 4b, S. 1545-1590

Camerer, Colin F., 2003, *Behavioral Game Theory: Experiments in Strategic Interaction.* Princton University Press

Croson, Rachel, Gneezy, Uri, 2009. *Gender Differences in Preference*. Journal of Economic Literature 47:2, S.1-27

Gneezy, Uri, Niederle, Muriel, Rustichini, Aldo, 2003. *Performance in Competitive Environments: Gender Differences.* Quarterly Journal of Economics 118, S. 1049-1074

Niederle, M., Vesterlund, L., (2007). *Do women shy away from competition? Do men compete too much?* Quarterly Journal of Economics 2007 / 3 p. 1067-1101

Sprietsma, M. (2010), Explaining *the persistent mathematics test score gap between boys and girls*. ZEW discussion paper 10-101

7. Appendix

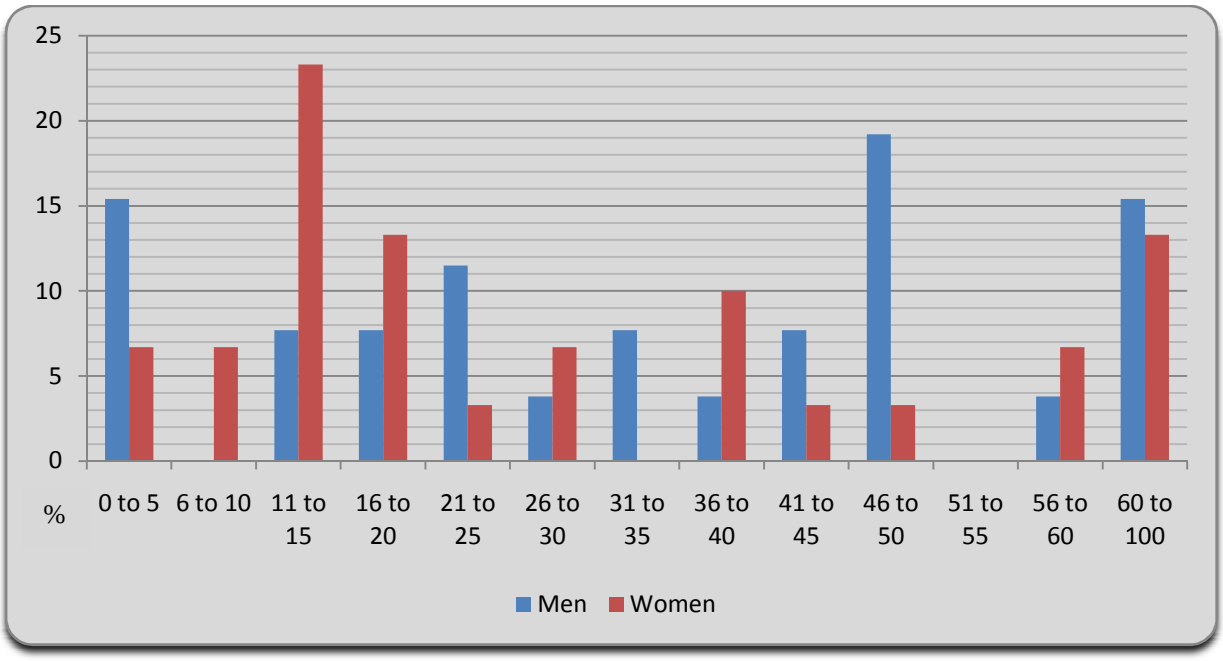


Figure 6 - Distribution of male and female submission values in round 1

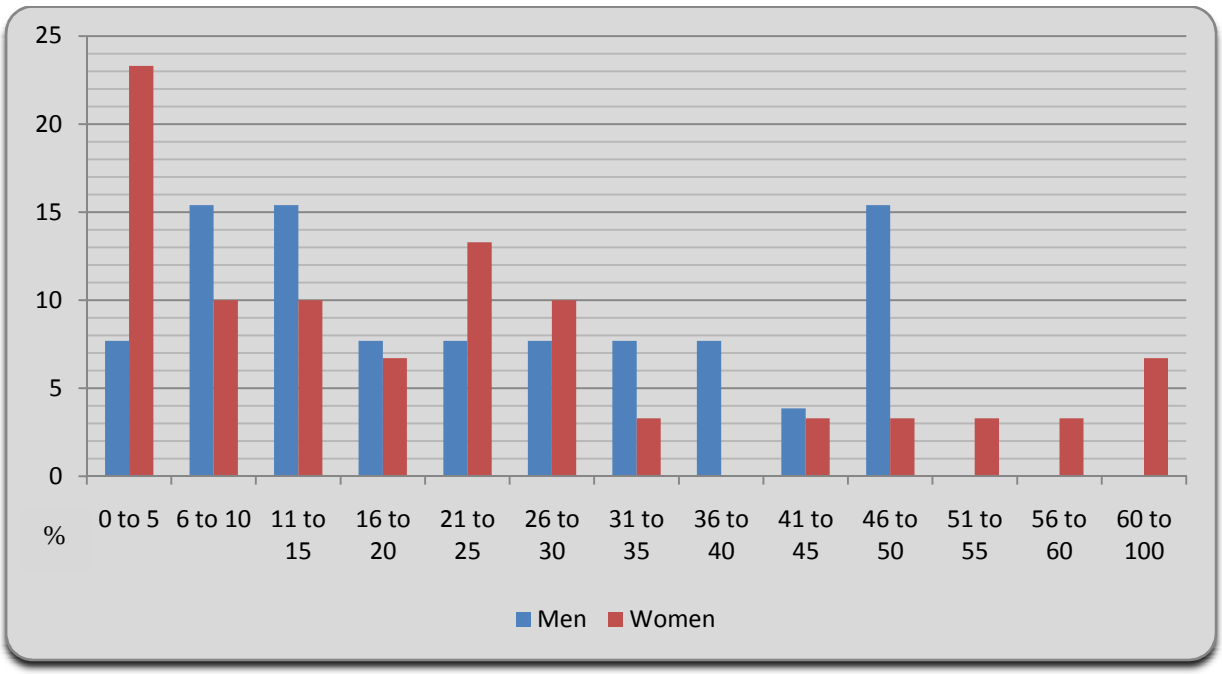


Figure 7 - Distribution of male and female submission values in round 2

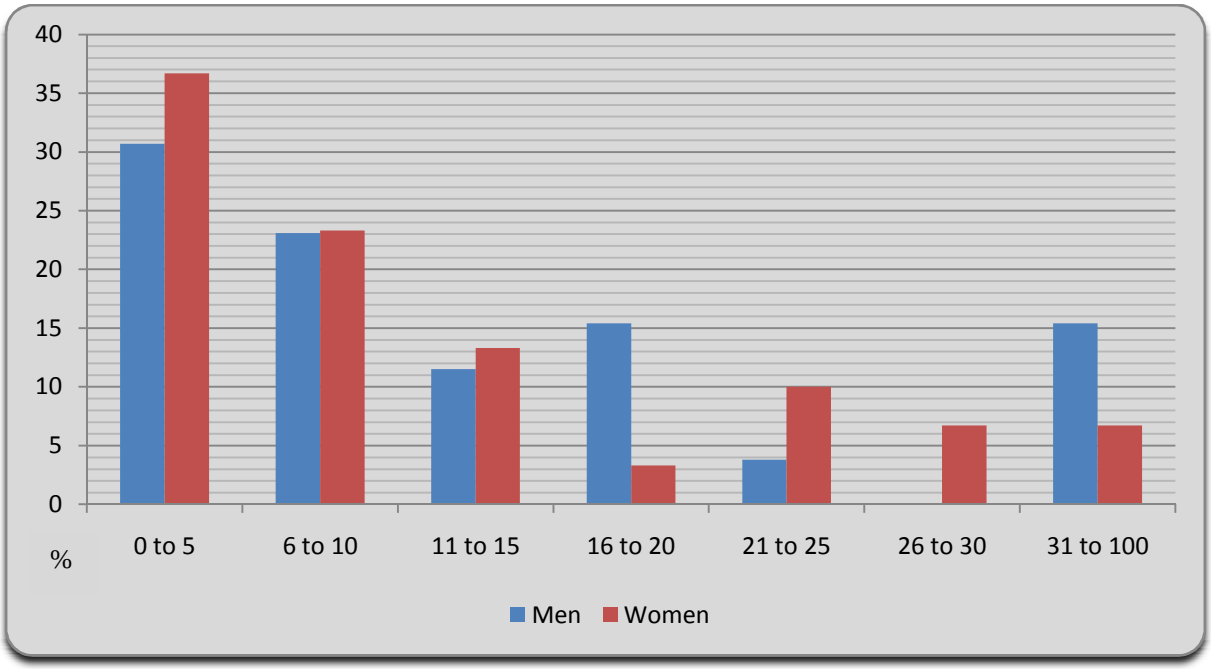


Figure 8 - Distribution of male and female submission values in round 3

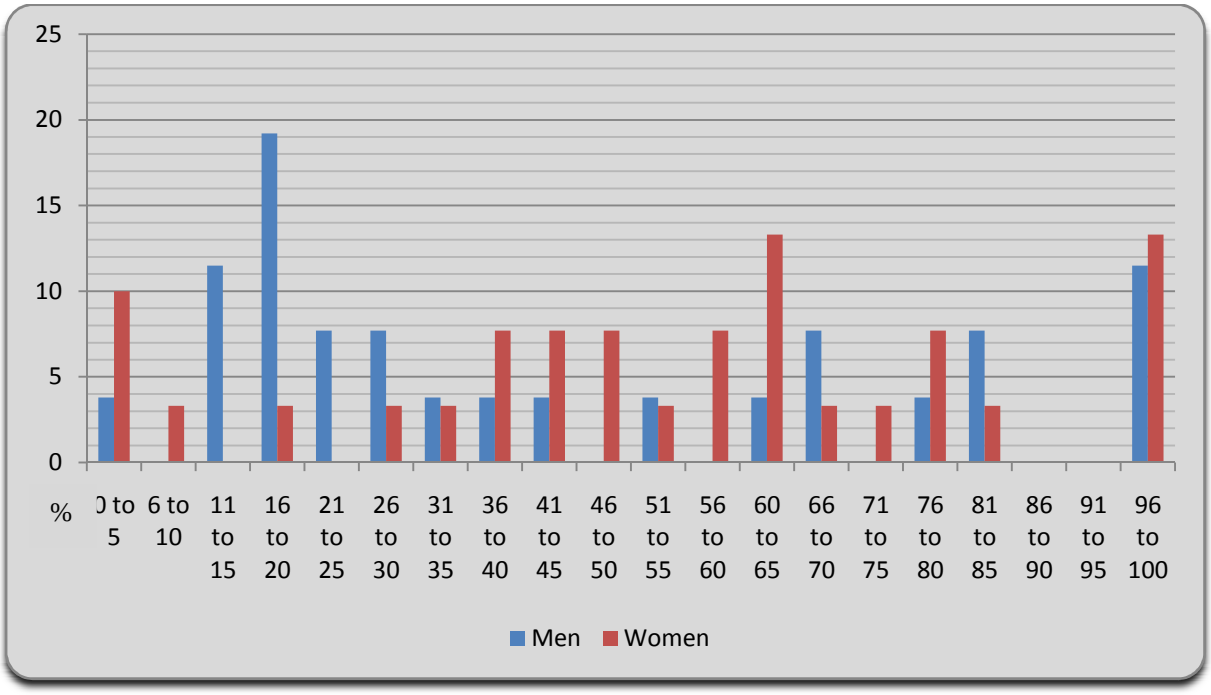


Figure 9 - Distribution of male and female submission values in round 4

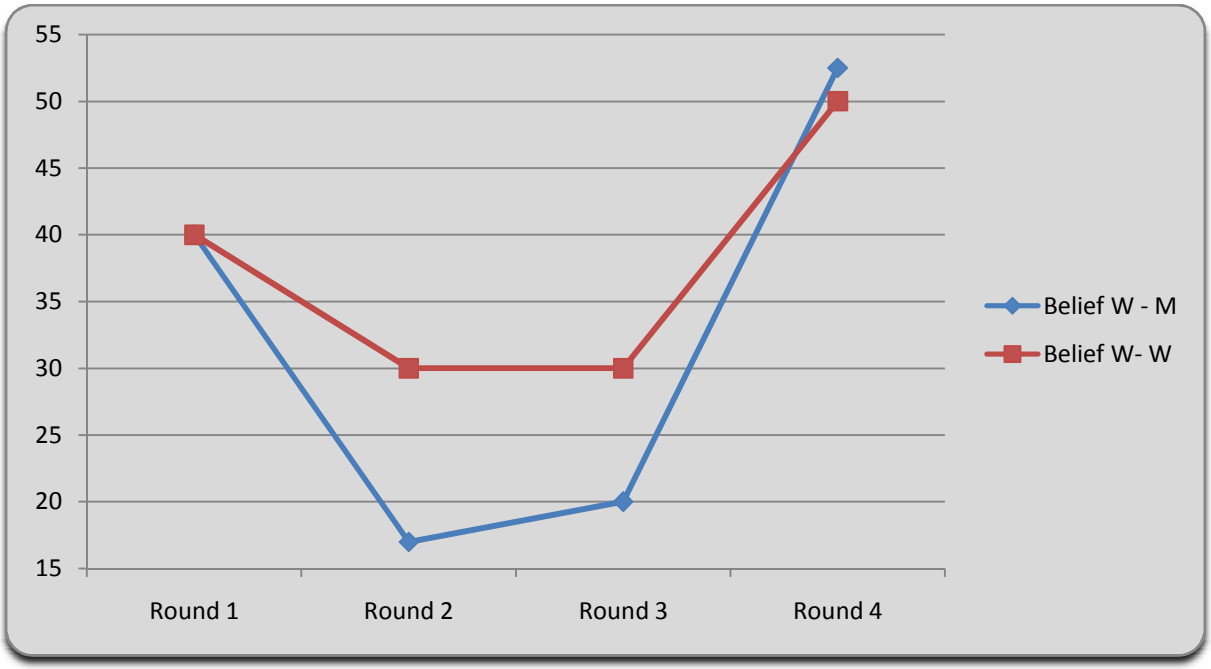
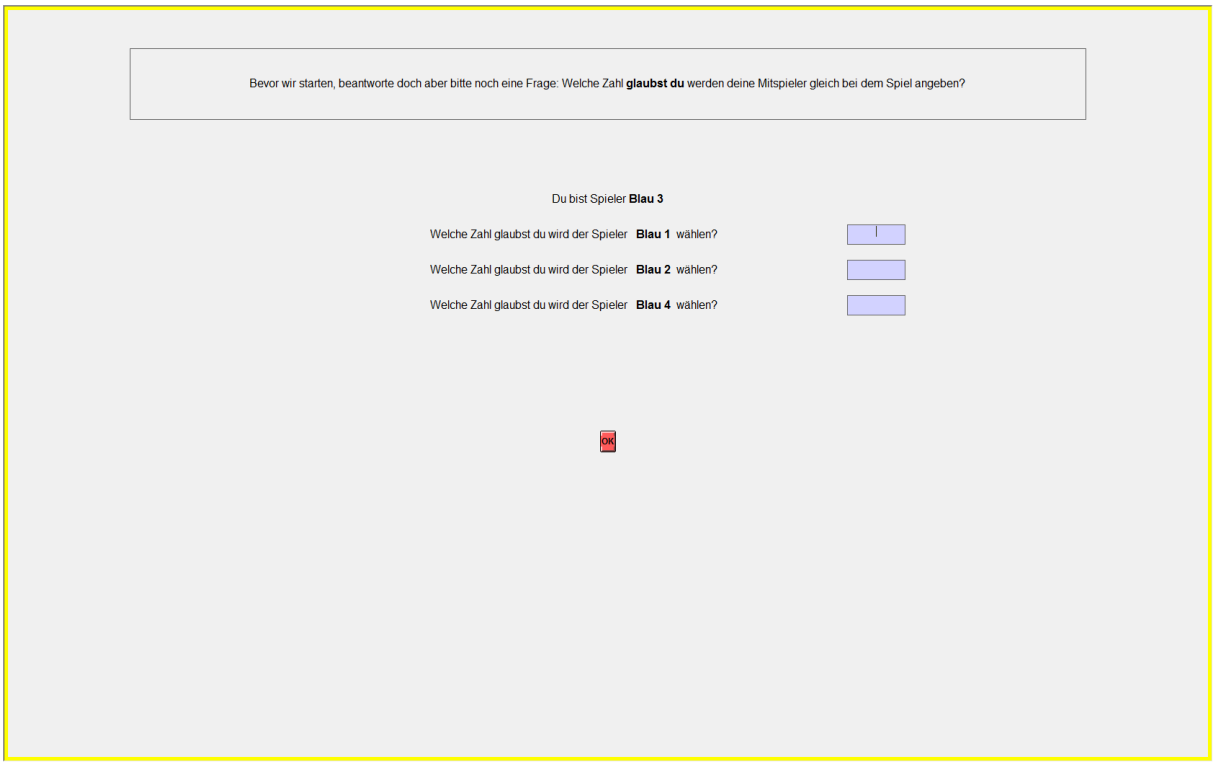
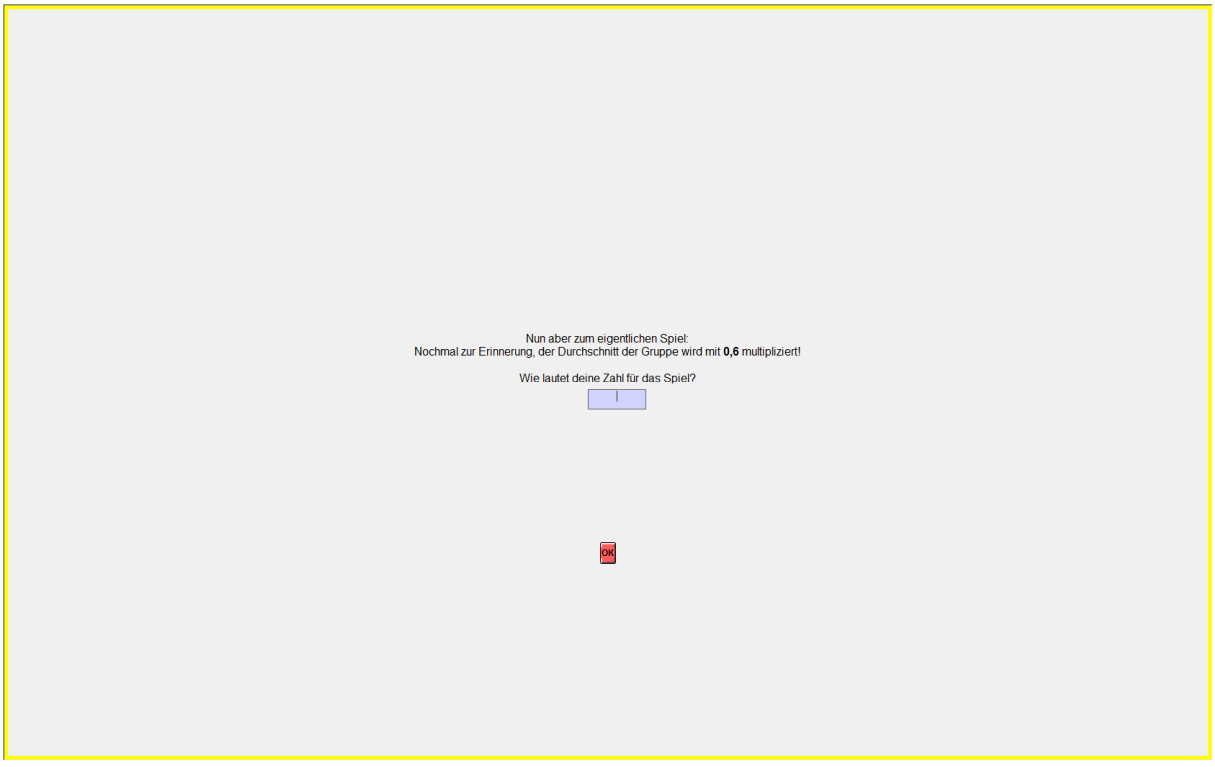


Figure 10 - Beliefs of women (median)

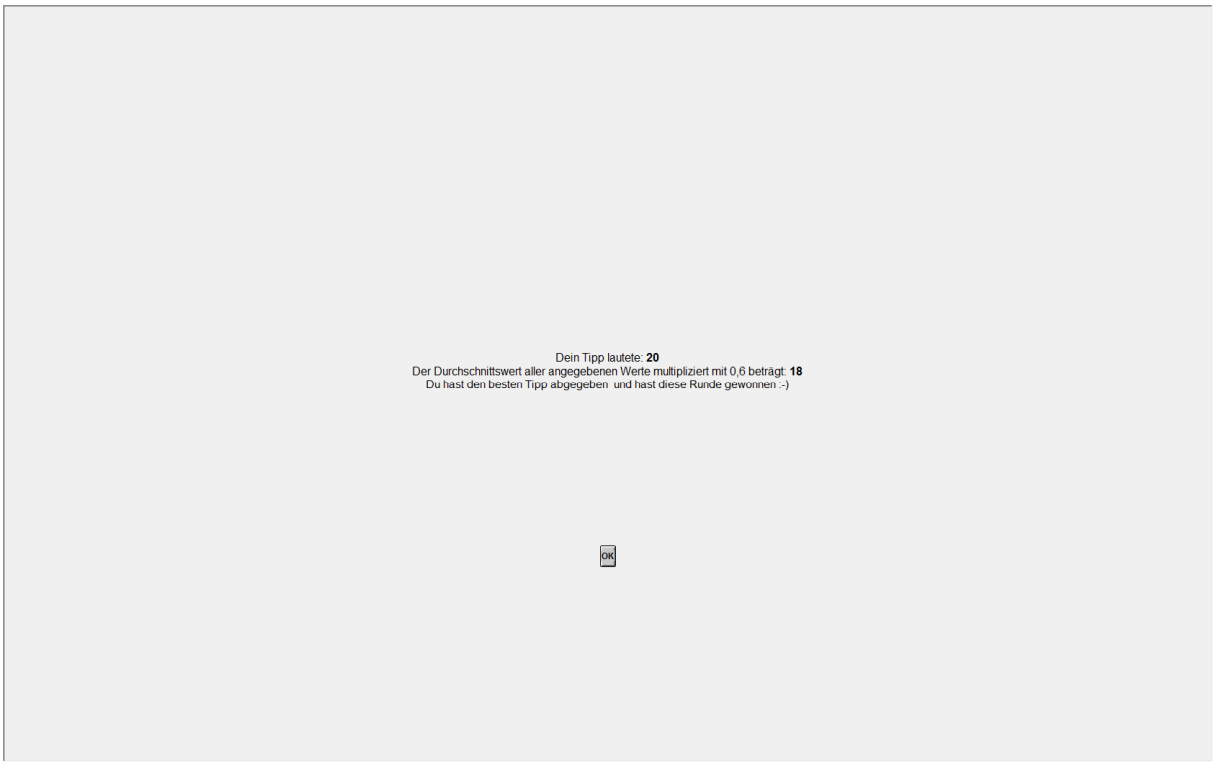
Screenshot 1 - Instructions



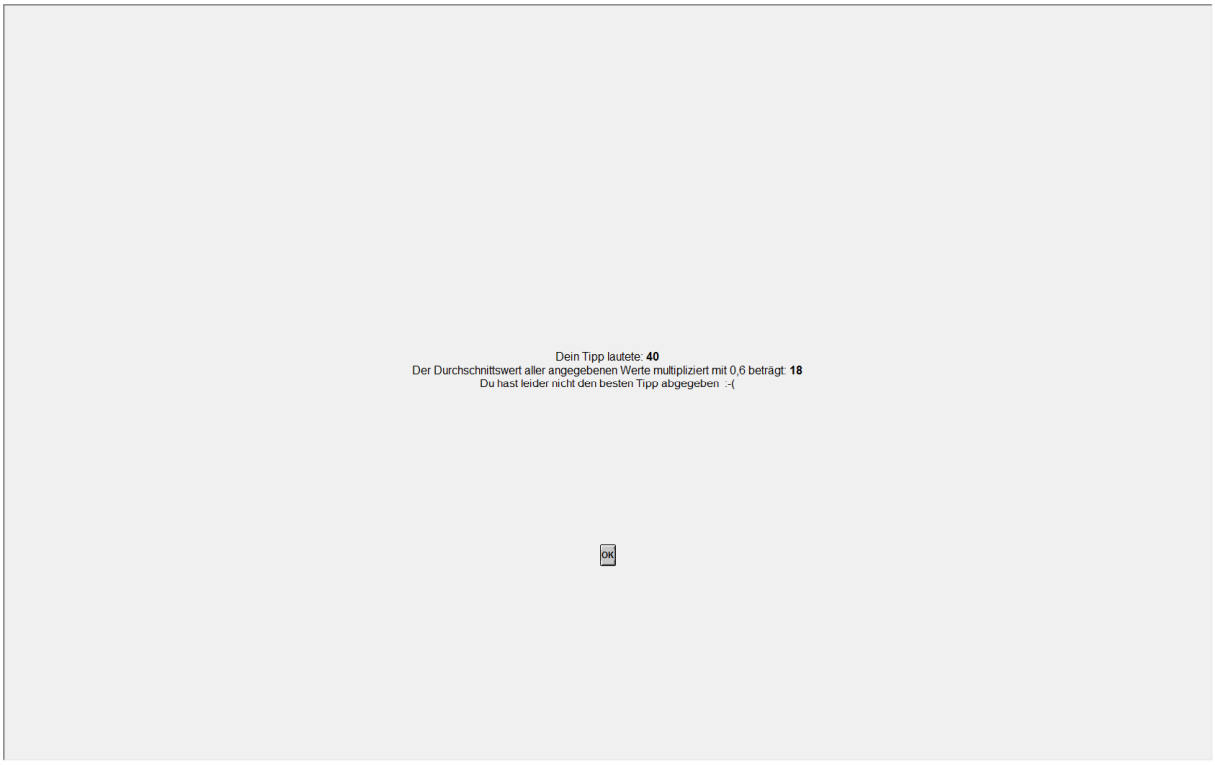
Screenshot 2 - Belief submission



Screenshot 3 - Actual submission for the beauty contest



Screenshot 4 - Message for winners



Screenshot 5 - Message for losers