Center for European, Governance and Economic Development Research

Discussion Papers

Number 188 – July 2016

ARE GROUPS 'LESS BEHAVIORAL'? THE CASE OF ANCHORING

Lukas Meub
Till Proeger

Georg-August-Universität Göttingen

ISSN: 1439-2305

ARE GROUPS 'LESS BEHAVIORAL'? THE CASE OF ANCHORING

Lukas Meub [a], **Till Proeger** [a,*],

[a] Faculty of Economic Sciences, Chair of Economic Policy and SME Research, University of Goettingen, Platz der Goettinger Sieben 3, 37073, Goettingen, Germany

* Corresponding author, till.proeger@wiwi.uni-goettingen.de;

phone: +49 551 39 10173; fax: +49 551 39 19558.

Abstract: Economic small group research points to groups as more rational decision-makers in numerous economic situations. However, no attempts have been made to investigate whether groups are affected similarly by behavioral biases that are pervasive for individuals. If groups were also able to more effectively avoid these biases, the relevance of biases in actual economic contexts dominated by group decision-making might be questioned. We consider the case of anchoring as a prime example of a well-established, robust bias. Individual and group biasedness in three economically relevant domains are compared: factual knowledge, probability estimates and price valuations. In contrast to previous anchoring studies, we find groups to successfully reduce, albeit not eliminate, anchoring in the factual knowledge domain. For the other two domains, groups and individuals are equally biased by external anchors. Group cooperation thus reduces biases for predominantly intellective tasks only, while no such reduction is achieved when judgmental aspects are involved.

Revised Version June 2016

Keywords: anchoring bias; group decision-making; heuristics and biases; incentives; laboratory experiment

JEL Classification: C91; C92; D8

1. INTRODUCTION

Economic research on group performance has evolved significantly in recent years, accounting for the fact that most economically and politically relevant decisions are taken by cooperating teams rather than individual actors. In their literature reviews, Kugler et al. (2012) as well as Charness and Sutter (2012) describe the general trend emerging from the growing body of literature on group performance. Across a broad range of experimental settings, it is shown that groups are more likely to follow game theoretic predictions and, as put by Charness and Sutter (2012, p. 159), are "*less behavioral than individuals*". Team cooperation is consequently interpreted as a means of effectively overcoming individual cognitive and motivational limitations and leading to more frequent rational behavior. Groups' increased rationality compared to individuals may serve as a partial vindication of the assumption of rational choice theory in reality (Charness and Sutter, 2012). This argument lends strong support to those strands of literature arguing that market conditions tend to eliminate irrational behavior through monetary incentives and learning effects. Widespread team decision-making might thus further support the argument of markets as "*Catalyst for Rationality and Filter of Irrationality*" (List and Millimet, 2008, p.1).[1] However, while numerous economic games have been considered in terms of group cooperation and rationality, the area of heuristics and biases has been neglected with respect to group performance for economic experimental contexts. Despite being assumed by Kugler et al. (2012), it remains open to question whether groups more effectively overcome individual cognitive biases. Although the current economic literature on team performance might lend support for this view, experimental evidence has yet to be provided.

In this paper, we investigate group decision-making in the domain of biases by drawing on the case of the anchoring-and-adjustment heuristic, initially presented by Tversky and Kahneman (1974). Their seminal experiment showed that subjects, when asked to provide their best estimates for the percentage of African nations in the United Nations, were biased towards numbers randomly generated by a wheel of fortune. Despite the obvious irrelevance of the anchor values, they systematically influenced subjects' estimations. Forty years of variations on the classical experimental design have found the anchoring bias to be fairly robust against experimental variations (cp. the literature review by Furnham and Boo, 2011). Building on

[1] Other contributions questioning the robustness of behavioral biases under market conditions, mostly drawing on field evidence from well-functioning markets, include e.g. List (2003; 2004a; 2004b; 2006); Levitt and List (2007); Cecchi and Bulte (2013).

previous anchoring research, we focus on a single bias, yet derive generalizable results for the area of economic group decision-making and biases by investigating the effects of group cooperation given different task characteristics. This in turn allows for a transfer of our findings to other biases and a broader comparison of group and individual behavior.

We chose to consider anchoring for our investigation into groups and biases due to its prominent application in explaining distortions in quite diverse economic situations. Assuming a robust influence of anchoring effects from previous laboratory experiments, numerous real-world behavioral effects have been attributed to anchoring. Recent examples for these applications include real estate pricing (Bucchianeri and Minson, 2013), art and online auctions (Beggs and Graddy, 2009; Dodonova and Khoroshilov, 2004), sports betting (Johnson et al., 2009; McAlvanah and Moul, 2013), earnings forecasts (Cen et al., 2013), financial forecasts (Fujiwara et al., 2013), macroeconomic forecasts (Bofinger and Schmidt, 2003; Campbell and Sharpe, 2009; Hess and Orbe, 2013) and sales forecasting (Lawrence and O'Connor, 2000).

In contrast to the view of a strong influence of anchoring in non-experimental settings, a number of economic field and laboratory experiments on anchoring in price valuations find no or only moderate effects (Simonson and Drolet, 2004; Tufano, 2010; Alevy et al., 2015; Fudenberg et al., 2012; Maniadis et al., 2014.) Based upon these results, it could be argued that different parameters prevalent in actual markets could successfully correct irrationalities of individual heuristics. Therefore, rationality-increasing teamwork as a ubiquitous form of decision-making in actual markets might be an additional filter for biased decisions that has not been considered in previous experimental studies.

We test whether groups are more or less susceptible to externally provided anchors than individuals in three distinct economic domains. We cover factual knowledge, probability estimations and price valuations and implement strong monetary incentives for unbiased decisions for groups and individuals alike. We argue that these three domains of decision-making cover well the range of economically relevant situations prone to irrationally anchored decisions as described in the recent non-experimental studies of anchoring on real-world decisions such as pricing, judgmental forecasting or auctions. While our anchoring exercises closely resemble the seminal anchoring studies, a competitive scheme of monetary incentives distinctly rewards cognitive effort and adjustment away from given anchors. While these conditions are rarely implemented in psychological anchoring studies, they seem necessary to test the robustness of biases under market conditions. Our results can thus add both to the current literature on the prevalence of anchoring in actual markets and to the question of whether groups might avoid biases that are found to be robust for individuals. In the following, we review the related literature.

Following Tversky and Kahnemann's (1974) seminal paper, a large body of literature has dealt with the causes and consequences of the anchoring bias.[2] With regard to the central features of our design, the results for individual decisions are fairly unambiguous. Exercises on factual knowledge have been repeated numerous times and lead to robust and substantial anchoring effects (Blankenship et al., 2008; McElroy and Dowd, 2007). The same holds true for probability estimations, as shown by Chapman and Johnson (1999) and Plous (1989). Price valuations and willingness to pay are covered in the studies by Sugden et al. (2013), Adaval and Wyer (2011), Bateman et al. (2008), Critcher and Gilovich (2008), Nunes and Boatwright (2004), Simonson and Drolet (2004), whereby rather moderate anchoring effects are shown for valuation tasks when compared to disparities in WTA/WTP due to anchoring. Consequently, in the domain of economic valuations, anchoring effects appear to be more fragile than in judgmental domains. The effect of monetary incentives remains disputed: Tversky and Kahnemann (1974), Wilson et al. (1996) and Epley and Gilovich (2005) offer prices for unbiased decisions and find no debiasing effects. In contrast, Wright and Anderson (1989), Simmons et al. (2010) and Meub and Proeger (2015) find subjects to be less biased when given monetary incentives and a realistic opportunity of achieving better solutions through increased cognitive effort.

Bias reduction through group cooperation in psychological experiments is reviewed by Kerr and Tindale (2004). While for some domains groups are less biased, such as for the hindsight bias (Stahlberg et al., 1995) and the overconfidence bias (Sniezek and Henry, 1989), the overall heterogeneity of the results and experimental paradigms preclude general predictions as to whether groups more effectively avoid behavioral biases. For anchoring in groups, studies in two domains have been carried out. In a study of legal juries, Hinsz and Indahl (1995) report that legal judgments are as biased by anchors provided during trials as for individuals. Whyte and Sebenius (1997) also find that for a non-incentivized negotiation exercise, groups acting as

---

[2] Furnham and Boo (2011) provide a general overview, Epley and Gilovich (2010) sum up the discussions on the theoretical foundations of anchoring. There are several contradictory methodological notions within psychological research regarding the reasons and influences on anchoring. For a more detailed view into this recent discussion, we refer to Frederick and Mochon (2012), Critcher et al. (2014), Wegener et al. (2010a) and Russo (2010), Frederick et al. (2010), Epley and Gilovich (2010) and Wegener et al. (2010b).

a single party are equally biased as individuals. Accordingly, groups fail to effectively use competing anchors to debias their judgment; rather, they compromise between various distorted individual judgments, making the overall result similarly biased. Nonetheless, while these studies might provide some perspective on group behavior when confronted with external anchors, we argue that they hold limited relevance for economic group research for the reasons outlined by Kugler et al. (2012): firstly, the lack of clear theoretical and consistent paradigms that allow for some benchmarking of expected and actual behavior; and secondly, the ubiquitous lack of financial incentives for cognitive effort that would induce more reasonable answers (for an elaboration of this aspect regarding anchoring, see Alevy et al., 2015).

In contrast to psychological studies, economic small group research offers fairly clear predictions for group performance, yet so far provides no evidence in terms of biases and groups. Reviewing the past ten years of economic group experiments, Charness and Sutter (2012) and Kugler et al. (2012) summarize that groups overall are more successful than individuals in achieving game-theoretical requirements for rational decision-making by alleviating cognitive limitations of individuals, as collaboration enables more rational decisions through the transfer of insight from cognitively superior individuals to the group. The effectiveness of this mechanism crucially depends on the demonstrability of task solutions (Laughlin et al., 2002). Hence, groups consistently outperform individuals in intellective tasks with a clear and demonstrable correct solution. The counterpart are judgmental tasks that have more ambiguous answers, whose solutions are not easily demonstrable to other persons (Cox and Hayne 2006). Group performance then depends on the respective task's position on a continuum from intellective to judgmental (Laughlin, 1980). Consequently, groups can mitigate individuals' bounded rationality through the transfer of information and it can be hypothesized that groups circumvent the anchoring bias through improved intra-group information availability. The positive effect of "more heads" on the overall cognitive performance thus leads to the expectation that groups will be less biased by external anchors.

Accordingly, there are two contradictory notions to be derived from previous research. The bulk of psychological research on anchoring effects in individuals and groups leads to the prediction that groups are unlikely to avoid the bias, regardless of monetary incentives. Group cooperation would thus fail to alleviate the bias regardless of task characteristics. Also, the active discussion of anchor values might as well foster the activation of anchor-consistent knowledge and even increase anchoring effects, e.g. through group polarization (Luhan et al, 2009). By contrast, following economic small group research, the cognitive superiority of groups would predict that groups successfully avoid external anchors as additional information becomes available within groups.

To account for these contradictory notions, we present an anchoring study comprising three different anchoring exercises that are compatible with previous psychological experimental designs and also with the economic domains discussed recently in anchoring field experiments. Additionally, we implement strong monetary incentives for unbiased decisions. In this setting, we compare the performance of individuals and three-person teams in terms of their ability to avoid anchors.

We find that groups are significantly less biased for an intellective factual knowledge task. For probability estimates and price valuations, individual and group decisions are equally biased; accordingly, individual biases are perpetuated by group cooperation. It appears that a group's ability to reduce individual biases depends on task characteristics. In the case of intellective tasks that have a clearly defined correct solution, debiasing is effective. For tasks with judgmental elements, groups approach the performance of average individuals. Overall, we suggest that groups are 'less behavioral' than individuals in certain domains, yet that this optimistic assumption prevalent in the literature should clearly not be generalized to all domains of decision-making and biases.

The remainder of the paper is structured as follows. In section two, our experimental design is described, while section three details our results and section four concludes.

2. EXPERIMENTAL DESIGN

2.1 THE GAME

We keep in line with previous anchoring studies by implementing five exercises from three domains that have been covered extensively for individuals, namely factual knowledge questions, probability estimates and price valuations. The factual knowledge questions featured topics related to the city of Goettingen and were chosen to ensure that experimental participants were somewhat familiar with these topics regardless of their field of study or age. The price valuations were based on pictures of several used articles sold on the internet.[3]

Subjects have 90 seconds for every exercise to enter their answer.[4] We determined a random order of exercises under the constraint that every subset of three periods comprises one exercise from each domain. Table 1 shows all 15 exercises of the three domains in the previously determined order, whereby the column *sequence* refers to two strictly opposing orders of high and low anchor values (A and B), which is explained below in detail.

[3] The pictures shown to participants are documented in Appendix B. All respective brands were erased from the pictures to prevent subjects from being influenced by brand names. All items were sold on web-based platforms comparable to eBay. Although smartphones were banned during the experiment, we additionally ensured that the correct answers to our questions could not be easily looked up.

[4] While 90 seconds may appear to be a too short as a timeframe, our pilot experiment indicated that the majority of groups did not use the entire time. Some groups would, however, decide on a correct solution within the first minute and spend the remainder of the time with idle talk on personal topics. We thus decided not to further extend the time frame for group discussions.

No.	Sequence A(B)	Domain	Exercise
1	H (L)	Factual knowledge	What percentage of students in Goettingen in 2011 were originally from North-Rhine Westphalia?
2	L (H)	Probability	What is the likelihood of a European team winning the football world championship in Brazil in 2014?
3	L (H)	Price	What is the price for this used bike? (see photo below)
4	H (L)	Probability	What is the likelihood of the Euro still being the currency in Germany in 5 years?
5	L (H)	Factual knowledge	How many students in Goettingen were between 18-21 years old in 2011?
6	H (L)	Price	What is the price for this used washing machine? (see photo below)
7	H (L)	Factual knowledge	How many students were living in the Goettingen district of Weende in 2011?
8	L (H)	Price	What is the price for this used coat? (see photo below)
9	H (L)	Probability	What is the likelihood of an unemployed person in Germany not finding a new job within 4 weeks?
10	L (H)	Factual knowledge	How many students were officially registered as residents in Goettingen in 2011?
11	H (L)	Probability	What is the likelihood of being stuck in a traffic jam for at least twenty minutes when driving from Munich to Hamburg?
12	H (L)	Price	What is the price for this used TV set? (see photo below)
13	L (H)	Probability	What is the likelihood of a white Christmas in Munich in 2014?
14	L (H)	Price	What is the price for this used ring? (see photo below)
15	H (L)	Factual knowledge	How many students in Goettingen received Bafoeg (government study grants) in 2011?

Table 1. Questions in the same order as presented in the experiment. Order refers to the sequences of high (H) and low (L) anchors

2.2 TREATMENTS

We run two experiments: one featuring individual decisions and another with groups of three players. Groups are randomly matched at the beginning; they are required to find a unitary solution using chat communication.

Using a between subjects design, each experiment comprises two treatments: *calibration* and *anchor*. This implements the procedure initially established in the seminal design by Jacowitz and Kahneman (1995), which allows for distinctively quantifying the anchoring bias.

Calibration

Calibration is the benchmark treatment conducted with both individuals and groups to determine the anchor values used later on in *anchor*. *Calibration* has subjects merely enter their answer for the 15 estimation exercises. The resulting estimates are then used to derive the anchor values, whereby the low and the high anchors represent the 15^{th} or the 85^{th} percentile of the estimations' distribution, respectively. Individuals exclusively received the anchor values derived in the individual *calibration* and groups those from group *calibration*. Accordingly, the absolute anchor values displayed in *anchor* are different for group and individual players, although anchors are the same in relative terms.

Anchor

In *anchor*, the values derived in *calibration* are shown to participants as external anchors. Additionally, in line with the seminal experimental paradigm for anchoring (Tversky and Kahneman 1974), participants indicate whether they assume the correct answer to be higher or lower than the anchor prior to their estimate. Both groups and individuals are displayed the respective anchor values at the center of the decision screen. The anchor value is included in the higher/lower question and thus has to be read by each participant in order to give an estimation.[5] Therefore, our design compares individual and group decision-making when players are confronted with anchors in the immediate context of the situation. We do not assess the difference between individuals confronted with an anchor who subsequently decide either autonomously or join a group discussion. This second approach, while plausible, would induce a time lag between presenting the anchor and the actual decision, which our design choice helps to avoid. Our design thus represents, for instance, the scenario of a team decision process where members jointly evaluate relevant information and are thus exposed to anchors simultaneously. This simultaneous exposition to anchors might be especially relevant if members discuss a decision in a personal meeting and information – as well as irrelevant numbers – occurs on slides, posters or on a board.

It is to note that we refrain from testing whether group or individual players provide estimates that are closer to the correct answers. Instead, we investigate if and to what extent the distribution of estimates differs between groups that are shown anchors and those in *calibration*. The same procedure is applied for individuals. We can then compare the change in behavior for

[5] Screenshots of the decision screens presented to participants in the individual and group experiments in both treatments are provided in Appendix C.

groups to the change in behavior for individuals. This calibration procedure, introduced by Jacowitz and Kahneman (1995), yields the advantage that exercise specifics hold little relevance, as we can control for subjects' capabilities to answer the questions by running *calibration.* Therefore, our results on the anchoring bias do not depend on exercise specific knowledge (which is likely to be higher for groups than for individuals), given that we compare groups' and individuals' estimates in *anchor* to groups' and individuals' estimates regarding the same exercises in *calibration* and then compare differences between groups and individuals. This approach entails the additional advantage that the order of exercises can be held constant so that learning effects have no impact on the validity of our findings.

While the order of exercises is held constant for both experiments, there are two strictly opposing sequences of high and low anchors (A and B). This gives us estimates for each exercise for the cases of a low anchor, a high anchor and no anchor. The two sequences of anchor values are shown in Table 1 and guarantee that each subject takes at least two exercises for both high and low anchors in each of the three domains.

Further, table 2 summarizes our experiments and the number of participants.

	Calibration	***Anchor***		**Total**
		Sequence A	Sequence B	
Individual	N=24	N=24	N=24	N=72
Group	N=72	N=60	N=60	N=192

Table 2. Overview of treatments and number of participants

Finally, following the 15 exercises, we assess the perceived relevance of the anchors by asking participants how important the reference (anchor) values were for their decisions, on a scale of one to nine. We thereby aim at identifying differences in the perception of anchor values between individuals and groups, which might help to explain potential differences in the extent of the anchoring bias. We thereby draw on a current discussion on the perceived relevance of the anchor values as a determinant of the magnitude of anchoring (Gloeckner and Englich, 2014; Meub and Proeger, 2015), which has so far led to mixed evidence.

2.3 PAYOFFS

We digress from the classical psychological anchoring experiments by providing monetary incentives for unbiased behavior, whereby giving the most accurate estimates possible becomes the optimal, payoff-maximizing strategy. However, in the important domains of price valuations and probability estimates there are no unambiguously correct answers. We therefore chose the five factual knowledge questions to determine payoffs as they have unambiguously correct.

It was explained in the instructions that only five of the fifteen exercises were rewarded monetarily, but it did not state to which of the 15 exercises this applied. To avoid participants from identifying the exercises that actually determined their payoff (which would influence their behavior across exercises and potentially confound with our treatment variable), we chose, besides factual knowledge questions, price valuations and probability estimates, for which several sources propose distinct numbers as correct answers. For each price valuation, there is a distinct price set by the respective supplier on the internet; for each probability estimate, there are sources claiming that they can provide an estimation of probabilities. Subjects could therefore not assume that these tasks were unanswerable per se. Consequently, the payoff could also have been determined using the non- factual knowledge exercises; thus, participants could not be sure that only a single domain of exercises was payoff-relevant but had to put effort into answering all exercises to the best of their knowledge. We thus elicit the same cognitive effort in participants for all three domains. Also, this design choice might only cause problems if it affects the potential change in behavior due to the anchors in systematically different manner for groups and individuals, which we assume to be rather unlikely.

A payoff scheme based upon relative or absolute deviations from the correct values across the different categories of exercises would have made the experiment more difficult to comprehend for subjects. It also would have become very difficult to guarantee a reasonable payoff (distribution) as the accuracy of estimates might have become quite heterogeneous and weak overall. We therefore implemented a more intuitively understandable incentive structure, whereby subjects are paid more if they perform better relative to the other players. Hence, our payoffs depend on the relative precision of players' estimates, i.e. their absolute deviation from the correct answers in comparison to the respective deviation of all other players. For each of the five questions, the top three ranked individual players are thereby awarded with 25 ECU, ranks four to six earned 20 ECU, ranks seven to nine received 15 ECU and all remaining subjects earned 10 ECU. In the group experiment, each of the three subjects from the best ranked group earned 25 ECU, the members of the second best gained 20 ECU and the third 15

ECU. Again, all others received 10 ECU. For both experiments, 10 ECU converted to €1. Consequently, our payment scheme has a competitive component. As previous experiments have shown that groups are more competitive than individuals (Wildschut et al., 2003), this effect could drive group-individual differences. However, as we exclusively compare the biasedness of groups and individuals with their respective counterparts in *calibration*, the differences in competitiveness between individuals and groups are irrelevant in our case. The competitive component could confound our results only if groups in calibration were affected systematically differently than groups in anchor, which we do not find to be realistic. Additionally, in the group experiment, all participants received a fixed payment of €2.5 to account for the potentially increased duration of the respective sessions due to chat communication.[6] Accordingly, the minimum payoff for individuals (groups) amounted to €5 (€7.5), while the maximum was €12.5 (€15.0).

2.4 PROCEDURE

The experiments took place in 12 sessions within one week in January 2014. They were conducted using z-Tree (Fischbacher, 2007) in the Laboratory for Behavioral Economics at the University of Goettingen. Participants were recruited using ORSEE (Greiner, 2004) and were only allowed to participate in one session. Understanding of the game and the payoff mechanism was ensured through control questions before the experiment. The sessions lasted around 35 minutes for individuals and groups. On average, individual participants earned €6.8 and group members €9.0. Participants were on average 24.1 years old, 54.7% were female.[7]

[6] The fixed payment for groups was implemented for the case that groups would regularly fail to provide matching responses within 90 seconds. In this case, group members were also shown another screen for 30 seconds asking them to provide a common answer to avoid missing values. However, this case rarely occurred.

[7] We recorded demographic information, gender and fields of study of all participants in a questionnaire after the experiment, finding that there are no significant or systematic differences with respect to the distribution over treatments. In the individual (group) player experiment 52% (56%) were female, 91% (86%) native speakers and 83% (89%) already participated in at least one experiment (Fisher's exact test, two-sided, p=.452 for gender; p=.345 for language; p=.135 for participation). On average they were 23.6 (24.4) years old and studied for 6.2 (5.5) semesters (Wilcoxon rank-sum test, z=-1.349, p=.1172 for age; z=.836, p=.4034 for semester). Also, there is no significant difference regarding the distribution across the major fields of study (Fisher's exact test, two-sided, p=.26). The original instructions were in German and are available from the authors upon request. A translation is provided in Appendix A.

3. RESULTS

To comprehensively explain our results, we first present the derivation of the anchor values. We then outline the *anchor coefficient* as a rough measure to provide an overview of the results and introduce the *anchor effectiveness index* as our key measurement for differences between groups and individuals. Subsequently, the results for the three exercise domains are presented in detail; finally, we present our findings on the perceived anchor relevance.[8]

3.1 ANCHOR DERIVATION

As explained in section two, we run a calibration treatment for both individual and group players. *Calibration* gives us the anchor values and serves as a benchmark to which we can compare the answers given in the anchor treatments. Recall that anchor values represent the 15th or the 85th percentile of the estimations' distribution in *calibration*. Jacowitz and Kahneman (1995) outline this method to allow for quantifying and testing the anchoring bias. By applying this procedure, we can further identify differences between individual and group players by analyzing their reaction to anchors in comparison to their unbiased behavior drawn from the respective calibration treatments. Table 3 summarizes descriptive statistics for *calibration*.

		***Individuals* (n=24)**				***Groups* (n=72, 24 groups)**			
exercise	**correct**	**mean**	**median**	**15th pct.**	**85th pct.**	**mean**	**median**	**15th pct.**	**85th pct.**
1	13	23.3	20	10	28	18.9	18	9	25
2		54.9	52	20	90	59.8	65	35	80
3		244	155	85	500	365	252	110	800
4		95.5	98	90	99	94.2	97	85	100
5	3335	5743	6206	3000	8000	7130	7000	5000	10000
6		136	120	75	200	140	115	65	200
7	3186	5131	4980	2100	8500	5201	5000	1000	8000
8		46.5	40	25	69	49.8	42	30	65
9		57.9	61	22	81	65.7	70	30	87
10	12705	17797	18000	14000	23200	19912	18625	14000	21000
11		71.3	75	50	85	78.4	80	62	90
12		209	184	100	300	196	175	120	280
13		45.9	48	17	73	37.0	38	20	55
14		52824	1778	200	13000	46530	8000	350	62000
15	4948	8960	8000	4300	15200	11487	9500	5500	16000

Table 3. Descriptive statistics for the calibration treatments

[8] There are 2.5% missing values for individual decisions, i.e. no estimate was entered before the time limit. For groups, 67% of all decisions were taken within the 90 second time limit, 27% took up the additional 30 seconds and 6% are missing values.

The anchor values derived for group and individual players are quite similar. Exercise 14 is an exception, in which the high anchor for groups far exceeds that for individual players. Note that values for the 15th and 85th percentile, i.e. the anchor values, are rounded such that there are no differences between the individual and group experiment with respect to the mere appearance of the anchors.[9] Furthermore, the 85th percentile for groups in exercise 4 equals 100. Implementing 100 as the anchor value for the probability estimate would violate our setting, which incorporates the higher/lower question. We therefore use 99 as the high anchor value for individuals in exercise 4. In factual knowledge questions, both individuals and groups tend to overestimate the correct answers, while there is no significant difference between individuals and groups in terms of estimation accuracy.[10]

3.2 THE ANCHOR COEFFICIENT

Before analyzing the results for each exercise domain, we initially provide a descriptive overview of the anchoring bias in group and individual players. We therefore calculate the *anchor index* as introduced by Jacowitz and Kahnemann (1995). The anchor index (AI) is defined as the difference between the median estimate in the high and the low anchor condition divided by the distance between the high and low anchor. Let $\tilde{x}_{ij}^{high}$ denote the median estimate of player i={individual, group} of exercise $j \in [1,15]$ when faced with the high anchor condition and $\tilde{x}_{ij}^{low}$ denote the respective median in the low anchor condition; A_{ij}^{high} gives the high anchor value and A_{ij}^{low} the low anchor value derived from *calibration*. We can then write for the calculation of the anchor index:

$$(1) \qquad AI_{ij}=[(\tilde{x}_{ij}^{high} - \tilde{x}_{ij}^{low}) / (A_{ij}^{high} - A_{ij}^{low})]$$

An index value AI_{ij} equal to 0 implies that there is no difference in median estimations between the low and high anchor condition for player i and exercise j. A value of 1 indicates a difference equal to the distance in the anchor values, while values greater than 1 correspond to a difference

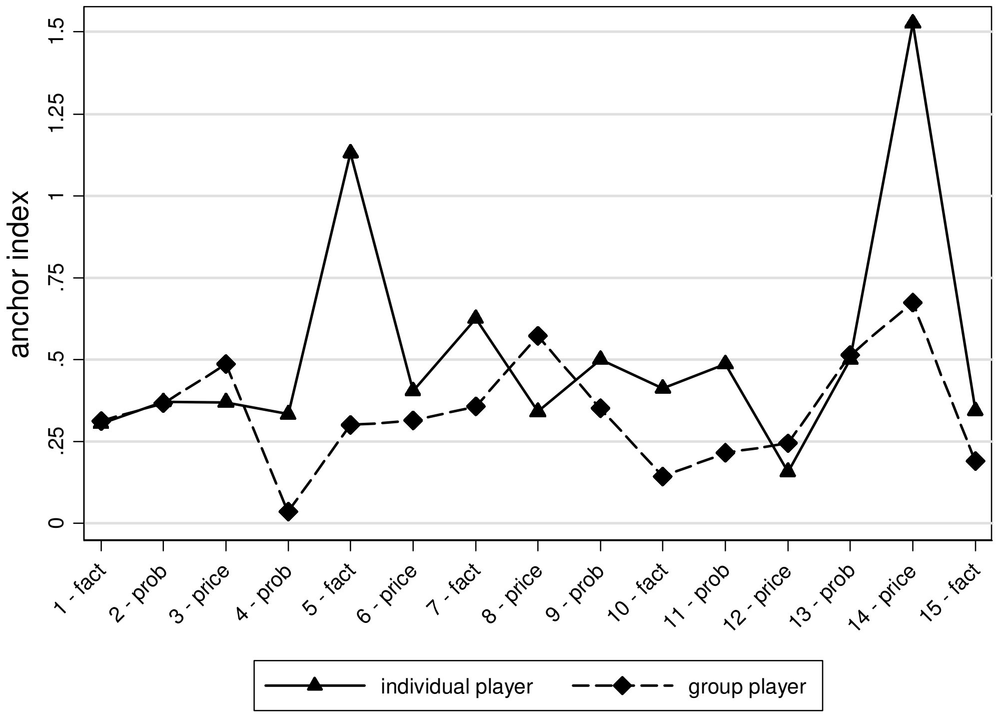
[9] E.g. for individuals players, the low anchor at exercise 7 is exactly 2121, for groups it is 1000. It cannot be ruled out that 2121 seems more or less plausible than a round number like 1000 per se, which might interfere with the anchoring bias. Accordingly, we round off thousands to hundreds and hundreds to tens.

[10] We measure accuracy by estimations' absolute deviations from correct values divided by these correct values to obtain absolute deviations in percentage points, which allow for pooling the results. For the five factual knowledge questions, this ratio on average amounts to .821 for individuals with a median of .563; for group players, the mean is .912 and the median .595. This difference is not statistically significant (Wilcoxon rank-sum test, z=.206, p=.8366).

in medians greater than the distance of the respective anchors. The anchor index values are shown for the 15 exercises in their actual order in Figure 1.

The graphs indicate an apparent anchoring bias in both individual and group players for all exercises. However, the magnitude of the bias varies over exercises. Figure 1 provides some indication that groups tend to be less biased, given that the overall average anchor index is .52 for individual players and only .34 for group players.

Figure 1. Anchor index (AI) for group and individual player



Note: “fact” refers to factual knowledge questions; “prob” refers to probability estimates and “price” to price valuations.

While this descriptive analysis can provide an initial impression that points to an anchoring bias for both individuals and groups, it is based merely on the median estimates and might thus obliterate important differences. We therefore analyze our results in detail in the following subsections by first explaining the anchor effectiveness index and subsequently applying it to the different exercise domains.

3.3 THE ANCHOR EFFECTIVENESS INDEX

As described by Jacowitz and Kahneman (1995) to allow statistical testing of the extent of the anchoring bias, we can range in estimates in *anchor* by assigning point values according to the corresponding percentile of the distribution of estimations in *calibration*: estimations equal to the median are assigned 50 points, estimations equal to the low (high) anchor 15 (85) points and so forth. If responses are below or above the bounds of the *calibration's* range of estimations, they transform to 0 or 100 points, respectively. This ordinal transformation procedure allows for pooling the results of varying exercises. Thereby, point values smaller than 50 indicate a downward bias and values greater than 50 point to an upward bias.[11]

To comprise the anchoring bias in both high and low anchor condition within a single key figure, we define the *anchor effectiveness index*, by which we measure the average deviation from *calibration's* median. We therefore calculate deviations depending on the anchor condition using the following procedure: in the high anchor condition, we subtract 50 from each prediction's assigned point value; while in the low anchor condition we subtract assigned point values from 50. The average of the corresponding deviations gives the *anchor effectiveness index (AEI)*.

Hence, an *AEI* of 0 indicates that the distribution of estimates is identical to the calibration treatment, given that positive and negative deviations from the median cancel out. A positive value hints at a systematic anchoring bias, which is increasing in strength for higher values, while a negative value would represent an asymmetrically biased behavior with respect to anchor values. Following this procedure, we can make a straightforward comparison of individual and group players' susceptibility toward anchors.

We additionally report *extreme values*, which are defined as estimations smaller (greater) than the anchors in the low (high) anchor condition. These values are relevant to more accurately identify the pattern of the anchoring bias (Jacowitz and Kahneman, 1995). Due to the definition of anchor values as the 15th and 85th percentile of estimations' distribution in *calibration*, a share of estimations greater than 15% exceeding these anchors would indicate that the bias moves some estimates above (below) the anchors that would not otherwise exceed these values. We can thus distinguish between an effect of the anchoring bias that merely causes estimations to be shifted towards the anchors and an effect that might be characterized as overshooting adjustment. Furthermore, we define an estimation to be an *outlier* if it is smaller (greater) in the

[11] Due to the distributions of estimations in the calibration treatments, we have to assign estimations in the anchor conditions to the closest available estimation value in *calibration* and match the respective point value. For example, if an estimated value of 20 forms the 20th percentile and an estimation of 30 the 30th percentile, values in the anchor condition smaller or equal to 25 are assigned 20 points, while values greater than 25 transform to 30 points.

low (high) anchor condition than the minimal (maximal) estimation of the respective calibration treatment. We thereby account for a shortcoming of the ordinal point transformation procedure: all values that are not within the distribution of the calibration treatments are uniformly assigned 0 or 100 points. Not considering these *outliers* could be misleading if their number differed between individuals and groups and they were additionally distributed asymmetrically to the average point values. Based on this procedure, we report the results for the three different task domains before turning to the perceived anchor relevance.

3.4 FACTUAL KNOWLEDGE

Table 4 summarizes the performance for the five factual knowledge questions that are also used for determining the payoffs. All tests presented are carried out by treating each group and individual player as one observation only.

	median estimation (median points)			**average points (std. dev.)**		**extreme values in % (outliers)**		***AEI***
	no anchor	**low anchor**	**high anchor**	**low anchor**	**high anchor**	**low anchor**	**high anchor**	
				Individuals				
E1	20	15 (37.5)	20 (62.5)	37.50 (20.23)	61.46 (26.01)	16.67 (0)	25 (0)	**11.98**
E5	6206	4350 (36.11)	10000 (100)	41.90 (23.96)	86.81 (18.88)	4.17 (0)	62.5 (62. 5)	**22.45**
E7	4980	3000 (26.32)	7000 (73.68)	33.99 (27.37)	67.54 (29.27)	29.17 (8.33)	33.33 (12.5)	**16.78**
E10	18000	17500 (50)	21293 (83.33)	51.39 (29.41)	78.47 (20.55)	20.83 (0)	33.33 (16.67)	**13.54**
E15	8000	8000 (61.11)	11750 (72.22)	56.94 (24.26)	68.98 (19.92)	4.17 (0)	8.33 (4.17)	**6.02**
total		**(37.5)**	**(75.73)**	**44.34 (26.27)**	**72.65 (24.58)**	**15 (1.67)**	**32.5 (19.17)**	**14.15**
				Groups				
E1	18	13 (30)	18 (50)	38.42 (25.44)	46.56 (14.57)	15.79 (15.79)	0 (0)	**4.71**
E5	7000	7000 (54.5)	8500 (72.23)	58.18 (23.94)	71.77 (19.61)	10 (0)	21.05 (15.79)	**11.44**
E7	5000	4000 (52.63)	6500 (68.42)	46.26 (18.36)	66.32 (22.87)	0 (0)	20 (0)	**10.19**
E10	18625	18000 (45)	19000 (60)	52.65 (25.93)	57.78 (25.04)	5.88 (5.88)	16.67 (0)	**2.71**
E15	9500	9000 (50)	11000 (63.64)	45.46 (25.92)	56.70 (23.10)	10 (10)	0 (0)	**5.59**
total		**(50)**	**(63.16)**	**47.00 (24.31)**	**60.35 (22.71)**	**8.24 (7.06)**	**11.96 (3.26)**	**6.82**

Table 4. Descriptive statistics for factual knowledge questions

The median estimations indicate that both group and individual players are prone to the anchoring bias. For all questions, given a low (high) anchor, players' median prediction is equal or smaller (greater) than the median in the calibration treatment. This finding corresponds to a gap in the transformed point averages and medians between the high and low anchor condition.

Individuals

When individuals were shown high anchor values, they estimated higher values relative to players in the calibration treatments, leading to point values greater than 50. There are some exceptions on the exercise level in the low anchor condition, as indicated by median point values not smaller than 50. Pooling the results for the five exercises for individual players, we find a systematic anchoring bias, i.e. a deviation from 50 points toward the anchor values, only in the high anchor condition (Sign test, one-sided, for low anchors $p=.1215$; for high anchors $p<.0001$). Nonetheless, there is as significant difference in point values between the high and low anchor condition (Wilcoxon signed-ranks test, $z=5.159$, $p<.0001$). Thus, the anchor values strongly bias individuals' estimates.

Groups

Considering group players, we find a significant bias in both conditions (Sign test, one-sided, for low anchors $p=.0717$; for high anchors $p=.0003$). Not surprisingly, estimates, as measured by point values, are significantly different between the low and high anchor condition (Wilcoxon signed-ranks test, $z=2.8899$, $p=.0039$). For both groups and individuals, high anchors prove to be more effective than low ones, which can be seen by the higher deviation from the 50 points representing unbiased behavior.

Group – individual differences

Overall, groups are apparently less biased, given that their point averages and medians fall within the interval established by those of individuals. The general susceptibility to the anchoring bias, as measured by the *AEI*, is lower for groups in all questions. In total, the *AEI* is only about half for group players (Wilcoxon rank-sum test, $z=2.967$, $p=.0030$).

The distribution of *extreme values* further supports the notion that group players are less biased. For low anchors, individual players' share of *extreme values* is almost doubled when compared to groups. For high anchors, it is tripled and significantly larger (Wilcoxon rank-sum test, for low anchors $z=1.209$, $p=.2266$; for high anchors $z=2.743$, $p=.0061$). Based on the definition of the anchor values and given the same behavior as observed in the calibration group, we should expect about 15% of estimations to be *extreme values*. For individuals in the high anchor condition, the share of 32.5% clearly indicates that the bias shifts estimations above the anchor value, which would otherwise be smaller than the anchor.[12] The higher share of *extreme values* in the high anchor condition for individuals supports the finding of more effective high anchors (Wilcoxon signed-ranks test, z=2.576, p=.0100). There is no such effect for group players, for whom the share of *extreme values* is identical in the high and low anchor condition (Wilcoxon signed-ranks test, z=.896, p=.3704).

While there are more *outliers* in the low anchor condition for groups (Wilcoxon rank-sum test, z=-1.848, p=.0645), there are more *outliers* for individuals in the high anchor condition (Wilcoxon rank-sum test, z=2.940, p=.0033). Most importantly, the distribution of outliers does not contradict the finding of less biased groups; on the contrary, the ordinal transformation rather weakens the strong differences in the anchoring bias. In total, there are about twice as many *outlier* estimates for individual players.

Economic relevance

We suggest that our findings translate to relevant differences in actual economic contexts. Recall that both groups and individuals tend to overestimate the correct answers for the factual knowledge questions. Individual players' absolute deviation in the low anchor condition amounts to 56.84%, which is not significantly smaller than 63.64% for groups (Wilcoxon rank-sum test, z=-1.252, p=.2105). In the high anchor condition, the stronger bias for individuals leads to a much higher average absolute deviation of 128.9% when compared to 93.43% for group players (Wilcoxon rank-sum test, z=1.852, p=.0640). Consequently, the stronger anchoring bias for individuals might have a positive effect if the anchor were to draw estimates toward the correct answers. By contrast, a highly negative effect on the accuracy results when the anchor draws estimations in the opposite direction. Groups' reluctance to adjust estimates toward anchor values renders their overall performance more robust. In sum, groups show significantly weaker anchoring bias. For factual knowledge questions, they are more resistant to adjust their responses towards the anchor values and much less prone to overshooting estimations.

Result 1: *Group cooperation reduces the anchoring bias for factual knowledge questions.*

[12] For theoretical consideration about the process underlying the overshooting adjustment according to the anchors, see e.g. Jacowitz and Kahneman (1995), who also find higher anchors to be more effective.

3.5 PROBABILITY ESTIMATES

Table 5 summarizes the main findings for probability estimates.

	median estimation (median points)			average points (std. dev.)		extreme values in % (outliers)		*AEI*
	no anchor	low anchor	high anchor	low anchor	high anchor	low anchor	high anchor	
				Individuals				
E2	52	45 (45.45)	71 (68.18)	45.45 (21.78)	67.80 (10.55)	8.33 (4.17)	4.17 (0)	**11.18**
E4	98	97 (54.17)	100 (100)	49.83 (23.43)	73.61 (34.02)	8.33 (0)	58.33 (0)	**11.89**
E9	61	40 (30.43)	70 (56.52)	30.98 (15.61)	61.59 (22.99)	12.5 (4.17)	16.67 (4.17)	**15.31**
E11	75	60 (20.83)	77 (60.42)	32.81 (25.40)	55.73 (27.74)	33.33 (0)	20.83 (8.33)	**11.46**
E13	48	25 (25)	53 (70.83)	34.55 (24.34)	66.15 (27.21)	16.67 (8.33)	25 (0)	**15.80**
total		**(37.5)**	**(65.155)**	**38.72 (23.25)**	**64.98 (25.99)**	**15.83 (3.33)**	**25 (2.5)**	**13.13**
				Groups				
E2	65	54 (45)	70 (60)	41.5 (27.53)	67.22 (28.81)	20 (20)	33.33 (22.22)	**12.63**
E4	97	98 (62.5)	98 (62.5)	59.17 (18.37)	71.49 (25.92)	0 (0)	36.84 (0)	**5.77**
E9	70	56 (29.17)	76 (62.5)	33.75 (23.33)	66.04 (22.71)	10 (0)	10 (5)	**16.15**
E11	80	79 (54.17)	85 (79.17)	47.29 (29.75)	66.01 (29.14)	15 (5)	31.58 (0)	**9.19**
E13	38	30 (41.67)	48 (75)	44.17 (20.92)	69.08 (20.09)	10 (5)	21.05 (0)	**12.29**
total		**(41.67)**	**(66.67)**	**45.18 (25.27)**	**67.96 (25.05)**	**11 (6)**	**26.32 (5.2)**	**11.22**

Table 5. Descriptive statistics for probability estimates

Again, both players' median estimations and point averages clearly indicate biased behavior in comparison to the calibration treatments.

Individuals

For the probability estimates, individual players' point values are again significantly different from 50. Accordingly, their estimates are biased (Sign test, one-sided, for low anchors $p<.0001$; for high anchors $p<.0001$). The point values for the high anchor condition significantly exceed the low ones (Wilcoxon signed-ranks test, $z=5.374$, $p<.0001$).

Groups

For group players, we find the same systematic deviation from the behavior in the calibration treatment (Sign test, one-sided, for low anchors p=.0717; for high anchors p<.0001) and significantly different transformed point values with respect to the anchor condition (Wilcoxon signed-ranks test, z=4.362, p<.0001).

Group – individual differences

For both anchor conditions, we find no difference between groups and individuals regarding the deviations from their respective calibration groups. Point values are not significantly different (Wilcoxon rank-sum test, for low anchors z=-1.316, p=.1882; for high anchors z=-.784, p=.4333). Groups are not significantly less biased, as can best be shown by the insignificant difference in the *AIE* (Wilcoxon rank-sum test, z=.360, p=.7186). In addition, the occurrence of *extreme values* is not significantly different between individual and group players (Wilcoxon rank-sum test, for low anchors z=.652, p=.5141; for high anchors z=-.510, p=.6098), while there is also no difference with respect to *outliers* (Wilcoxon rank-sum test, for low anchors z=-.99, p=.3223; for high anchors z=-1.109, p=.2672).

Overall, groups are not capable of avoiding the anchoring bias more effectively when compared to individual players for probability estimations.

Result 2: *Group cooperation does not reduce the anchoring bias for probability estimations.*

3.6 PRICE VALUATIONS

Table 6 presents the results for the five price valuations.

	median estimation (median points)			average points (std. dev.)		extreme values in % (outliers)		*AEI*
	no anchor	low anchor	high anchor	low anchor	high anchor	low anchor	high anchor	
				Individuals				
E3	155	150 (50)	304 (79.17)	49.13 (25.30)	77.26 (17.16)	8.33 (8.33)	12.5 (8.33)	**14.06**
E6	120	100 (41.67)	150 (66.67)	39.41 (23.41)	64.76 (19.19)	25 (0)	12.5 (4.17)	**12.67**
E8	40	35 (45.83)	50 (66.67)	44.10 (26.58)	69.97 (20.37)	29.17 (4.17)	29.17 (0)	**12.93**
E12	184	168 (47.92)	199 (62.5)	49.65 (25.12)	59.20 (23.47)	16.67 (0)	8.33 (0)	**4.77**
E14	1778	473 (20.83)	20000 (91.67)	29.34 (26.96)	86.98 (12.12)	25 (4.17)	62.5 (0)	**28.82**
total		**(45.83)**	**(79.17)**	**42.33** **(26.17)**	**71.63** **(20.94)**	**20.83** **(3.33)**	**25** **(2.5)**	**14.65**
				Groups				
E3	252	165 (29.17)	500 (79.17)	38.13 (15.72)	79.86 (10.23)	5 (0)	11.11 (5.56)	**20.39**
E6	115	88 (35.42)	130 (66.67)	36.67 (19.38)	64.69 (19.81)	15 (5)	10.53 (0)	**13.40**
E8	42	30 (25)	50 (75)	36.46 (24.78)	65.00 (20.43)	50 (0)	10 (0)	**14.27**
E12	175	161 (45.83)	200 (70.83)	47.71 (26.88)	68.75 (21.01)	20 (0)	15 (0)	**10.52**
E14	8000	925 (23.81)	42500 (80.95)	29.76 (13.85)	68.25 (31.84)	0 (0)	33.33 (0)	**19.19**
total		**(29.17)**	**(76.19)**	**38.08** **(21.32)**	**69.16** **(21.93)**	**18.75** **(1.04)**	**15.79** **(1.05)**	**15.52**

Table 6. Descriptive statistics for price valuations

For all exercises, the median valuations of the calibration treatments fall within the range of medians for the low and high anchor condition. Groups and individuals are both clearly biased by the anchor values.

Individuals

Individual players' point values systematically deviate from 50 for both anchor conditions (Sign test, one-sided, for low anchors p=.0129; for high anchors p<.0001). Point values in the high anchor condition are again significantly greater than in the low one (Wilcoxon signed-ranks test, z=5.641, p<.0001).

Groups

For groups, we find the same pattern, given that point values are significantly different from 50 (Sign test, one-sided, for low anchors p<.0001; for high anchors p<.0001) and there is a difference in point values between the high and low anchor condition (Wilcoxon signed-ranks test, z=5.162, p<.0001).

Group – individual differences

While the high anchors again seem to be more effective than the low ones, there are no systematic differences between group and individual players for both anchor conditions (Wilcoxon rank-sum test, for low anchors z=1.136, p=.2561; for high anchors z=.578, p=.5630). Moreover, there are no significant differences with respect to the distribution of *extreme values* (Wilcoxon rank-sum test, for low anchors z=.136, p=.8915; for high anchors z=1.533, p=.1252) nor for *outliers* (Wilcoxon rank-sum test, for low anchors z=.842, p=.4098; for high anchors z=.848, p=.3967). Consequently, these findings are reflected in the (almost) equal *AEI* for groups and individuals (Wilcoxon rank-sum test, z=-.616, p=.5379).

Overall, there is no evidence showing that groups are less biased than individuals in a price valuation task.

Result 3: *Group cooperation does not reduce the anchoring bias for price valuations.*

3.7 PERCEPTION OF THE ANCHOR VALUES

Finally, we consider the perceived anchor relevance as an explanatory variable for differences across treatments. We therefore asked all participants in treatment groups to rate how relevant they perceived the anchor values to be, on a scale from 1 to 9. The results might give some indication as to how anchors were processed by the players and whether the anchors' potentially misleading effect is acknowledged more appropriately by groups rather than by individuals. However, we find no significant differences between subjects playing individually and those who had been assigned to a group. Please note that all group members were asked individually. For individual (group) players, 33.33% (37.5%) of participants perceived the anchor values as not relevant at all, indicated by a scale value of 1. The median for both player types is equal to 5; the mean for individual player is 4.125 and 3.979 for group players. These small difference are not significant (Wilcoxon rank-sum test, z=-.358, p=.7207), which gives the indication that individual and group players did not vary in the perceived relevance of the anchors. Therefore, group decision-making does not seem to lead to greater awareness concerning the irrelevance of the anchor values.

Result 4: *Group cooperation does not reduce the perceived relevance of the anchor values.*

Summing up, we show that groups do not ubiquitously reduce the anchoring bias; rather, the task characteristics have to be taken into account. For factual knowledge questions, groups are evidently less biased, while there is no such evidence for probability estimates or price valuations.

4. CONCLUSION

In this study, we set out to investigate whether groups are able to reduce the anchoring bias. Given the large evidence of superior group rationality when compared to individual decision-making, it appears reasonable to assume that groups are able to avoid individually persistent biases. This finding would in fact reduce the relevance of heuristics and biases in actual situations, given that a substantial part of decisions in economic domains are taken collaboratively. However, no study in experimental economics to date has made a direct comparison between individual and group performance with a focus on bias-reduction. We therefore present an anchoring design similar to the majority of psychological anchoring studies, yet implement strong monetary incentives and group decision-making.

While groups are biased by the anchor, they are in fact able to reduce anchoring in the domain of factual knowledge. By contrast, for probability estimations and price valuations, groups are equally biased by the external anchors as individual players. Thus, stating that groups are generally less affected by behavioral biases is not accurate. Rather, the group's ability to debias decisions primarily depends on the task characteristics.

Our results can be interpreted when drawing on the differentiation between intellective and judgmental tasks (Cox and Hayne, 2006). It is a common result in psychological small group research that groups primarily outperform individuals in tasks that have easily demonstrable correct solutions (Laughlin et al., 2002). For tasks requiring judgment that goes beyond straightforward intellective reasoning, group performance tends to approach the level of average individuals. Apparently, in rather intellective tasks such as factual knowledge questions, groups' enhanced cognitive performance enables them to more successfully refrain from external anchors. Once judgmental aspects are involved, as with probability estimates or price valuations, the individual adherence to anchors is reproduced during the group decision process. Overall, the ubiquitous character of the anchoring bias can be asserted, even given group cooperation and monetary incentives. Our results suggest that the relevance of biases in market contexts involving groups cannot be negated altogether. Rather, a differentiated view is required, considering group performance as conditional on specific domains of decision-making. This might more closely show the robustness of heuristics and biases in various market contexts.

APPENDIX A

Instructions for calibration and anchor treatments. Note that the instructions for the respective calibration and anchor treatments are similar. The differing information for the respective group experiments are indicated in braces.

The Game

In this game, you will answer 15 questions {*along with two other players who will be assigned randomly to you at the beginning of the game*}. In each period, you will have ninety seconds to enter your answer. {*You must enter a common answer within your group. To find a common answer, you will communicate with your group members via chat. If your group members enter different solutions, you will have an additional thirty seconds to find a common answer. If you fail to do so, you will receive no payment for this answer.*} If you do not enter a solution, you will receive no payment for this answer.

Your Payoff

Your payoff depends on how close your {*group's*} answer is to the correct value, compared to all other players {*groups*}. However, only 5 of the 15 questions are relevant for your payment. You do not know which these 5 questions are. The payments are calculated as follows:

First to third best answer {*best group-answer*}	25 ECU
Forth to sixth best answer {*second best group-answer*}	20 ECU
Seventh to ninth best answer {*third best group-answer*}	15 ECU
All other answers {*group-answer*}	10 ECU

10 ECU converts to €1. Your payments in every period will be summed up and paid to you after the game. {*You will receive an additional basic payment of €2.5.*}

APPENDIX B

Pictures of used articles that were shown to subjects for the price valuations.

i) question no. 3

ii) question no. 6

iii) question no. 8

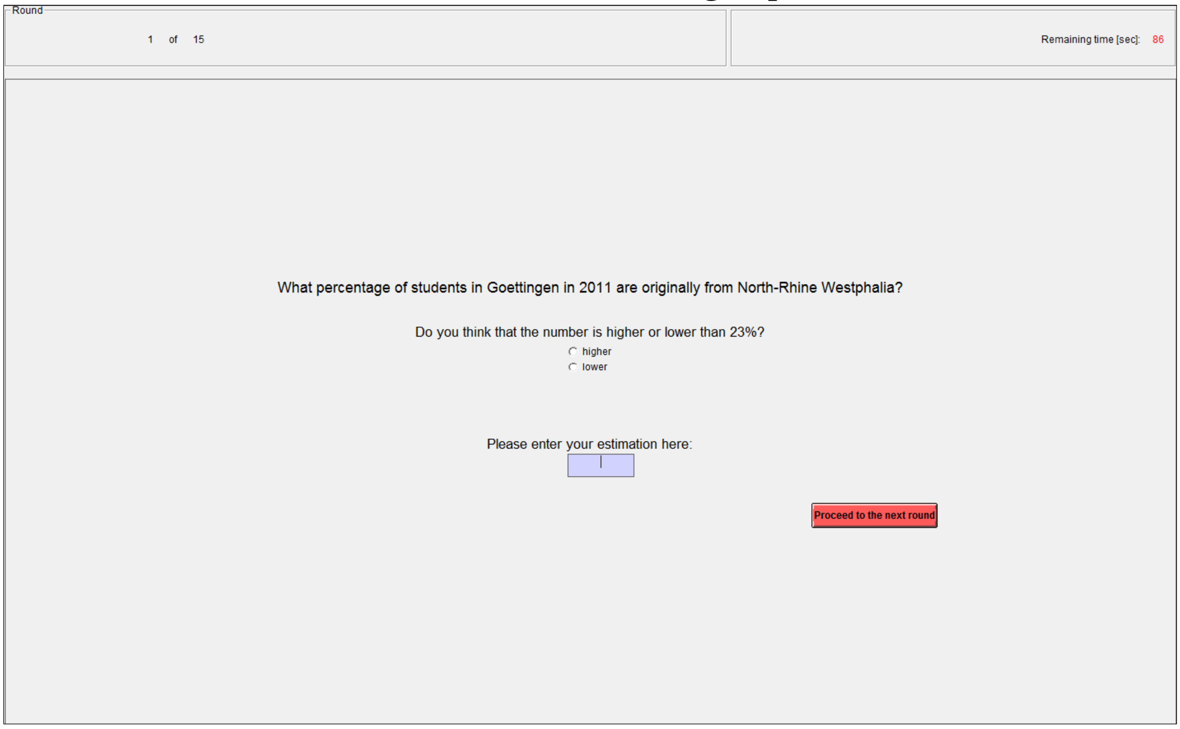
iv) question no. 12

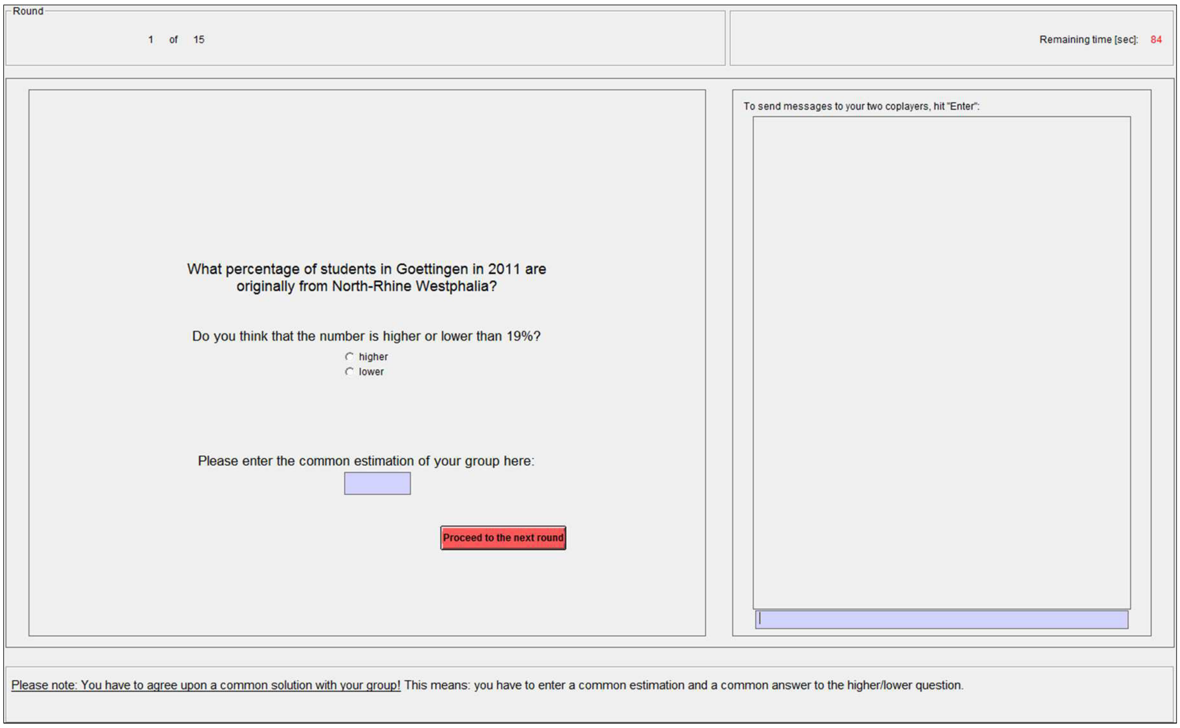
v) question no. 14

APPENDIX C

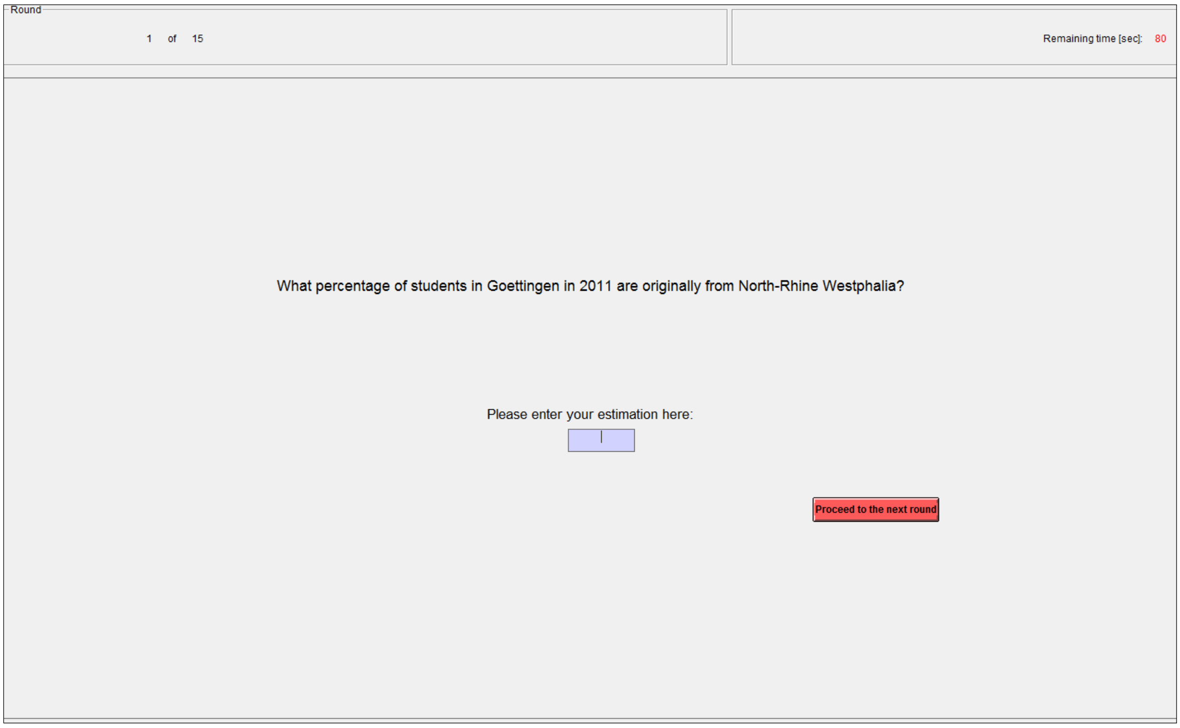
Screenshots of the main game interface for groups and individuals. Note that the calibration treatments only differs with respect to the higher/lower question and the anchor value at the center of the decision screen. All other aspects between calibration and anchor treatments are identical.

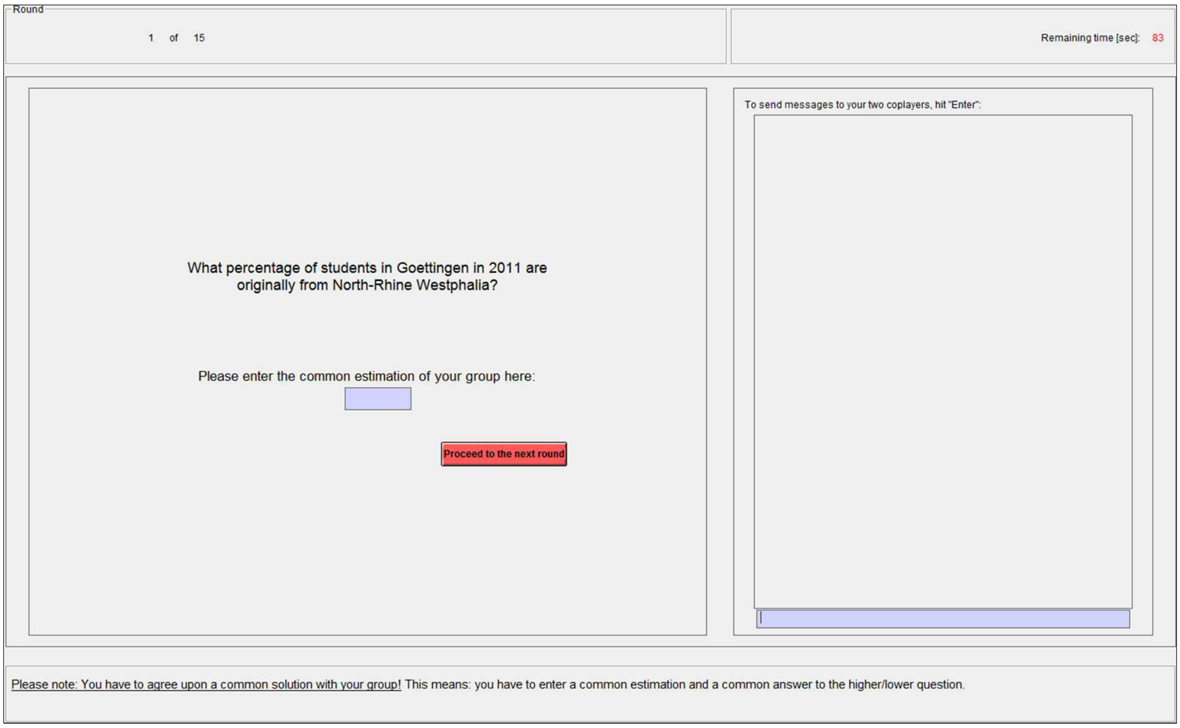
Screenshots for the anchor treatments (individual / groups)





Screenshots for the calibration treatments (individual / groups)





REFERENCES

Adaval, R., Wyer, R.S. (2011). Conscious and Nonconscious Influences of a Price Anchor: Effects on Willingness to Pay for Related and Unrelated Products. *Journal of Marketing Research* 48, 355-365.

Alevy, J.E., Landry, C.E., List, J.A. (2015). Field Experiments on the Anchoring of Economic Valuations. *Economic Inquiry* 53 (3), 1522-1538. doi: 10.1111/ecin.12201.

Bateman, I.J., Burgess, D., Hutchinson, W.G., Matthews, D.I. (2008). Learning Design Contingent Valuation (LDCV): NOAA Guidelines, Preference Learning, and Coherent Arbitrariness. *Journal of Environmental Economics and Management* 55, 127-41. doi: 10.1016/j.jeem.2007.08.003.

Beggs, A., Graddy, K. (2009. Anchoring effects: evidence from art auctions. *American Economic Review* 99 (3), 1027-1039. doi: 10.1257/aer.99.3.1027.

Blankenship, K.L., Wegener, D.T., Petty, R.E., Detweiler-Bedell, B., Macy, C.L. (2008). Elaboration and consequences of anchored estimates: an attitudinal perspective on numerical anchoring. *Journal of Experimental Social Psychology* 44, 1465-1476. doi: 10.1016/j.jesp.2008.07.005.

Bofinger, P., Schmidt, R. (2003). On the reliability of professional exchange rate forecasts: an empirical analysis for the €/US-$ rate. *Financial Markets and Portfolio Management* 17, 437-449. doi: 10.1007/s11408-003-0403-z.

Bucchianeri, G.W., Minson, J. (2013). A homeowner's dilemma: Anchoring in residential real estate transactions. *Journal of Economic Behavior & Organization* 89, 76-92. doi: 10.1016/j.jebo.2013.01.010.

Campbell, S.D., Sharpe, S.A. (2009). Anchoring bias in consensus forecasts and its effect on market prices. *Journal of Financial and Quantitative Analysis* 44, 369-390. doi: 10.1017/S0022109009090127.

Cecchi, F., Bulte, E. (2013). Does market experience promote rational choice? Experimental evidence from rural Ethiopia. *Economic Development and Cultural Change* 61 (2), 407-429. doi: 10.1086/668275.

Cen, L., Hilary, G., Wei, K.C.J. (2013). The Role of Anchoring Bias in the Equity Market: Evidence from Analysts' Earnings Forecasts and Stock Returns. *Journal of Financial and Quantitative Analysis* 48, 47-76. doi:10.1017/S0022109012000609.

Chapman, G.B., Johnson, E.J. (1999). Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes* 79, 115-153. doi:10.1006/obhd.1999.2841

Charness, G., Sutter, M. (2012). Groups make better Self-Interested Decisions. *Journal of Economic Perspectives* 26 (3), 157-176. doi: 10.1257/jep.26.3.157.

Cox, J., Hayne, S. (2006). Barking up the right tree: are small groups rational agents?. *Experimental Economics* 9, 209-222. doi: 10.1007/s10683-006-9123-3.

Critcher, C.R., Gilovich, T. (2008). Incidental environmental anchors. *Journal of Behavioral Decision Making* 21, 241-251. doi: 10.1002/bdm.586.

Critcher, C.R., Rosenzweig, E.L., Gilovich, T. (2014). Beyond Anchoring: Contextual Influence on Adjustment. *University of California, Haas School of Business Working Paper*.

Dodonova, A., Khoroshilov, Y. (2004). Anchoring and transaction utility: evidence from online auctions. *Applied Economic Letters* 11, 307-310. doi: 10.1080/1350485042000221571.

Epley, N., Gilovich, T. (2005). When effortful thinking influences judgmental anchoring: differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making* 18, 199-212. doi: 10.1002/bdm.495.

Epley, N., Gilovich, T. (2010). Anchoring unbound. Journal of Consumer Psychology 20 (1), 20-24. doi: 10.1016/j.jcps.2009.12.005.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10, 171-178.

Frederick, S., Kahneman, D., Mochon, D. (2010). Elaborating a simpler theory of anchoring. *Journal of Consumer Psychology*, 20 (1), 17-19. doi: 10.1016/j.jcps.2009.12.004.

Frederick, S.W., Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General* 141 (1), 124-133. doi: 10.1037/a0024006.

Fudenberg, D., Levine, D.K., Maniadis, Z. (2012). On the robustness of anchoring effects in WTP and WTA experiments. *American Economic Journal: Microeconomics* 4, 131-145. doi: 10.1257/mic.4.2.131.

Fujiwara, I., Ichiue, H., Nakazono, Y., Shigemi, Y. (2013). Financial markets forecasts revisited: Are they rational, stubborn or jumpy?. *Economics Letters* 118 (3), 526-530. doi: 10.1016/j.econlet.2012.12.037.

Furnham, A., Boo, H.C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics* 40, 35-42. doi: 10.1016/j.socec.2010.10.008.

Gloeckner, A., Englich, B. (2014). When Relevance Matters. Anchoring Effects can be larger for relevant than for irrelevant anchors. *Social Psychology* 46(1), 4-12. doi: 10.1027/1864-9335/a000214.

Greiner, B. (2004). An online recruiting system for economic experiments. In: Kremer, K., Macho, V. (Eds.), Forschung und wissenschaftliches Rechnen 2003. *GWDG Bericht 63. Ges. für Wiss. Datenverarbeitung, Goettingen*, 79-93.

Hess, D., Orbe, S. (2013). Irrationality or efficiency of macroeconomic survey forecasts? Implications from the anchoring bias test. *Review of Finance* 17 (6), 2097-2131. doi: 10.1093/rof/rfs037.

Hinsz, V.B., Indahl, K.E. (1995). Assimilation to anchors for damage awards in a mock civil trial. *Journal of Applied Social Psychology* 25, 991-1026. doi: 10.1111/j.1559-1816.1995.tb02386.x.

Jacowitz, K.E., Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin* 21, 1161-1166. doi:10.1177/01461672952111004.

Johnson, J.E.V., Schnytzer, A., Liu, S. (2009). To what extent do investors in a financial market anchor their judgements excessively? Evidence from the Hong Kong horserace betting market. *Journal of Behavioral Decision Making* 22, 410-434. doi: 10.1002/bdm.640.

Kerr, N.L., Tindale, R.S. (2004). Group performance and decision making. *Annual Review of Psychology* 55, 623-655. doi: 10.1146/annurev.psych.55.090902.142009.

Kugler, T., Kausel, E., Kocher, M. (2012). Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley Interdisciplinary Reviews: Cognitive Science* 3 (4), 471-482. doi: 10.1002/wcs.1184.

Laughlin, P.R. (1980). Social combination processes of cooperative problem-solving groups on verbal intellective tasks, in: Fishbein, M. (Ed.), *Progress in Social Psychology*, Erlbaum, Hillsdale NJ, 127-155.

Laughlin, P.R., Bonner, B.L., Miner, A.G. (2002). Groups perform better than the best individuals on Letter-to-Numbers problems. *Organizational Behavior and Human Decision Processes* 88, 606-620.

Lawrence, M., O'Connor, M. (2000). Sales forecasting updates: how good are they in practice?. *International Journal of Forecasting* 16 (3), 369-382. doi: 10.1016/S0169-2070(00)00059-5.

Levitt, S.D., List, J.A. (2007). What do laboratory experiments measuring social preferences reveal about the real world?. *The Journal of Economic Perspectives*, 21 (2), 153-174. doi: 10.1257/jep.21.2.153.

List, J.A. (2003). Does Market Experience Eliminate Market Anomalies?. *Quarterly Journal of Economics* 118, 41-71. doi: 10.1162/00335530360535144.

List, J.A. (2004a). Testing neoclassical competitive theory in multilateral decentralized markets. *Journal of Political Economy*, 112 (5), 1131-1156. doi: 10.1086/422564.

List, J.A. (2004b). Neoclassical theory versus prospect theory: Evidence from the marketplace. *Econometrica*, 72 (2), 615-625. doi: 10.1111/j.1468-0262.2004.00502.x.

List, J.A. (2006). The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *Journal of Political Economy*, 114 (1), 1-37. doi: 10.1086/498587.

List, J.A., Millimet, D.L. (2008). The market: Catalyst for rationality and filter of irrationality. *The B.E. Journal of Economic Analysis & Policy* 8, 1935-1682. doi: 10.2202/1935-1682.2115.

Luhan, W.J., Kocher, M.G., Sutter, M. (2009). Group polarization in the team dictator game reconsidered. *Experimental Economics* 12, 26-41.

Maniadis, Z., Tufano, F., List, J.A. (2014). One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects. *The American Economic Review* 104 (1), 277-290. doi: 10.1257/aer.104.1.277.

McAlvanah, P., Moul C.C. (2013). The House Doesn't Always Win: Evidence of Anchoring Among Australian Bookies, *Journal of Economic Behavior & Organization* 90, 87-99. doi: 10.1016/j.jebo.2013.03.009.

McElroy, T., Dowd, K. (2007). Susceptibility to anchoring effects: how openness-to-experience influences responses to anchoring cues. *Judgment and Decision Making* 2 (1), 48-53.

Meub, L., Proeger, T. (2015). Anchoring in social context. *Journal of Behavioral and Experimental Economics* 55, 29-39. doi: 10.1016/j.socec.2015.01.004.

Meub, L., Proeger, T. (2015). Can anchoring explain biased forecasts? *CEGE Discussion Paper*, No. 166.

Nunes, J.C., Boatwright, P. (2004). Incidental prices and their effect on willingness to pay. *Journal of Marketing Research* 41 (4), 457-466. doi: 10.1509/jmkr.41.4.457.47014.

Plous, S. (1989). Thinking the unthinkable: the effects of anchoring on likelihood estimates of nuclear war. *Journal of Applied Social Psychology* 19, 67-91. doi: 10.1111/j.1559-1816.1989.tb01221.x.

Russo, J.E. (2010). Understanding the effect of a numerical anchor. *Journal of Consumer Psychology* 20 (1), 25-27. doi: 10.1016/j.jcps.2009.12.006.

Simmons, J.P., LeBoeuf, R.A., Nelson, L.D. (2010). The Effect of Accuracy Motivation on Anchoring and Adjustment: Do People Adjust From Provided Anchors?. *Journal of Personality and Social Psychology* 99, 917-932. doi: 10.1037/a0021540.

Simonson, I., Drolet, A. (2004). Anchoring Effects on Consumers' Willingness-to-Pay and Willingness-to-Accept. *Journal of consumer research* 31 (3), 681-90. doi: 10.1086/425103.

Sniezek, J.A., Henry, R.A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes*, 43 (1), 1-28. doi: 10.1016/0749-5978(89)90055-1.

Stahlberg, D., Eller, F., Maass, A., Frey, D. (1995). We knew it all along: Hindsight bias in groups. *Organizational Behavior and Human Decision Processes* 63 (1), 46-58. doi: 10.1006/obhd.1995.1060.

Sugden, R., Zheng, J., Zizzo, D.J. (2013). Not all anchors are created equal. *Journal of Economic Psychology* 39, 21-31. doi: 10.1016/j.joep.2013.06.008.

Tufano, F. (2010). Are 'True' Preferences Revealed in Repeated Markets? An Experimental Demonstration of Context-Dependent Valuations. *Experimental Economics* 13 (1), 1-13. doi: 10.1007/s10683-009-9226-8.

Tversky, A., Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124-1131. doi: 10.1126/science.185.4157.1124.

Wegener, D.T., Petty, R.E., Blankenship, K.L., Detweiler-Bedell, B. (2010a). Elaboration and numerical anchoring: implications of attitude theories for consumer judgment and decision making. *Journal of Consumer Psychology* 20, 5-6. doi: 10.1016/j.jcps.2009.12.003.

Wegener, D.T., Petty, R.E., Blankenship, K.L., Detweiler-Bedell, B. (2010b). Elaboration and numerical anchoring: Breadth, depth, and the role of (non-) thoughtful processes in anchoring theories. *Journal of Consumer Psychology* 20 (1), 28-32. doi: 10.1016/j.jcps.2009.12.007.

Whyte, G., Sebenius, J.K. (1997). The effect of multiple anchors on anchoring in individual and group judgment. *Organizational Behavior and Human Decision Processes* 69 (1) 74-85. doi: 10.1006/obhd.1996.2674.

Wildschut, T., Pinter, B., Vevea, J.L., Insko, C.A., Schopler, J. (2003). Beyond the group mind: a quantitative review of the interindividual-intergroup discontinuity effect. *Psychological Bulletin 129* (5), 698-722. doi: 10.1037/0033-2909.129.5.698.

Wilson, T.D., Houston, C.E., Etling, K.M., Brekke, N. (1996). A new look at anchoring effects: basic anchoring and its antecedents. *Journal of Experimental Psychology* 125, 387-402. doi: 10.1037/0096-3445.125.4.387.

Wright, W.F., Anderson, U. (1989). Effects of situation familiarity and financial incentives on use of the anchoring and adjustment heuristic for probability assessment. *Organizational Behavior and Human Decision Processes* 44, 68-82. doi: 10.1016/0749-5978(89)90035-6.