

ECONOMIC FORECASTING
EXPERIMENTAL EVIDENCE ON INDIVIDUAL BIASES,
GROUP COOPERATION AND HERD BEHAVIOR

Dissertation

zur Erlangung des Doktorgrades
der Wirtschaftswissenschaftlichen Fakultät
der Georg-August-Universität Göttingen

vorgelegt von

Lukas Meub

geboren in Lich

Göttingen, 2014

Erstgutachter: Prof. Dr. Kilian Bizer
Zweitgutachter: Prof. Dr. Markus Spiwoks

Weiteres Mitglied der Prüfungskommission: Prof. Dr. Claudia Keser

CONTENTS

1	INTRODUCTION AND SUMMARY	1
2	ANCHORING: A VALID EXPLANATION FOR BIASED FORECASTS WHEN RATIONAL PREDICTIONS ARE EASILY ACCESSIBLE AND WELL INCENTIVIZED? <i>(with Till Proeger and Kilian Bizer)</i>	21
3	AN EXPERIMENTAL STUDY ON SOCIAL ANCHORING <i>(with Till Proeger)</i>	49
4	ARE GROUPS LESS BEHAVIORAL? THE CASE OF ANCHORING <i>(with Till Proeger)</i>	75
5	OVERCONFIDENCE AS A SOCIAL BIAS: EXPERIMENTAL EVIDENCE <i>(with Till Proeger)</i>	109
6	STRATEGIC COORDINATION IN FORECASTING: AN EXPERIMENTAL STUDY <i>(with Kilian Bizer, Till Proeger and Markus Spiwoks)</i>	123
7	A COMPARISON OF ENDOGENOUS AND EXOGENOUS TIMING IN A SOCIAL LEARNING EXPERIMENT <i>(with Till Proeger and Hendrik Hüning)</i>	169
8	THE IMPACT OF COMMUNICATION REGIMES ON GROUP RATIONALITY: EXPERIMENTAL EVIDENCE <i>(with Till Proeger)</i>	201

Chapter I

INTRODUCTION AND SUMMARY

“Much has been written about the doubtful accuracy of economists’ predictions. [...] They are better at predicting the direction than the actual magnitude of events. [...] This is disappointing, but it does not mean that economics is not a science.”
(‘Economics’, Encyclopædia Britannica¹)

The forecasting of economic developments guides decisions in all areas of public interest, whether in governments, finance or corporations.² Accurate predictions hold an essential role for private and public actors’ ability to enable the efficient planning and execution of their strategic objectives. Accordingly, the development and evaluation of models to enhance accurate forecasting has always been among the central areas of economic research interest. The rapid progress in information technology in recent times has intensified the search for the empirically sound modelling of economic developments. However, this “*virtual revolution in how economists compute, apply, and evaluate forecasts*” (Elliot and Timmermann, 2008, p.3) has not altered the deplorable finding that naïve “no change” forecasts regularly outperform professional predictions in terms of their accuracy (Spiwoks, 2004).

The basic neoclassical explanation for this empirical finding is markets’ information efficiency. Following the efficient market hypothesis, forecasting is futile by definition, given that asset prices incorporate new information perfectly as soon as it becomes accessible. With current prices mirroring all available information, forecasting – necessarily based on private information – is either obsolete or a gamble on future events independent of actual developments. Despite being a constant companion during the development of neoclassical economics, the evident criticism of the axioms underlying market efficiency has received additional attention lately.³ Following the recent macroeconomic turmoil, neoclassical assumptions have been criticized in favor of a behavioral understanding of decision-making. Criticism on the deterministic perspective of rational expectations has opened up a broad field of research that enables a more realistic understanding regarding agents’ incentives, as well as the institutions and processes involved in economic forecasting.⁴ Consequently, behavioral evidence from laboratory studies has become influential for the interpretation of potentially

¹ Quoted from Ottaviano and Sorensen (2006a, p. 120).

² Please note that this introduction summarizes the studies that constitute the dissertations of both Till Proeger and Lukas Meub. Thus, while the shared topic of research is discussed from different angles in the respective introductions, the chapters’ summaries are necessarily closely related.

³ Comprehensive overviews of the development of discussions on the efficient market hypothesis are provided by Fama (1970; 1991) and Spiwoks (2002); a critical account on its influence in the years leading up to the recent financial crisis is given by Krugman (2009).

⁴ For a general overview on the development of behavioral economics and its connections to neoclassical economics, please refer to Berg and Gigerenzer (2010).

dysfunctional markets for forecasts, as it yields a central advantage, in that analyzing actual time series always leaves open the theoretical possibility that low forecasting accuracy is due to markets' information efficiency – an argument that can logically never be dismissed. At the same time, there are distinct behavioral approaches equally suitable to explain the status quo bias in forecasting, that also cannot be tested explicitly using empirical forecasting data. Introducing laboratory experiments to forecasting research firstly enables the construction of decision situations in which achievable outcomes and optimal behavior can be unambiguously benchmarked. Changing distinct parameters subsequently allows the identification and weighing of determinants for forecasting quality compared to the predefined benchmark. Thus, experiments enable a clean *ceteris paribus* analysis of the determinants of behavior in forecasting, which otherwise remains inaccessible.

In this book, the authors pursue particular aspect of this broad research program. It has been shown that many forecasting time series share the characteristic tendency of predictions being biased towards current values. This implies that a major share of forecasters rely on the strategy of making (quasi-) naïve predictions. Denominated as a “status quo bias in forecasting” (Gubaydullina et al., 2011), this finding has been confirmed in numerous empirical studies.⁵ Naturally, merely forecasting present states precludes the revelation of private expectations on future states and accordingly disqualifies the respective predictions as a useful means for planning economic activities. To investigate the factors determining this finding, the authors build on two of the major behavioral explanations for poor forecasting quality: individual heuristics and biases, as well as rational herding.

Heuristics and Biases

The first paradigm drawn upon in this book is the psychological research on biases and heuristics and its application to forecasting. While psychological research considers numerous heuristics,⁶ the anchoring bias is most prominent when behavioral anomalies in forecasting are discussed.⁷ Tversky and Kahnemann (1974) were the first to show that individuals'

⁵ Among the recent examples are Welch (2000), Gallo et al. (2002), Bofinger und Schmidt (2003), Spiwoks (2004), Clement and Tse (2005), Batchelor (2007), Spiwoks und Hein (2007), Spiwoks et al. (2008; 2010), Ager et al. (2009), Jegadeesh and Kim (2009), as well as Gubaydullina et al. (2011).

⁶ See Harvey (2007) for a comprehensive review of forecasting research in psychology.

⁷ For applications see e.g. real estate price forecasts (Bucchianeri and Minson 2013), financial forecasts (Fujiwara et al. 2013), sports betting (Johnson et al. 2009; McAlvanah and Moul 2013), earnings forecasts (Cen et al. 2013), macroeconomic forecasts (Campbell and Sharpe 2009; Hess and Orbe 2013), as well as sales forecasting (Lawrence and O'Connor 2000).

assessments can be systematically influenced by random numbers, which are completely irrelevant for the respective tasks. A large body of psychological experiments has since addressed the validity of this result across specific settings, leading to the conclusion that anchoring is “*exceptionally robust, pervasive and ubiquitous*” (Furnham and Boo, 2011, p. 41). Transferring this finding to forecasting, it is claimed that forecasters “*use the last data point in the series as a mental anchor and then adjust away from that anchor to take account of the major feature(s) of the series. However, as adjustment is typically insufficient, their forecasts are biased.*” (Harvey, 2007, p.17). Accordingly, it is argued that forecasts clustered around current values do not necessarily reflect strategic decisions, but rather a subconscious, irrational bias towards the current state, which cannot be alleviated through contradictory incentives or alternative market designs.

A second behavioral anomaly that has regularly been considered, particularly in the context of financial forecasting, is the overconfidence bias. The term refers to individuals’ systematic inability to realistically evaluate their capabilities and the resulting tendency of overly optimistic self-assessments across a wide variety of decision situations. As with anchoring, the bias has been comprehensively investigated in experimental psychology and has also been applied to economic contexts for some time.⁸ While assuming robust overconfidence among analysts is not considered as a reason for biased forecasts, it can serve to explain the continuation of forecasting despite their poor accuracy. Self-denial of the obvious failure to produce correct predictions, fueled by overconfidence, might hold considerable influence on the persistence of forecasts’ uniformity.⁹ Overconfidence can thus be interpreted as a psychological mechanism that individually reinforces and justifies uninformative predictions.

Rational Herding Behavior: Reputation and Cascades

The second paradigm built upon in this book is forecasters’ herding on public information as another explanation for uninformative homogenous predictions. Within this paradigm, two distinct lines of reasoning applicable to herding in forecasting have been considered. The first one, based on the seminal model by Scharfstein and Stein (1990), emphasizes reputational

⁸ An overview of the discussions within experimental psychology is given by Moore and Healy (2008). Examples for the application to economic contexts include e.g. investor experience (Menkhoff et al., 2013), CEOs’ overconfidence and investment (Malmendier and Tate, 2005), gender and investment (Barber and Odean, 2001), monetary policy decisions (Claussen et al., 2012), financial trading (Biais et al., 2005) and pricing decisions for consumer goods (Grubb, 2009).

⁹ A notable exception to this effect is given by the DIW, who in 2009 admitted to be incapable of making a prediction for the economic development in 2010 due to the complexity of the financial crisis (Plickert, 2009).

concerns that prompt analysts to herd.¹⁰ By making homogenous predictions and conforming to consensus values, forecasters are assumed to generate a “sharing-the-blame” effect. By adapting to the current market consensus, individual analysts appear well informed and ultimately more capable; at the same time, false forecasts can easily be explained as a community-wide inability to anticipate future developments. Overall, reputational herding explains the status quo bias in forecasting as a result of rational, strategic coordination, which serves to uphold a favorable individual reputation. This of course leads to an incomplete overall information aggregation as “*the desire of analysts to be perceived as good forecasters turns them into poor forecasters*” (Ottaviani and Sorensen, 2006a, p.443). Experimental studies by Cote and Sanders (1997), Ackert et al. (2008) and Bedke et al. (2009) have taken up this concept, with the empirical evidence emphasizing the strong influence of reputational concerns on herding behavior.

The second explanation within the paradigm of rational herding is the concept of informational cascades, whose fundamental difference to reputational herding lies in the assumption that analysts’ honestly strive for correct predictions. Following the seminal models by Welch (1992), Bikhchandani et al. (1992) and Banerjee (1992), it has been pointed out that discarding private information can become rational given that prior public information is acknowledged and processed appropriately. Hence, conforming to the preceding consensus might become the rational strategy for agents trying to truthfully predict future states. These “cascade situations” evolve once public information becomes overwhelming and private information is rationally discarded by all following agents in the decision sequence (Graham, 1999). Under the assumptions of a strictly consecutive line of decision making and a separation of public and private information, the concept of informational cascades serves as a reasonable explanatory approach to homogenous forecasts. Following the seminal experiment by Anderson and Holt (1997), numerous experimental studies have been conducted to investigate subjects’ decision making within this framework. It has been shown that despite the overall rational weighing of private and public information, subjects display the tendency to irrationally stick to private information in cascade situations (Weizsäcker, 2010). Accordingly, while cascade situations still occur considerably often, more information is revealed compared to perfect Bayesian updating and decisions become less homogenous than proposed by the theory.

¹⁰ Extensions to the basic concept are suggested in the models by Froot et al. (1992), Trueman (1994), Ehrbeck and Waldmann (1996), Ottaviani and Sorensen (2006a; b; c).

Methodological remarks

While the objective of elaborating behavioral foundations for biased forecasts is worthwhile, the overall trend towards using evidence from experiments only distantly related to the actual topic has also been subject to criticism. Clark and Friesen (2009) pointed out economists' tendency to uncritically use stylized facts derived from experimental studies to interpret all kinds of empirical phenomena. They argue that experimental evidence can only be validly transferred if the respective experimental design closely reproduces the real-world situation that it is supposed to explain. This comprehensible and constructive argument for experimental settings oriented towards external validity is taken up in this book to add to a behavioral understanding of economic forecasts. The authors thereby build on existing experimental settings from the two distinct paradigms presented above and introduce prevalent factors for real-world decision situations in the field of forecasting. Consequently, the studies focus on merging and advancing insights originating from psychological and game theoretical experimental paradigms. By implementing parameters that increase the internal and external validity of the respective experimental paradigms, a compromising "level of realism" regarding the realities of forecasting is sought. This serves to build a bridge between the empirical studies on forecasting accuracy and the related experimental discussions. While a comprehensive and theoretically satisfying application of this ambitious goal is beyond the reach of this dissertation, seven distinct experimental studies are presented. The resulting stylized facts are able to further economic research concerning individual forecasting behavior, the interaction of analysts and obstacles for rational forecasting overall. In particular, this general notion is applied to the two main paradigms used to explain the status quo bias in forecasting.

The first paradigm discussed in this book uses psychological experiments on heuristics and biases to explain characteristics of forecasting time series and analysts' behavior. Relying on psychological experiments in explaining economic contexts is problematic insofar as they regularly fail to incorporate the prevalent conditions in actual markets. Such conditions can be seen in the implementation of monetary incentives and the observation of other market participants, both of which foster learning effects that might lead to a correction of biased choices (List and Millimet, 2008). However, these factors are largely missing in the literature on biases and heuristics regularly applied to forecasting research. In the experimental studies presented in the first part of this book, such factors as monetary incentives, rational strategies, group decisions and the observation of other decision-makers are connected to the

experimental paradigms of the anchoring- and overconfidence bias. This serves at increasing the internal validity of the respective experimental paradigms – from an economic perspective – and thus also the applicability to forecasting.

The second paradigm drawn upon with the goal of increased realism is that of rational herding. From an economic perspective, studies on rational herding have a high internal validity by featuring rational strategies, monetary incentives and ample opportunities of learning to play optimally; nonetheless, the external validity for economic forecasting is arguably low. This results from the strict assumptions of the respective experimental settings, which guarantee a clear comparability with game theoretical benchmarks of rational decision-making, yet hinders the transfer to real-world applications. In the studies presented in the second part of this book, the authors modify economic experiments concerned with rational herding to enable a closer applicability for the explanation of a status quo bias in forecasting. This involves the loosening of several restrictions guiding individual decision-making, particularly endogenizing the timing and order of decisions, opening action spaces, modifying modes of communication or implementing coordination games in a forecasting setting.

Summary of the studies on Anchoring and Overconfidence (Chapter 2 to 5)

The first four studies in this book are concerned with the anchoring- and overconfidence bias, behavioral effects considered as robust human traits by psychological research in the past decades. While both biases are widely applied to economic analyses, consideration using experimental economics' paradigms has only recently started. In the following contributions, doubts concerning the alleged robustness of these biases under economic conditions have been convincingly voiced.¹¹ The present studies connect to these doubts and further apply prevalent conditions in markets to assert the robustness of both anchoring and overconfidence in forecasting.

Chapter two, entitled “*Anchoring: a valid explanation for biased forecasts when rational predictions are easily accessible and well incentivized?*”, presents the first experimental setting investigating anchoring within a forecasting task featuring a rational strategy. The task is implemented in the form of a simple formula based on several known and one random determinant. Subjects are asked to predict future values, maximizing their payoffs by ignoring

¹¹ For anchoring, this argument is made by Maniadis et al. (2014); Fudenberg et al. (2012); Alevy et al. (2011); Bergman et al., (2010); Tufano (2010); Simonson and Drolet (2004). For the case of the overconfidence bias, Clark and Friesen (2009) make a similar point.

the random determinant as its expected value is zero. The traditional anchoring paradigm is built upon as subjects are asked prior to their actual forecast to estimate whether the future value will be higher or lower than the anchor value, in this case defined as the previous round's value. Thus, a rational strategy, strong monetary incentives, feedback to enable learning and finally a realistic chance of making unbiased predictions are integrated, criteria that are not met in most psychological anchoring experiments. The authors further investigate the role of cognitive load by making subjects' tasks more complex and the impact of higher risk exposure. By testing anchoring within a forecasting setting implementing economic experimental paradigms, the authors provide new insight into whether anchoring is a valid explanation for biased forecasts. It is shown that monetary incentives reduce the average anchoring bias to one third of its value in a setting without monetary incentives. In turn, task complexity doubles and increased risk quadruples the anchoring bias. Moreover, a strong reduction of the forecasts' variance results when compared to a non-anchor treatment, which results from the overall biasedness of forecasts. The results are interpreted as a confirmation of empirical studies claiming that forecasts biased towards current values accrues from anchoring. While monetary incentives, obviously prevalent in actual markets, reduce the bias, higher cognitive load and risk increase the subconscious reliance on anchor values and reduce the overall variance of forecasts. The transfer of psychological anchoring research to an experimental economic paradigm thus supports the widespread assumption that biased forecasts results from anchoring.

The third contribution, entitled "*An experimental study on social anchoring*", expands the second chapter of this book by introducing the element of public observation of other subjects' forecasts. Psychological studies regularly use experimenter-given anchors without an inner connection to the respective task. Contradictory, real-world anchors such as current consensus values are constituted by prior publicly observable and aggregated decisions of all market participants, reflecting an aspect known to forecasters. Implementing an endogenous, realistically derived "social" anchor further increases the external validity for anchoring in forecasting. While the general experimental setting as presented in chapter two remains unchanged, anchor values are now transparently derived from the mean forecast of all participants in the previous period. Again introducing monetary incentives and feedback on performance, it is shown that the "social" anchor significantly increases the bias when compared to the neutral anchor of the experimental setting presented in chapter two. Consequently, a relevant additive influence of conformity pressure in the context of anchoring in forecasting is assumed by the authors.

Chapter four, entitled “*Are Groups ‘Less Behavioral’? The Case of Anchoring*”, introduces a further characteristic of actual markets into the investigation concerning the robustness of anchoring effects. Experimental economics has arrived at a consensus that groups are more rational decision makers than individuals in most settings (Kugler et al., 2012). It is further argued that group cooperation is at the core of most relevant economic decision processes, which is interpreted as a partial vindication of the assumptions of rational decision-making in actual markets. However, there is little evidence on the effect of group cooperation with respect to behavioral biases. If groups were able to avoid anchoring, the assumption of its ubiquitous influence in real-world decision making, particularly in the case of forecasting, could not be upheld. Consequently, an experimental study is run, comparing individuals’ and groups’ anchoring bias in the domains of factual knowledge, probability estimations and price estimates, all of which characterize forecasting tasks. This again serves to further add external validity to the classical anchoring studies, enabling a connection to empirical studies on forecasting accuracy. The authors show that groups are only partly able to reduce the anchoring bias. Groups are less biased than individuals in tasks requiring factual knowledge, while no difference is found for price- and probability estimates. As in the previous studies, the dismal influence of anchoring on decisions under economic conditions featuring group decision-making cannot be altogether negated.

The fifth contribution, entitled “*Overconfidence as a social bias: Experimental evidence*” deals with the robustness of the overconfidence bias under economic conditions. Clark and Friesen (2009) find that learning effects and monetary incentives succeed in quickly eliminating the overconfidence bias in individual decision-making. Using a forecasting task as in the first three experiments, Clark and Friesen’s main result is replicated. However, the authors demonstrate that even a very simple observational setting again fosters overconfident self-assessments and eliminates subjects’ underconfidence, which is prevalent in the individualistic setting. Accordingly, by implementing the observation of other participants as a very basic feature of actual markets, it is shown that market conditions can also aggravate biased behavior, in this case overconfidence.

Overall, the first part of this book can be considered as reinforcing the stability of biases in forecasting. Based on psychological experimental paradigms that are largely inconsiderate of economic conditions, it is shown that conditions such as risk, cognitive load and observability lead to increased individual biasedness. In turn, the availability of rational strategies, monetary incentives learning effects and group cooperation only partially reduces biasedness.

Studies on rational herding, information aggregation and coordination (Chapter 6 to 8)

The second part of this book deals with the opposing side of the discussion, i.e. economic experiments with very restrictive experimental assumptions impeding the application to the context of forecasting.

In chapter six, named “*Strategic coordination in forecasting. An experimental study*”, a coordination game is integrated into a forecasting setting. While coordination games have been extensively conducted in experimental economics, no application has shown considerable external validity regarding a forecasting context. The given set-up fills this gap by having subjects perform a forecasting task with incentives for accurate predictions and contradictory incentives for coordination on focal points, which precludes accurate forecasts. Thereby the experimental economic discussions on focal point coordination are connected to an applied setting, thus enabling a transfer to forecasting research. The authors find that, despite strong incentives for correct predictions, prediction accuracy is low and frustration with the complexity of the forecasting task fosters learning to play the coordination equilibrium. This mechanism leads to high levels of coordination, predominantly on current values that constitute the payoff dominant equilibrium. Consequently, the experimental design explores rational coordination and the connected abandoning of forecasting efforts as a necessary condition of analysts’ reputational herding on current values. The loosening of a number of restrictive parameters given in ordinary coordination games effectively enables the transfer to a forecasting setting, providing insights into mechanisms of strategic coordination among analysts.

The seventh contribution, “*A comparison of endogenous and exogenous timing in a social learning experiment*”, furthers the discussions on social learning as the second major explanatory approach for rational herding in forecasting. The majority of studies on rational herding refer to the paradigm of informational cascades, following Anderson and Holt’s (1997) experiment featuring sequential binary decisions. By contrast, this study presents a two-player social learning game that implements endogenous timing and a continuous action space, whereby outcomes are compared to a set-up with an exogenously fixed order of decisions. Waiting costs are implemented, which are positively correlated with private signal’s strength, hence outwaiting the other player is costly. While leading to the loss of a strict internal benchmark of rational decision making, this change enables a more realistic perspective on sequential forecasting, in which analysts can outwait others to gather additional information by observing their decisions. However, waiting to become the second-

mover is costly and endogenous ordering might also become disadvantageous in terms of overall welfare. Additionally, the authors emphasize that continuous action spaces more accurately capture actual circumstances, e.g. in situations in which point estimations are required rather than binary forecasts merely indicating the direction of future developments. Thus, a continuum of potential values can be predicted, replacing the binary decision framework crucial for the occurrence of cascade situations. Overall, it is shown that endogenous timing in decision making fosters informational efficiency, compared to an exogenously fixed order of predictions. However, the waiting costs incurred compensate the welfare gains from the more accurate predictions, which were enabled through the additional information revealed by the predictions' specific timing. Consequently, the findings resulting from a more realistic social learning game in the context of forecasting suggest that information aggregation becomes more effective, leading to more accurate predictions. Yet, at the same time, costs of delay arise, thus keeping overall welfare unchanged after all. This core result is robust against the specific design of waiting costs and accuracy rewards, since subjects succeed in considerably adjusting to the configuration of these key parameters. These findings add further experimental evidence to the discussions on social learning by investigating the welfare effects of loosening restrictive experimental parameters, thus getting closer to real-world applications such as forecasting.

Chapter eight, entitled "*The impact of communication regimes on group rationality: Experimental evidence*", introduces an experiment that builds on the seminal social learning game by Anderson and Holt (1997) and the closely related study on group performance by Fahr and Irlenbusch (2011), thus marking a contrast to the previous chapter. In this benchmark paper, it is shown that group cooperation enhances rationality in cascade situations. While this brings greater external validity to the investigation of rational herding, a relevant influence on group rationality remains unclear: it can be argued that small group research in experimental economics has not featured realistic means of communication. The majority of studies rely on face-to-face communication, regularly yielding superior results.¹² However, the strong assumption of universal personal discussions in decision-making is problematic. In real-world organizations, a continuum of modes of decision-making are applied, from personal to purely virtual, which in turn might lead to very different degrees of effectiveness in conveying and aggregating information (Balliet, 2010). Accordingly, by checking the results of Fahr and Irlenbusch (2011) under conditions of virtual communication,

¹² See Kugler et al. (2012) for an elaboration of this aspect and a comprehensive literature review on group performance in economic experiments.

the authors investigate the fragility of group superiority in rational herding situations. Overall, once decision-making is made using virtual communication, group rationality deteriorates for more complex tasks, while superior group rationality only replicates for cognitively less challenging tasks. Linking the results to forecasting research, the authors suggest that the assumption of groups' superior rationality in situations of social learning is overoptimistic, given that it fails to hold for even slightly impeded conditions of decision-making. Thus, as already discussed in chapter four, it should not be assumed that group cooperation universally corrects individual fallacies.

The second part of this book deals with the loosening of restrictive assumptions in experimental settings applicable to behavioral forecasting research. Overall, the authors argue that applying established experimental settings to conditions closer to the actual realities of forecasting offers the potential to provide greater external validity regarding rational herding in forecasting. Of course, the present studies provide evidence confined to specific fields. Nevertheless, their results point to the merits of building a bridge between empirical results and behavioral studies by striving for the consolidation of different experimental paradigms.

REFERENCES

- Ackert, L.F., Church, B.K., Ely, K., 2008. Biases in Individual Forecasts: Experimental Evidence. *The Journal of Behavioral Finance* 9 (2), 53-61.
- Ager, P., Kappler, M., Osterloh, S., 2009. The accuracy and efficiency of the Consensus Forecasts: A further application and extension of the pooled approach. *International Journal of Forecasting* 25 (1), 167-181.
- Alevy, J.E., Landry, C.E., List, J.A., 2011. Field Experiments on Anchoring of Economic Valuations. University of Alaska Anchorage, Department of Economics, Working Paper No. 2011-02.
- Anderson, L.R. Holt, C.A., 1997. Information Cascades in the Laboratory. *The American Economic Review* 87 (5), 847-862.
- Balliet, D., 2010. Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution*, 54(1), 39-57.
- Banerjee, A.V., 1992. A simple model of herd behavior. *Quarterly Journal of Economics* 107 (3), 797-817.
- Barber, B.M., Odean, T., 2001. Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics* 116 (1), 261-292.
- Batchelor, R., 2007. Bias in macroeconomic forecasts. *International Journal of Forecasting* 23 (2), 189-203.
- Bedke, N., Bizer, K., Spiwoks, M., 2009. Gregarious Analysts - Experimental Evidence for Reputational Herding. *Journal of Money, Investment and Banking* 12, 26-36.
- Berg, N., Gigerenzer, G., 2010. As-If Behavioral Economics: Neoclassical Economics in Disguise?. *History of Economic Ideas* 18 (1), 133-166.
- Bergman, O., Ellingsen, T., Johannesson, M., Svensson, C., 2010. Anchoring and cognitive ability. *Economics Letters* 107 (1), 66-68.
- Biais, B., Hilton, D., Mazurier, K., Pouget, S., 2005. Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market. *The Review of Economic Studies* 72 (2), 287-312.

- Bikhchandani, S., Hishleifer, D., Welch, I., 1992. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy* 100 (5), 992-1026.
- 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 (4), 437-449.
- Bucchianeri, G.W., Minson, J., 2013. A homeowner's dilemma: Anchoring in residential real estate transactions. *Journal of Economic Behavior & Organization* 89 (6), 76-92.
- 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 (2), 369-390.
- 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 (1), 47-76.
- Clark, J., Friesen, L., 2009. Overconfidence in Forecasts of Own Performance: An Experimental Study. *The Economic Journal* 119 (534), 229-251.
- Claussen, C.A., Matsen, E., Røisland, Ø., Torvik, R., 2012. Overconfidence, monetary policy committees and chairman dominance. *Journal of Economic Behavior & Organization* 81 (2), 699-711.
- Clement, M.B., Tse, S.Y., 2005. Financial analyst characteristics and herding behavior in forecasting. *The Journal of Finance* 60 (1), 307-341.
- Cote, J., Sanders, D., 1997. Herding behavior: Explanations and implications. *Behavioral Research in Accounting* 9, 20-45.
- Ehrbeck, T., Waldmann, R., 1996. Why are professional forecasters biased? Agency versus behavioral explanations. *The Quarterly Journal of Economics* 111 (1), 21-40.
- Elliott, G., Timmermann, A., 2008. Economic forecasting. *Journal of Economic Literature* 46 (1), 3-56.
- Fahr, R., Irlenbusch, B., 2011. Who follows the crowd – Groups or individuals?. *Journal of Economic Behavior & Organization* 80 (1), 200-209.
- Fama, E.F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance* 25 (2), 383-417.
- Fama, E.F., 1991. Efficient Capital Markets II. *Journal of Finance* 46 (5), 1575-1617.

- Froot, K.A., Scharfstein, D.S., Stein, J.C., 1992. Herd on the Street: Informational Inefficiencies in a Market with Short-Term Speculation. *The Journal of Finance* 47 (4), 1461-1484.
- Fudenberg, D., Levine, D.K., Maniadis, Z., 2012. On the robustness of anchoring effects in WTP and WTA experiments. *American Economic Journal: Microeconomics* 4 (2), 131-145.
- 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.
- Furnham, A., Boo, H.C., 2011. A literature review of the anchoring effect. *The Journal of Socio-Economics* 40 (1), 35-42.
- Gallo, G.M., Granger, C.W., Jeon, Y., 2002. Copycats and Common Swings: The Impact of the Use of Forecasts in Information Sets. *IMF staff Papers* No 49 (1).
- Graham, J.R., 1999. Herding among Investment Newsletters: Theory and Evidence. *The Journal of Finance* 54 (1) 237-268.
- Grubb, M.D., 2009. Selling to overconfident consumers. *The American Economic Review* 99 (5), 1770-1807.
- Gubaydullina, Z., Hein, O., Spiwoks, M., 2011. The Status Quo Bias of Bond Market Analysts. *Journal of Applied Finance & Banking* 1 (1), 31-51.
- Harvey, N., 2007. Use of heuristics: Insights from forecasting research. *Thinking & Reasoning* 13 (1), 5-24.
- 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.
- Jegadeesh, N., Kim, W., 2009. Do analysts herd? An analysis of recommendations and market reactions. *Review of Financial Studies* 23 (2), 901-937.
- Johnson, J.E.V., Schnytzer, A., Liu, S., 2009. To what extent do investors in a financial market anchor their judgments excessively? Evidence from the Hong Kong horserace betting market. *Journal of Behavioral Decision Making* 22 (4), 410-434.
- Krugman, P., 2009. How Did Economists Get It So Wrong?. *New York Times*, see: http://www.ie.ufrj.br/hpp/intranet/pdfs/krugman_september_6_2009_howdideconomistsdidsowrong.pdf, last accessed: 03/14/14.

- Kugler, T., Kausel, E.E., Kocher, M.G., 2012. Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley Interdisciplinary Reviews: Cognitive Science* 3 (4), 471-482.
- Lawrence, M., O'Connor, M., 2000. Sales forecasting updates: how good are they in practice?. *International Journal of Forecasting* 16 (3), 369-382.
- 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 (1), 1-53.
- Malmendier, U., Tate, G., 2005. CEO Overconfidence and Corporate Investment. *The Journal of Finance* 60 (6), 2661-2700.
- 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.
- 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.
- Menkhoff, L., Schmeling, M., Schmidt, U., 2013. Overconfidence, experience, and professionalism: An experimental study. *Journal of Economic Behavior & Organization* 86, 92-101.
- Moore, D.A., Healy, P.J., 2008. The trouble with overconfidence. *Psychological Review* 115 (2), 502-517.
- Ottaviani, M., Sorensen, P.N., 2006a. Professional advice. *Journal of Economic Theory* 126 (1), 120-142.
- Ottaviani, M., Sorensen, P.N., 2006b. The strategy of professional forecasting. *Journal of Financial Economics* 81 (2), 441-466.
- Ottaviani, M., Sorensen, P.N., 2006c. Reputational cheap talk. *The Rand Journal of Economics* 37 (1), 155-175.
- Plickert, P., 2009. Ökonomen im Erklärungsnotstand. *Frankfurt Allgemeine Zeitung* (Online), see: <http://www.faz.net/aktuell/wirtschaft/konjunktur/keine-prognose-oekonomen-im-erklaerungsnotstand-1791987.html>, last access: 03/12/14.
- Scharfstein, D.S., Stein, J.C., 1990. Herd Behavior and Investment. *The American Economic Review* 80 (3), 465-479.
- Simonson, I., Drolet, A., 2004. Anchoring Effects on Consumers' Willingness-to-Pay and Willingness-to-Accept. *Journal of Consumer Research* 31 (3), 681-90.

- Spiwoкс, M., 2002. Ansätze zur Überprüfung der Hypothese informationseffizienter Kapitalmärkte - Ein Literaturüberblick. Sofia- Studien 02-5.
- Spiwoкс, M., 2004. External triggered herding bei Rentenmarkt-Analysten. *Financial Markets and Portfolio Management* 18 (1), 58-83.
- Spiwoкс, M., Bedke, N., Hein, O., 2008. Forecasting the past: the case of US interest rate forecasts. *Financial Markets and Portfolio Management* 22 (4), 357-379.
- Spiwoкс, M., Bedke, N., Hein, O., 2010. Topically Orientated Trend Adjustment and Autocorrelation of the Residuals - An Empirical Investigation of the Forecasting Behavior of Bond Market Analysts in Germany. *Journal of Money, Investment and Banking* 14, 16-35.
- Spiwoкс, M., Hein, O., 2007. Die Währungs-, Anleihen- und Aktienmarktprognosen des Zentrums für Europäische Wirtschaftsforschung. *AStA Wirtschafts- und Sozialstatistisches Archiv* 1 (1), 43-52.
- Trueman, B., 1994. Analyst forecasts and herding behavior. *Review of Financial Studies* 7 (1), 97-124.
- Tufano, F., 2010. Are 'True' Preferences Revealed in Repeated Markets? An Experimental Demonstration of Context-Dependent Valuations. *Experimental Economics* 13 (1), 1-13.
- Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. *Science* 185 (4157), 1124-1131.
- Weizsäcker, G., 2010. Do we follow others when we should? A simple test of rational expectations. *The American Economic Review* 100 (5), 2340-2360.
- Welch, I., 1992. Sequential sales, learning, and cascades. *The Journal of Finance* 47 (2), 695-732.
- Welch, I., 2000. Herding among security analysts. *Journal of Financial Economics* 58 (3), 369-396.

Chapter II

ANCHORING: A VALID EXPLANATION FOR BIASED FORECASTS WHEN RATIONAL PREDICTIONS ARE EASILY ACCESSIBLE AND WELL INCENTIVIZED?

with Till Proeger and Kilian Bizer

Published:

Cege Discussion Papers, no. 166, Göttingen University, July 2013.

Download: <http://www.uni-goettingen.de/de/60920.html>

ABSTRACT

Behavioral biases in forecasting, particularly the lack of adjustment from current values and the overall clustering of forecasts, are increasingly explained as resulting from the anchoring heuristic. Nonetheless, the classical anchoring experiments presented in support of this interpretation lack external validity for economic domains, particularly monetary incentives, feedback for learning effects and a rational strategy of unbiased predictions. We introduce an experimental design that implements central aspects of forecasting to close the gap between empirical studies on forecasting quality and the laboratory evidence for anchoring effects. Comprising more than 5,000 individual forecasts by 455 participants, our study shows significant anchoring effects. Without monetary incentives, the share of rational predictions drops from 42% to 15% in the anchor treatment. Monetary incentives reduce the average bias to one third of its original value. Additionally, the average anchor bias is doubled when task complexity is increased, and quadrupled when the underlying risk is increased. The variance of forecasts is significantly reduced by the anchor once risk or cognitive load is increased. Subjects with higher cognitive abilities are on average less biased toward the anchor when task complexity is high. The anchoring bias in our repeated game is not influenced by learning effects, although feedback is provided. Our results support the studies that ascribe biased forecasts and their specific variance to anchoring effects.

Keywords: anchoring; cognitive ability; forecasting; heuristics and biases; incentives; laboratory experiment

JEL classification: C90; D03; D80; G17

HIGHLIGHTS

- We find significant anchoring effects in a repeated numerical forecasting task.
- Monetary rewards for accuracy and subjects' cognitive abilities decrease anchoring.
- Scaling up risk and cognitive load to capture actual volatility increases anchoring.
- Anchoring is not reduced by feedback and learning effects.
- Our results suggest biased forecasts with low variance are driven by anchoring.

1. INTRODUCTION

The anchoring heuristic (Tversky and Kahnemann, 1974) is increasingly considered when explaining biased forecasts with examples including financial forecasts (Fujiwara et al., 2013), real estate price forecasts (Northcraft and Neale, 1987; Bucchianeri and Minson, 2013), sports betting (Johnson et al., 2009; McAlvanah and Moul, 2013), earnings forecasts (Cen et al., 2013), macroeconomic forecasts (Nordhaus, 1987; Frankel and Froot, 1987; Bofinger and Schmidt, 2003; Campbell and Sharpe, 2009; Hess and Orbe, 2013) and sales forecasting (Lawrence and O'Connor, 2000). The findings point to two core empirical patterns: an excessive influence of current values and a clustering of forecasts, reflected in a low overall variance. The underlying mechanism is typically described as in Harvey (2007, p. 17), who states that forecasters tend to “*use the last data point in the series as a mental anchor and then adjust away from that anchor to take account of the major feature(s) of the series. However, as adjustment is typically insufficient, their forecasts are biased.*” Given that almost 40 years of psychological studies show the robustness of anchoring (cp. Furnham and Boo, 2011 for a review), it provides a reasonable explanation for biased individual forecasts.¹ There is, however, substantiated criticism concerning the immediate applicability of psychological evidence to explain economic data. On a general level, markets are expected to rule out behavioral biases as individuals gain expertise and face real financial stakes (Levitt and List, 2007; List and Millimet, 2008). Persistent biases subsequently result from specific laboratory conditions and experimenter demand effects, and ultimately hold little relevance outside the lab (Zizzo, 2012; for anchoring, see Chapman and Johnson, 1999). In the specific case of anchoring, this is suggested in the field experiments of Alevy et al. (2010) and Fudenberg et al. (2012), who show only minor anchoring effects on subjects’ willingness-to-pay/-accept. Their results resonate well with Clark and Friesen’s (2009) criticism of economists’ tendency to casually adopt psychological biases as stylized facts without supportive experimental studies that implement economic conditions. In the classic psychological studies cited in support of anchoring in forecasting, subjects take uninformed and non-incentivized guesses (“How many African countries in the UN?”). Thus, anchoring cannot be seen as a deviation from the rational strategy. In contrast, anchoring might actually

¹ Another prominent explanation of systematically biased forecasts points to reputational concerns of forecasters trying to strategically conceal their inability to predict future values. This results in strong incentives for herding behavior among forecasters. For this approach, see e.g. Ottaviani and Sorensen (2006) or Lamont (2002) and the experimental study by Ackert et al. (2008).

increase – if only slightly – the likelihood of a correct guess when subjects lack task specific knowledge and are not provided any information. While the external validity might hold for situations of purely intuitive decision-making, it is insufficient proof for forecasting settings where distinctly non-intuitive decision processes and strong incentives for correct predictions prevail. Taking up the doubts concerning the transferability of anchoring, McAlvanah and Moul (2013) investigate anchoring “in the wild” (ibid. p. 88) for the case of horseracing bookmakers. They find anchoring effects, stating that they are robust for an economic domain as the experienced bookmakers face real financial incentives for unbiased predictions under time pressure.

However, controlled laboratory studies are needed to systematically assess the robustness of anchoring in forecasting settings. This includes timely feedback to enable learning effects, a chance of correct predictions by providing a rational strategy of avoiding the anchor, a non-intuitive high cognitive effort task and finally monetary incentives. Our experimental design implements these factors. We thus close the gap between economic empirical studies on anchoring and the respective psychological lab-based studies in order to improve the external validity of the anchoring heuristic for economical domains. We introduce a simple numerical forecasting task that distinctly facilitates unbiased decisions as the rational strategy. The respective last values of the time series serve as anchors and thus have a dual function: they reveal the previous rounds’ correct value to enable learning effects, as well as provide the external anchor for the current round. In this setting, we investigate the influence of monetary incentives, cognitive abilities, task-specific risk and cognitive load on the extent of the anchoring bias. In contrast to previous forecasting experiments (Leitner and Leopold-Wildburger, 2011 provide a review), a correct prediction is considerably easy to achieve.² Unlike regular anchoring experiments, we facilitate the rational strategy to test for anchoring under conditions that offer an easily accessible strategy of unbiased forecasts. While this evidently contradicts the complexities of actual forecasting, we argue that a test of anchoring

² There are many time series forecasting experiments investigating individual prediction behavior (see Harvey 2007, for a literature review). However, these studies are not designed to capture anchoring itself. While they point to anchoring as a potential explanation of behavior, the designs do not give specific evidence comparable to previous research on anchoring. They are also defined by excessive complexity of the forecasting tasks and varying sources of information. As we are not interested in these aspects, but rather the anchoring effect itself, we refrain from basing our setting on the classic forecasting experiments. For examples of time series forecasting experiments, see e.g. Bolger and Harvey (1993); Lawrence and O’Connor (1995); Becker et al. (2005, 2007, 2009); Leitner und Schmidt (2006); Reimers and Harvey (2011).

in forecasting should implement a low-complexity task. If anchoring occurs when avoiding it is simple and incentivized, we assume that its impact on actual forecasts in a complex environment is even more relevant. In the following, the respective literature is reviewed to deduct our behavioral hypotheses.

Tversky and Kahnemann's (1974) seminal paper presented the 'anchoring-and-adjustment' heuristic, from which numerous studies have evolved that show a pervasive influence of anchoring in decision-making. The aspects tested are diverse and range from factual knowledge (Blankenship et al., 2008; Wegener et al., 2001) to probability calculations (Chapman and Johnson, 1999) to price estimations after monetary reforms (Amado et al., 2007). Task-specific expertise is shown to be irrelevant for the anchoring bias, as in English and Soder (2009), for a juridical context supporting the assumption that forecasting experts are susceptible to anchor heuristics. Overall, the influence of the anchoring heuristic proved to be "*exceptionally robust, pervasive and ubiquitous*" (Furnham and Boo, 2011, p. 41) regarding experimental variations.

The only experimental study of anchoring in a forecasting context was presented by Critcher and Gilovich (2008), who investigated the influence of incidental anchors in real life; for example, by attempting to forecast the capabilities of athletes with high and low shirt numbers. They find that subjects are subconsciously biased by the closest incidental anchor in their environment for their estimations.

Regarding incentives for accurate predictions, Tversky and Kahnemann (1974), Wilson et al. (1996) and Epley and Gilovich (2005) offer prizes as rewards for the most accurate, unbiased estimations but find only minor effects of such an incentive. Chapman and Johnson (2002) summarize these findings, concluding that "*incentives reduce anchoring very little if at all*" (p. 125). Wright and Anderson (1989) find a reduction in the bias using performance-related financial incentives, if subjects are familiar with the tasks. Simmons et al. (2010) show that incentives for accuracy work, once subjects are given certainty about the correct direction of adjustment for their initial predictions. We interpret these contradictory findings as resulting from a varying availability of strategies for solving the given tasks and the information at hand. Once participants are given the realistic chance of issuing more accurate predictions, monetary incentives are able to reduce anchoring effects. This is in line with standard assumptions concerning the introduction of monetary incentives in economic experiments (see e.g. Smith and Walker, 1993), which are expected to induce more rational behavior.

There are two contradictory results concerning the role of cognitive abilities in anchoring. Stanovich and West (2008) do not find a correlation between the susceptibility to anchoring

and higher cognitive abilities, based upon individually stated SAT results. Oechssler et al. (2009) come to the same conclusion using the cognitive reflection test (Frederick, 2005). Using a corporate cognitive ability test, Bergman et al. (2010) find a significant reduction of anchoring in subjects with higher cognitive abilities.

Blankenship et al. (2008) examine the effect of increased cognitive load, i.e. a systematic mental overload of subjects, by establishing time pressure and making the task more complex. They find significant anchoring effects once high cognitive load is established, which supports Wegener et al. (2001) as well as Wegener et al. (2010) in their assumption that at least two anchoring heuristics exist: one that occurs under high cognitive load and another that occurs under low cognitive load. They argue that different levels of cognitive effort induce anchoring, albeit due to different mechanisms. On the one hand, in simple tasks, the anchor is used intuitively as a cue to the correct answer; on the other, the anchor works in the framework of a more complex thinking process by activating anchor-consistent knowledge. Therefore, anchor biases can occur in the context of intuitive decisions and analytically challenging tasks. While the observable result is identical, the cognitive processes that elicit anchoring need to be differentiated in respect of the context investigated (Crusius et al., 2012). Consequently, a valid test of anchoring in forecasting has to implement high-cognitive-effort tasks that more closely resemble the actual cognitive processes of forecasting, in contrast to the classical anchoring studies that mostly induce intuitive responses. Accordingly, the anchoring task has to bring about non-intuitive decisions, yet provide a fairly simple rational strategy of unbiased decisions.

We contribute to the literature reviewed above by presenting new evidence on the influence of incentives for unbiased predictions, cognitive abilities, task complexity, cognitive load and learning effects in the context of anchoring. Despite the deliberately simple rational strategy for unbiased predictions, we find significant anchoring effects. Monetary incentives reduce the average anchoring bias to around one third compared with non-monetary conditions. Increased cognitive load doubles the average anchoring bias, while increased risk quadruples the distortion when compared to monetary conditions. The variance of forecasts is smaller when higher risk or cognitive load is induced. Participants with higher cognitive capabilities are less prone to the influence of external anchors in a setting with higher cognitive load. Despite the feedback in every round, the anchoring bias is not reduced by learning effects. In sum, we show that the core findings regarding biased forecasts – a lack of adjustment from current values and clustering – might very well be attributed to anchoring effects.

The remainder of this paper is organized as follows: in section 2, we describe the experimental design, section 3 introduces our behavioral hypotheses, section 4 presents the results and section 5 concludes.

2. EXPERIMENTAL DESIGN

We implement a forecasting task whereby participants are asked to predict future values using a simple formula comprising several determinants. The formula is known to participants and remains constant throughout the experiment. Subjects have to predict the correct value using this given formula and the determinants that change each round.³ One determinant is a random variable which is uniformly distributed over the interval $[-25,25]$. Its realizations are unknown and change every round, thus we induce an element of risk into the forecasting task. Its expected value is zero. The formula is $x_t = a_t + b_t - c_t + d_t$; x_t being the value participants are asked to predict, a_t, b_t, c_t are the known determinants and d_t is the random variable.

Each of our four experiments comprises two treatments. In the anchor treatments, subjects are shown the realized value of the previous round as an external anchor, and are asked whether the value of the current round will be higher or lower than the anchor value. In this way, the standard paradigm of traditional anchoring (Tversky and Kahnemann, 1974) is implemented. The design basically demands participants to give a directional forecast first, then a point forecast. Subjects in the respective control groups are not shown the realized value of the previous round and accordingly are not asked the higher/lower question.

The rational strategy for payoff maximization is the calculation of the expected value using the formula and determinants. Given that the expected value of the random determinant is zero, it should not affect predictions. Moreover, the external anchor of the previous value does not contain any additional information for the current round. Therefore, any bias toward the anchor value can be qualified as not rational.

In our first experiment (“basic”), we test if anchoring occurs when participants forecast without monetary incentives. Participants were asked to participate in a classroom experiment. Beforehand, every subject receives instructions⁴ along with the formula, as well as ten forms for entering his or her calculation in each round. Instructions are read aloud prior

³ Subjects in the classroom experiment were allowed to use a pocket calculator, whereas in the lab they were able to use the Windows calculator implemented in the z-Tree program.

⁴ The introductions were in German. A translation is provided in Appendix.

to the experiment. Before starting their calculations, subjects are asked to do the cognitive reflection test (Frederick, 2005) in a maximum of six minutes, two minutes for each question. Subsequently, the calculations begin. Note that the calculations are intentionally fairly easy to solve. For instance, the calculation in the first round is $100 + 40 - 50 = 90$; a task that every participant should be able to complete. Each round lasts one minute, during which the determinants and the last round's realized value (in anchor treatment only) are displayed on a PowerPoint sheet and read aloud. Participants are asked to write down their estimations on their forms. In the anchor treatment, they are additionally asked to estimate whether the current value is higher or lower than the previous value. Each treatment has ten rounds.

The second experiment ("monetary") introduces a monetary incentive for accurate predictions. The experiments 2-4 are conducted using the software 'z-tree' (Fischbacher, 2007) and carried out in an experimental lab.⁵ The formula and determinants remain identical, likewise the cognitive reflection test before the actual experiment. The time for calculating the current value remains one minute per round, with fifteen rounds played in the second experiment. The payoff in each round is fifty cents minus the absolute difference between the respective forecast and the correct value in cents. Payoffs cannot become negative. Subjects are given an additional Euro for correctly answering all three CRT questions at the beginning. The third experiment ("risk") increases the underlying risk by tripling the range of the random determinant's interval. Accordingly, the (d_t) 's are realizations of a random variable uniformly distributed over the interval $[-75,75]$. The expected value remains at zero. In order to account for the higher variance of d_t , the payoff in each round is eighty cents minus the absolute difference between the respective forecast and the correct value in cents.

The fourth experiment ("cognitive load") reduces the time that subjects have to make predictions to 30 seconds and introduces a more complex formula. The formula can now be written as $x_t = a_t + b_t - 0.5c_t + d_t^2 + e_t$; e_t being the random variable, again uniformly distributed over the interval $[-25,25]$. x_t is the value participants are asked to predict in each round, a_t , b_t , c_t , d_t are the known determinants in round t .

Given the realizations for all determinants, following the rational strategy of predicting the expected values of x_t yields on average 0.38€ (=50-12.1) per prediction in the *monetary* experiment (0.45€ in *risk* and 0.38€ in *cognitive load*). A naïve strategy of predicting the previous round's values, i.e. anchoring in the most extreme way, would yield on average

⁵ Since we run a new control group in each experiment, transferring the experiment to the lab should not lead to a misinterpretation of the results. This would only be true if the control and anchor groups were affected differently by the conditions in the lab.

0.20€ per prediction in *monetary* (0.33€ in *risk* and 0.22€ in *cognitive load*). Bearing in mind that subjects make 15 forecasts in total, there is obviously a strong monetary incentive for unbiased predictions. However, relying on the anchor values generates some payoff due to the weak autocorrelation of values to be predicted. We thus capture a key feature of real time series data: although no additional information can be obtained by observing the previous round's values, the naïve forecast yields some success.

Experiment 1 was conducted at the University of Göttingen in May 2012. Participants were undergraduate students in multiple tutorials of an introductory course in economics. Control and treatment groups were conducted in different tutorials. The experiment took on average eighteen minutes.

The lab-based experiments took place in twenty six sessions from May to July 2012 and were conducted in the Laboratory for Behavioral Economics at the University of Göttingen. Participants were recruited using the online recruiting system ORSEE (Greiner, 2004) and were only allowed to participate in one session, which lasted around thirty minutes. On average, each participant earned €6,86. Overall, participants were on average 23.3 years old, 54% were female. Table 1 provides an overview of the different experiments and the numbers of participants.⁶

Experiment	Variation			Number of participants		
	monetary	risk	cognitive load	control	anchor	total
<i>basic</i>	no	low	low	58	115	173
<i>monetary</i>	yes	low	low	44	53	97
<i>risk</i>	yes	high	low	39	53	92
<i>cognitive load</i>	yes	low	high	35	58	93
total				176	279	455

Table 1: Summary of experiments and participants.

⁶ Note that in *basic*, the treatment-specific difference in number of participants is due to the number of participants in the respective tutorials; in the laboratory experiments, differences occur because anchor treatments were conducted earlier on and yielded more attendees, while control treatments were conducted after the anchor treatments where attendance was weaker. However, our analysis of treatment comparison is not influenced or biased by these differences in any way.

3. HYPOTHESES

Given that anchoring has been shown to be “*extremely robust*” (Furnham and Boo, 2011, p. 41) in various settings, we expect a significant bias towards the external anchor values within our forecasting design.

Following Wright and Anderson (1989) and Simmons et al. (2010) and thus discarding Epley and Gilovich (2005), Wilson et al. (1996) and Tversky and Kahnemann (1974), monetary incentives can be expected to reduce anchoring, since a rational strategy is available. Increased cognitive load and risk exposure should further increase anchoring as subjects might act more intuitively (Blankenship et al., 2008). However, the existence of a simple rational strategy along with monetary incentives can be expected to induce more rational behavior on average (Rydal and Ortmann, 2004); also, time pressure might lead to better decisions as in Kocher and Sutter (2006). The two opposing tendencies of rational strategy versus anchoring bias are addressed in Hypothesis 1:

Hypothesis 1 (“*Rationality and anchoring bias*”). Subjects’ forecasts are biased towards the external anchor.

Based H1, we hypothesize that a systematic bias towards the anchor value can lead to a smaller variance of the forecasts in the treatment group. Therefore, the anchor heuristic would help to explain the empirical result of clustered forecasts. In order to test this assumption, we formulate Hypothesis 2:

Hypothesis 2 (“*Differences in variance*”). The external anchor reduces the variance in forecasts.

Furthermore, we examine the influence of subjects’ cognitive abilities on the extent of the anchoring bias. Therefore, we aim at furthering the ongoing discussion concerning the susceptibility to anchoring depending on cognitive abilities (see Bergman et al., 2010). Consequently, we formulate Hypothesis 3:

Hypothesis 3 (“*Cognitive abilities and anchoring bias*”). Higher cognitive abilities reduce the anchoring bias.

Finally, we are interested in the relevance of learning effects. As the task is repeated and feedback is given in the treatment groups, learning effects are fostered. However, studies on experts in a judicial context (Englich et al., 2005; Englich and Soder, 2009) and in time series forecasting (Harvey et al., 1994; Harvey and Fisher, 2005) suggest that anchoring is

independent of participants' prior knowledge or learning effects. Accordingly, we formulate Hypothesis 4:

Hypothesis 4 (“*Learning effects*”). The anchoring bias is not reduced by learning effects.

4. RESULTS

We structure the following results according to our Hypotheses. First, we investigate prediction accuracy for each experiment, in order to check whether subjects are prone to the anchoring bias. Furthermore, we compare treatment effects between experiments to identify the driving forces of the anchoring bias. Second, we look for differences in the variance of predictions between the treatments. Third, results are evaluated regarding influences of cognitive abilities. Fourth, we comment on learning effects in our experiment.

4.1 RATIONALITY AND ANCHORING BIAS

Recall that showing the correct value of the previous round in the treatment group does not change the profit-maximizing strategy of forecasting the expected value. Additionally, subjects in the control group do not answer the higher/lower-question.⁷ If forecasts in the anchor treatments are biased toward the values of previous rounds, we interpret this as evidence in support of Hypothesis 1.

Table 2 summarizes the main data for treatment comparison, indicating the mean absolute deviation of predictions from the expected values, the fraction of optimal forecasts and the share of subjects acting rationally by treatments for all experiments.⁸ Forecasts equal to the

⁷ In experiment 1, 77% of the higher/lower-questions were answered correctly (87% in experiment 2, 77% in experiment 3 and 68% in experiment 4).

⁸ Our dataset contains 253 missing values (predictions) because subjects did not enter a value in the respective round. Additionally, the dataset is corrected for subjects' forecasts if the task was obviously misinterpreted. We assume this to be true if the forecast of subject i in period t (y_{it}) is smaller than 25 or negative ($y_{it} < 25$), i.e. subjects tried to forecast the random determinant and not the realized value. Thus, 265 observations were deleted. Furthermore, we remove outliers, i.e. forecasts deviating by more than three times the maximum realization of the random determinant from the expected value. Accordingly, for experiments 1, 2 and 4, observations are defined as outliers and dropped if $y_{it} < [E(x_t) - 3 * 25]$ or $y_{it} > [E(x_t) + 3 * 25]$. For experiment 3, we chose a smaller multiplier for the interval due to the greater range of the random determinant. In this case, we drop forecasts if $y_{it} < [E(x_t) - 2 * 75]$ or $y_{it} > [E(x_t) + 2 * 75]$. In total, we removed 100 observations defined as outliers by the criterion described, which leaves us with a total of 5,342 forecasts.

expected value are characterized as optimal. A subject is defined as rational if not more than one forecast deviates from the expected value. Given that the previous round's values are by design first shown in the second period in the treatment group, we exclude values for the very first period. However, the results also hold when including the first period.

Experiment	<i>basic</i>			<i>monetary</i>		
	Control	Anchor	test statistic	Control	Anchor	test statistic
mean absolute deviation	10.2 (13.66)	13.1 (12.32)	-6.490*** (0.0000)	5.8 (10.18)	6.9 (8.77)	-4.394*** (0.0000)
share optimal forecasts	42.8	15.1	(0.0000)	57.4	43.5	(0.0000)
share rational subjects	27.9	6.1	(0.001)	50.0	23.5	(0.013)

Experiment	<i>risk</i>			<i>cognitive load</i>		
	Control	Anchor	test statistic	Control	Anchor	test statistic
mean absolute deviation	20.0 (24.74)	17.2 (17.4)	-0.56 (0.5752)	12.7 (17.21)	13.8 (16.53)	-2.107** (0.0351)
share optimal forecasts	41.0	29.7	(0.0000)	40.2	32.0	(0.0082)
share rational subjects	20.0	10.0	(0.224)	13.0	2.27	(0.113)

Table 2: Treatment comparison

Note: Test-statistics for the average absolute deviation derived by a two-sided Wilcoxon rank-sum test; p-values in parentheses. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$). For the share of optimal forecasts and the share of rational subjects a two-sided Fisher's exact is applied and the respective p-values are shown.

The fraction of optimal forecasts is higher for every experiment in the control group. For example, in *basic*, the average absolute deviation is increased by around 28% when the anchor value is shown (19% in *monetary*, 9% in *cognitive load*). In *risk*, there is no significant difference of the average absolute deviation between treatments, although there is a higher fraction of optimal forecasts. This can be explained by the anchor value's tendency to reduce the variance of deviations. There are more optimal decisions in the control groups, but the non-optimal ones deviate more from the expected value. These results will be discussed in more detail in the context of comparing the variance of forecasts over treatments (subsection 4.2).

However, one might interpret differences across treatments as accruing from the representativeness bias (Kahnemann and Tversky, 1973). The distribution of forecasts in the treatment groups might reflect the distribution of the value to be forecasted.⁹ This is due to the tendency of forecasters to replicate the distribution of a time series' noise, thus incorporating the uncertainty rather than ignoring it for an optimal prediction. (Harvey, 1995; Harvey et al., 1997; Harvey, 2007). We therefore have to show that deviations from the expected value are systematically related to the anchor values and do not stem from non-optimal behavior evoked by the representativeness bias. We test for a specific anchoring pattern in the forecasts of the treatment groups by running a regression.

Equation (1) presents the model to adequately explain the subjects' forecasts. Let y_{it} denote the forecast of subject i at time t , and x_t the realized value at time t , whereby $E(x_t)$ gives its expected value. A_i is a dummy, which is 1 for subjects in the treatment group.

$$y_{it} = \gamma_1 E(x_t) + \theta_1 [A_i(E(x_t) - x_{t-1})] + \theta_2 [E(x_t) - y_{it-1}] + u_{it} \quad (1)$$

In the given context, an optimal forecast of x_t can be explained by the expected value (*expected_value*) $E(x_t)$ only, i.e. ($\gamma_1=1$). However, we are interested in a potential bias caused by the anchor value, which is the realized value of the previous round. We include the term $\theta_1 [A_i(E(x_t) - x_{t-1})]$ (*anchor_deviation*) to control for an anchoring bias. It measures the deviation of the realized value of the previous round x_{t-1} and the expected value in the current round $E(x_t)$ for subjects in the treatment group ($A_i=1$). An unbiased forecast is given if $\theta_1=0$, whereas a forecast biased toward the anchor value is given if $\theta_1 < 0$. Additionally, we control for the influence of the deviation of the previous round's forecast y_{it-1} from the expected value of the current round $E(x_t)$ (*forecast_deviation*). Again, $\theta_2 < 0$ indicates a bias toward the forecast of the previous round, whereas in the absence of this bias θ_2 is equal to zero.¹⁰

In sum, information is used efficiently if a regression of (1) results in an estimation of γ_1 , which is not significantly different from 1. At the same time, all other variables should show an insignificant effect on the values forecasted ($\theta_1 = \theta_2 = 0$). In such a case, there would be no

⁹ The distribution of the values to be forecasted is common knowledge in both treatments. Nevertheless, the representativeness bias might be more relevant in the treatment groups because the noise in the realizations is far more obvious when feedback is given.

¹⁰ This control variable is required due to the possible correlation of forecasts made in consecutive rounds. Since the forecasts and realized values of previous rounds are definitely correlated, this would lead to an omitted variable bias and the inconsistent estimation of all other coefficients.

evidence for H1, indicating that on average and *ceteris paribus* forecasts are made optimally and are unbiased.

Table 3 provides the results of a fixed-effects regression on our unbalanced panel dataset of Eq. (1), applying robust Driscoll and Kraay standard errors. Hence, we control for unobservable heterogeneity, heteroskedasticity, serial correlation in the idiosyncratic errors and cross-sectional dependence.

Experiment	<i>basic</i>	<i>monetary</i>	<i>risk</i>	<i>cognitive load</i>
expected_value	0.853*** (0.030)	0.986*** (0.010)	0.841*** (0.025)	0.766*** (0.045)
anchor_deviation	-0.100*** (0.022)	-0.045** (0.019)	-0.130*** (0.029)	-0.101* (0.048)
forecast_deviation	0.012 (0.031)	-0.020 (0.017)	0.066 (0.041)	0.024 (0.016)
constant	15.44*** (2.92)	2.904*** (0.91)	21.52*** (2.121)	23.00*** (4.623)
F-Statistic ($\gamma_1=1$)	23.24***	1.85	40.27***	26.89***
Prob. > F	(0.001)	(0.197)	(0.000)	(0.000)
Observations	1344	1314	1245	961
No. of Groups	159	96	92	92

Table 3: Fixed-effects regression of Eq. (1) with forecast (y_{it}) as dependent variable.

Note: Robust Standard Errors in parentheses; for F-Statistics p-value in parentheses. (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

For all experiments, we find a significant effect of the deviation in the anchor value.¹¹

Notwithstanding, there are differences between the experiments in terms of the average quality of the forecast. A smaller marginal effect of a change in the expected value, i.e. a smaller γ_1 for $\gamma_1 < 1$, has to be associated with a lower average quality of the forecasts and less rational behavior. In *monetary*, the subjects adjust best compared to the other experiments and optimal on average according to a change in the expected value. The forecasting quality drops

¹¹ We checked the robustness of our results by only considering the first ten rounds played. This check was due to the temporal restriction in the classroom experiment, in which we were only able to play ten rounds. However, estimating Eq. (1) by the same procedure as in Table 3 with only the first ten rounds does not relevantly alter our results. Moreover, we estimated Eq. (1) while controlling for a treatment-specific influence of the *expected_value*. Again, we find a significant influence of the deviation in the anchor value for all experiments, whereby the magnitude of the coefficients changes only slightly.

if there are no monetary incentives (*basic*) or the underlying risk is increased (*risk*). The lowest quality on average is realized if the cognitive load is increased by a more complex definition of the task (*cognitive load*).

For all experiments, we find a negative and significant effect of the deviation in the anchor value ($\theta_1 < 0$), which has to be interpreted as an on average bias towards the realized value of the previous period in forecasts by the treatment group, as compared to the control group. For a decreasing (increasing) value in t compared to $t-1$, subjects in the treatment group give significantly higher (lower) forecasts. This fact has to be considered as a systematic inability to ignore the realized value of the previous round.

Besides the significance of the bias towards the anchor value, its relevance needs to be addressed. Based on the average absolute difference of the anchor values and the expected values of 24.6 points in *basic* (20.4 in *monetary*, 32.9 in *risk*, 20.4 in *cognitive load*), the estimated marginal effect of -0.1 (-0.045, -0.13 and -0.101) amounts to a ceteris paribus bias of 2.46 (0.92, 4.28 and 2.06) points on average. This corresponds to 2.53% (0.94%, 4.3% and 2.11%) of the average values to be forecasted.¹²

Obviously, implementing monetary incentives diminishes the influence of the anchoring bias. In *monetary*, the average bias in the treatment group is around one third of the bias in *basic*. In comparison to *monetary*, higher underlying risk more than quadruples the extent of the bias. Establishing a higher cognitive load through a more complex definition of the task at hand more than doubles the extent of the bias compared to *monetary*.

We conclude that the anchoring bias has a significant and relevant impact on subjects' forecasts. The information given is not used optimally. On average, subjects are unable to ignore the values of the previous rounds, as the rational strategy would suggest. The empirical finding of forecasts being frequently biased towards the respective current values can be motivated by the anchoring bias. Therefore, we interpret our results as presenting strong evidence in favor of H1.

¹² The differences in the average deviation of the anchor value and realized values in experiments 2, 3 and 4 accrue from the lower number of rounds being played in experiment 1, together with small adjustments as part of the formula modification in experiment 4 and changed realized values for the unknown determinant in experiment 3 due to the greater range of the interval of the random variable. The changes in experiment 4 became necessary to avoid subjects' calculations of the expected values becoming too complicated.

4.2 VARIANCE OF FORECASTS

In order to test for differences in the variance of forecasts (H2), we present the standard deviation over experiments and treatments, as well as the Brown and Forsythe statistic resulting from the procedure to test for equality in group variances in Table 4.

Experiment	Standard Deviation		Tests (H ₀ : equality)	
	control	anchor	B/F-statistic (W50)	B/F-statistic (W0)
<i>basic</i>	24.41	22.22	0.3 (0.58)	0.65 (0.419)
<i>monetary</i>	23.6	22.32	1.61 (0.205)	1.61 (0.204)
<i>risk</i>	37.42	29.51	18.96*** (0.000)	22.52*** (0.000)
<i>cognitive load</i>	28.55	24.67	9.88*** (0.002)	9.84*** (0.002)

Table 4: Summary of standard deviations and Brown/Forsythe statistics

Note: Asterisks representing p-values (in parentheses) of the B/F-statistic testing the null of equal variances. (***) p<0.01, ** p<0.05, * p<0.1) W50 denotes the results from the test procedure using the median; W0 when using the mean.

We find a smaller standard deviation in the anchor treatments for all experiments. The difference in *basic* and *monetary* turns out not to be significant at the conventional levels of significance. For *risk* and *cognitive load*, both test procedures point to a significantly lower variance in the anchor treatments. In the presence of monetary incentives and high underlying risk or high cognitive load, we conclude that the anchoring bias causes a smaller variance in the forecasts. Consequently, we find mixed results regarding H2. The uniformity or low variance of forecasts revealed by empirical studies might be explained to some extent by a systematic anchoring bias. The anchor value causes a higher frequency of deviations from rational forecasts, which in turn tend to be smaller compared to the control group.

4.3 COGNITIVE ABILITIES

To test for the influence of cognitive abilities on the anchoring bias, we group subjects using the rule proposed by Oechssler et al. (2009), according to which subjects correctly answering two or more questions of the CR-Test are classified as having “high cognitive abilities”

(HCA), and otherwise as having “low cognitive abilities” (LCA). In total, 29% of the subjects answered none of the questions correctly, 24% got one question right, 23% two questions and 23% all three questions. Accordingly, 53% of the subjects were grouped as having LCA, and 47% as having HCA. We expect LCA subjects to be more prone to the anchoring bias, owing to their tendency to answer intuitively (H3).

We find HCA subjects to predict more accurately and act rationally more often. For *basic (monetary/risk/cognitive load)*, the average absolute prediction error for HCA pooled over treatments is 10.5 points (4.6/14.2/10.5), while for LCA it is 13.3 points (8.7/25.1/17.0). The difference between LCA and HCA subjects in the control group amounts to 0.1 points (2.0/14.9/10.5). For the treatment group, the difference is given by 3.2 points (6.2/9.9/4.5).¹³ However, we are interested in the specific effect on the anchoring bias of having higher cognitive abilities.

Therefore, we modify Eq. (1) such that it allows the identification of a potential influence of a subject’s cognitive abilities on the anchoring bias. HCA_i denotes a dummy for subjects classified as having high cognitive ability.

$$y_{it} = \gamma_1 E(x_t) + \theta_1 [A_i(E(x_t) - x_{t-1})] + \vartheta_1 [A_i HCA_i (E(x_t) - x_{t-1})] + \theta_2 [E(x_t) - y_{it-1}] + u_{it} \quad (2)$$

The impact of the deviation in the anchor values is now to be interpreted according to the subjects’ cognitive abilities. θ_1 gives the marginal effect of a change in the deviation in the anchor values for subjects in the anchor treatment and the LCA group; $(\theta_1 + \vartheta_1)$ gives the marginal effect for the HCA group. The extent of the bias towards the anchor in the LCA group ($\theta_1 < 0$) is smaller for the HCA group if $\vartheta_1 > 0$. Table 5 illustrates the regression results of Eq. (2) using the analogue estimation routine as for Eq.(1).

Again, we find a significant effect on the forecasts of the deviation in the anchor value from the expected value for subjects in the treatment groups. The marginal effect or bias tends to be smaller for subjects in the HCA group; individually though, HCA is only significant in *cognitive load*. The average marginal effect on forecasts in *cognitive load* of a one-unit change in the deviation in the anchor values from the expected value is estimated to be -0.179 in the LCA group and only -0.0395 in the HCA group.

¹³ The control group in *basic* shows an average absolute prediction error of 10.1 points (4.9/15.7/7.4) for HCA subjects and 10.2 points (6.9/30.6/17.9) for LCA; the treatment group in *basic* shows an average absolute prediction error of 10.8 points (4.4/12.6/11.9) for HCA subjects and 14.0 points (10.6/22.5/16.4) for LCA.

Experiment	<i>basic</i>	<i>monetary</i>	<i>risk</i>	<i>cognitive load</i>
expected_value	0.853*** (0.031)	0.986*** (0.010)	0.841*** (0.025)	0.766*** (0.045)
anchor_deviation	-0.106*** (0.026)	-0.073*** (0.01)	-0.144*** (0.028)	-0.179*** (0.079)
anchor_deviation_HCA	0.019 (0.042)	0.0459 (0.033)	0.0252 (0.024)	0.140** (0.060)
forecast_deviation	0.012 (0.031)	-0.021 (0.017)	0.0664 (0.041)	0.024 (0.016)
constant	15.44*** (2.924)	2.888*** (0.901)	21.56*** (2.106)	22.98*** (4.592)
F-Statistic ($\theta_1 = \vartheta_1 = 0$)	10.71*** (0.006)	29.81*** (0.000)	12.83*** (0.001)	2.71 (0.104)
Prob. > F				
Observations	1344	1314	1245	961
No. of groups	158	96	92	92

Table 5: Fixed-effects regression of Eq. (2) with forecast (y_{it}) as dependent variable

Note: Robust Standard Errors in parentheses; for F-Statistics p-value in parentheses (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Overall, we find a non-optimal, biased behavior, even in the HCA group on average. Nevertheless, the extent of the bias tends to be lower in the HCA group, at least in the presence of a more complex definition of the task. Therefore, we also find evidence in support of H3 and conclude that cognitive abilities have an influence on the susceptibility to the anchoring heuristic.

4.4 LEARNING EFFECTS

We hypothesized that learning effects should be absent if anchoring subconsciously influences subjects as a behavioral bias. We find evidence in support of H4, which can best be seen when again considering the share of optimal forecasts or the average absolute deviation only for the last 5 rounds. For *basic* (*monetary/risk/cognitive load*), we find 43% (58.2/45.2/40.6) of forecasts to be rational in the control group and only 15% (45.7/32.2/32.7) in the treatment group. The average absolute deviation amounts to 9.7 (5.4/17.2/13.4) in the control group and 12.9 (6.6/17.5/13.0) in the treatment group. These numbers essentially resemble the results shown in Table 1 when considering all periods.

5. CONCLUSION

The present article extends the applied empirical studies on anchoring in various fields of forecasting with a counterpart laboratory study. Therefore, we implement economic conditions in an anchoring experiment to provide external validity for an economic domain. In contrast to classic anchoring experiments, our study introduces a rational strategy and further captures central features of forecasts, specifically feedback and learning effects, time pressure, a high cognitive effort task and strong monetary incentives for avoiding the anchoring bias.

We find a strong anchoring bias despite the existence of a rational strategy with monetary incentives, feedback and repeated decisions. On average, higher risk and cognitive load increase anchoring, which supports our notion that anchoring is bound to increase for actual forecasting where the cognitive load and complexity is higher. We advance the discussion on incentives for accuracy and show that monetary incentives reduce anchoring if a simple strategy for avoiding anchoring is available. We show a relevant reduction in the average orientation towards the external anchor among individuals performing well on the cognitive reflection test, at least if the task complexity is high. Finally, anchoring tends to reduce the variance of predictions.

Our results support the empirical studies that emphasize anchoring effects in forecasting. We find both a robust influence of the respective last correct value and clustered forecasts despite an accessible and incentivized strategy of avoiding it. It may be assumed that forecasters are generally exposed to significant levels of risks and uncertainty as well as high cognitive load in a complex and dynamic forecasting environment. Even if all relevant information were available to forecasters, as in our experiment, anchoring would prevent an optimal interpretation of data. Consequently, we assume that the effect of anchoring in forecasting demonstrated in our study is bound to increase for real-world predictions and can thus serve as a valid explanation for forecasters' lack of adjustment from current values.

ACKNOWLEDGEMENTS

The authors would like to thank Ingo Geishecker, Claudia Keser and Markus Spiwoks for their helpful comments on earlier drafts of this paper. We are also grateful to participants of the ESA World Meeting in Zurich of 2013, the *cege* and *GLOBE* doctoral research colloquium at the University of Goettingen. Joana Voigt and Tonia Sauerwald provided excellent student assistance.

REFERENCES

- Ackert, L.F., Church, B.K., Ely, K., 2008. Biases in Individual Forecasts: Experimental Evidence. *The Journal of Behavioral Finance* 9 (2), 53-61.
- Alevy, J.E., Craig Landry, C.E., List, J., 2011. Field Experiments on Anchoring of Economic Valuations. University of Alaska Anchorage, Department of Economics, Working Paper No. 2011-02.
- Amado, S., Teköznel, M., Topsever, Y., Ranyard, R., Del Missier, F., Bonini, N., 2007. Does “000,000” matter? Psychological effects of Turkish monetary reform. *Journal of Economic Psychology* 28 (2), 154-169.
- Becker, O., Leitner, J., Leopold-Wildburger, U., 2005. Modelling Judgmental Forecasts under Tabular and Graphical Data Presentation Formats, in: Schmidt, U., Traub, S. (Eds.), *Advances in Public Economics: Utility, Choice and Welfare*. Berlin: Springer, 255-266.
- Becker, O., Leitner, J., Leopold-Wildburger, U., 2007. Heuristic modeling of expectation formation in a complex experimental information environment. *European Journal of Operational Research* 176 (2), 975-985.
- Becker, O., Leitner, J., Leopold-Wildburger, U., 2009. Expectation formation and regime switches. *Experimental Economics* 12 (3), 350-364.
- Bergman, O., Ellingsen, T., Johannesson, M., Svensson, C., 2010. Anchoring and cognitive ability. *Economics Letters* 107 (1), 66-68.
- 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 (6), 1465-1476.

- 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 (4), 437-449.
- Bolger, F., Harvey, N., 1993. Context-sensitive heuristics in statistical reasoning. *Quarterly Journal of Experimental Psychology* 46 (4), 779-811.
- Bucchianeri, G.W., Minson, J., 2013. A homeowner's dilemma: Anchoring in residential real estate transactions. *Journal of Economic Behavior & Organization* 89 (6), 76-92.
- 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 (2), 369-390.
- 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 (1), 47-76.
- Chapman, G.B., Johnson, E.J., 1999. Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes* 79 (2), 115-153.
- Chapman, G.B., Johnson, E.J., 2002. Incorporating the irrelevant: Anchors in judgments of belief and value, in: Gilovich, T., Griffin, D., Kahneman, D. (Eds.), *The Psychology of intuitive Judgment: Heuristics and Biases*. New York: Cambridge University Press, 120-138.
- Clark, J., Friesen, L., 2009. Overconfidence in Forecasts of Own Performance: An Experimental Study. *The Economic Journal* 119 (534), 229-251.
- Critcher, C.R., Gilovich, T., 2008. Incidental environmental anchors. *Journal of Behavioral Decision Making* 21 (3), 241-251.
- Crusius, J., van Horen, F., Mussweiler, T., 2012. Why process matters: a social cognition perspective on economic behavior. *Journal of Economic Psychology* 33 (3), 677-685.
- Englich, B., Mussweiler, T., Strack, F., 2005. The last word in court: a hidden disadvantage for the defense. *Law and Human Behavior* 29 (6), 705-722.
- Englich, B., Soder, K., 2009. Moody experts: how mood and expertise influence judgmental anchoring. *Judgment and Decision Making* 4 (1), 41-50.
- 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 (3), 199-212.

- Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2), 171-178.
- Frankel J., Froot, K., 1987. Using Survey Data to Test Standard Propositions Regarding Exchange Rate Expectations. *The American Economic Review* 77 (1), 133-153.
- Frederick, S., 2005. Cognitive reflection and decision making. *The Journal of Economic Perspectives* 19 (4), 25-42.
- Fudenberg, D., Levine, D.K., Maniadis, Z., 2012. On the robustness of anchoring effects in WTP and WTA experiments. *American Economic Journal: Microeconomics* 4 (2), 131-145.
- 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.
- Furnham, A., Boo, H.C., 2011. A literature review of the anchoring effect. *The Journal of Socio-Economics* 40 (1), 35-42.
- Greiner, B., 2004. An online recruitment system for economic experiments. *GWDG Berichte* 63, 79-93.
- Harvey, N., 1995. Why are judgements less consistent in less predictable task situations? *Organizational Behavior and Human Decision Processes* 63 (3), 247-263.
- Harvey, N., 2007. Use of heuristics: Insights from forecasting research. *Thinking & Reasoning* 13 (1), 5-24.
- Harvey, N., Bolger, F., McClelland, A.G.R., 1994. On the nature of expectations. *British Journal of Psychology* 85 (2), 203-229.
- Harvey, N., Ewart, T., West, R., 1997. Effects of data noise on statistical judgement. *Thinking & Reasoning* 3 (2), 111-132.
- Harvey, N., Fischer, I., 2005. Development of experience-based judgment and decision making: The role of outcome feedback, in: Betsch, T., Haberstroh, S. (Eds.), *The routines of decision making*. Mahwah NJ: Lawrence Erlbaum Associates Inc., 119-137.
- 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.

- 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 (4), 410-434.
- Kahneman, D., Tversky, A., 1973. On the psychology of prediction. *Psychological Review* 80 (4), 237-251.
- Kocher, M.G., Sutter, M., 2006. Time is money - Time pressure, incentives, and the quality of decision-making. *Journal of Economic Behavior & Organization* 61 (3), 375-392.
- Lamont, O.A., 2002. Macroeconomic forecasts and microeconomic forecasters. *Journal of Economic Behavior & Organization* 48 (3), 265-280.
- Lawrence, M., O'Connor, M., 1995. The anchoring and adjustment heuristic in time series forecasting. *Journal of Forecasting* 14 (5), 443-451.
- Lawrence, M., O'Connor, M., 2000. Sales forecasting updates: how good are they in practice?. *International Journal of Forecasting* 16 (3), 369-382.
- Leitner, J., Leopold-Wildburger, U., 2011. Experiments on forecasting behavior with several sources of information - A review of the literature. *European Journal of Operational Research* 213 (3), 459-469.
- Leitner, J., Schmidt, R., 2006. A systematic comparison of professional exchange rate forecasts with the judgmental forecasts of novices. *Central European Journal of Operations Research* 14 (1), 87-102.
- Levitt, S.D., List, J.A., 2007. What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?. *Journal of Economic Perspectives* 21 (2), 153-174.
- 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 (1), 1-53.
- 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.
- Nordhaus, W.D., 1987. Forecasting efficiency: concepts and applications. *The Review of Economics and Statistics* 69 (4), 667-674.
- Northcraft, G.B., Neale, M.A., 1987. Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes* 39 (1), 84-97.

- Oechssler, J., Roider, A., Schmitz, P.W., 2009. Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization* 72 (1), 147-152.
- Ottaviani, M., Sorensen, P.N., 2006. The strategy of professional forecasting. *Journal of Financial Economics* 81 (2), 441-466.
- Reimers, S., Harvey N., 2011. Sensitivity to autocorrelation in judgmental time series forecasting. *International Journal of Forecasting* 27 (4), 1196-1214.
- Rydval, O., Ortmann, A., 2004. How financial incentives and cognitive abilities affect task performance in laboratory settings: an illustration. *Economics Letters* 85 (3), 315-320.
- 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 (6), 917-932.
- Smith, V.L., Walker, J., 1993. Monetary rewards and decision cost in experimental economics. *Economic Inquiry* 31 (2), 245-261.
- Stanovich, K.E., West, R.F., 2008. On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology* 94 (4), 672-695.
- Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. *Science* 185 (4157), 1124-1131.
- Wegener, D.T., Petty, R.E., Blankenship, K.L., Detweiler-Bedell, B., 2010. Elaboration and numerical anchoring: implications of attitude theories for consumer judgment and decision making. *Journal of Consumer Psychology* 20 (1), 5-16.
- Wegener, D.T., Petty, R.E., Detweiler-Bedell, B.T., Jarvis, W., Blair G., 2001. Implications of attitude change theories for numerical anchoring: anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology* 37 (1), 62-69.
- 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 (4), 387-402.
- 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 (1), 68-82.
- Zizzo, D.J., 2010. Experimenter demand effects in economic experiments. *Experimental Economics* 13 (1), 75-98.

APPENDIX

Instructions for Classroom (Experiment 1, *basic*)

Instructions

In this game, you will estimate 10 values. Each value accrues from the determinants A, B, C and D. The determinants A, B and C will be shown to you in each round. The determinant D varies arbitrarily between -25 and 25 from round to round.

Formula:

$$\text{value} = A + B - C + D$$

Speaking is not permitted during the game. The game will take approximately 15 minutes. Of course, your data will be treated anonymously.

x^{th} Round

1. Do you think that the value is higher or lower than the value of the preceding round?

Please tick the box:

<i>higher</i>	<i>lower</i>
---------------	--------------

2. Please enter your estimation:

Note: Question 1 does not apply for the control group.

The Game

In this game, you will estimate a value in each round. There are a total of 15 rounds in which you will give your estimation. In each round, the correct value results from the determinants **A, B, C and D** {Exp4: A, B, C, D and E}. The determinants A, B and C {Exp4: A, B, C and D} will be displayed to you in each round. The determinant D {Exp4: E} varies arbitrarily between -25 and 25 {Exp3: -75 and +75} in each round; you do not know its exact value.

The formula to calculate the value is:

$$\text{value} = \mathbf{A} + \mathbf{B} - \mathbf{C} + \mathbf{D} \text{ \{Exp4: } A+B-0.5C+D^2+E\}}$$

This formula is valid for every round of the game. {Exp2-4 Anchor Treatments: As soon as all players have submitted their estimation at the end of each round, the correct value for each round will be displayed. In the following round, you will also have to estimate whether the value will be higher or lower than that of the preceding round.}

Before the 15 rounds start, you will answer three questions. You have two minutes to answer each question. The game will start once all players have completed this task.

In each round, you will have one minute {Exp4: 30 seconds} to enter your estimations and click on OK to confirm them.

Please note: If you do not enter a number within this minute and confirm it with OK, your payment in the corresponding round will be 0 Euros.

The Payment

Your payment is calculated according to the accuracy of your estimation with regard to the value. The payment is calculated as follows: you receive 50 {Exp3: 80} cents in each round. The difference between your estimation and the value is deducted from your payment in cents. It is not possible for your payment to become negative.

Example: value = 100
 your estimation = 75
 difference between your estimation and the value = 25
 your payment: 50ct. – 25 ct. = **25ct.** {Exp3: 80ct. – 25 ct. = 55ct. }

The gains of each round are added together and paid to you after the end of the game. Furthermore, you will receive € 1 for providing the correct answers to all three preceding questions, as well as a basic payment of € 1.50.

Note: Original instructions were in German. Differences in experiments are indicated by {Exp#:...}. If not indicated, differences apply to both anchor and control treatments.

Chapter III

AN EXPERIMENTAL STUDY ON SOCIAL ANCHORING

with Till Proeger

Published:

Cege Discussion Papers, no. 196, Göttingen University, March 2014.

Download: <http://www.uni-goettingen.de/de/60920.html>

ABSTRACT

The anchoring-and-adjustment heuristic has been studied in numerous experimental settings and is increasingly drawn upon to explain systematically biased decisions in economic areas as diverse as auctions, real estate pricing, sports betting and forecasting. In these cases, anchors result from publicly observable and aggregated decisions of other market participants. However, experimental studies have neglected this social dimension by focusing on external, experimenter-provided anchors in purely individualistic settings. We present a novel experimental design with a socially derived anchor, monetary incentives for unbiased decisions and feedback on performance to more accurately implement market conditions. Despite these factors, we find robust effects for the social anchor, an increased bias for higher cognitive load, and only weak learning effects. Finally, a comparison to a neutral, external anchor shows that the social context increases the bias, which we ascribe to conformity pressure. Our results support the assumption that anchoring remains a valid explanation for systematically biased decisions within market contexts.

Keywords: anchoring; conformity; heuristics and biases; incentives; laboratory experiment

JEL classification: C9; D8; Z2

HIGHLIGHTS

- We present a novel anchoring experiment with an endogenous socially derived anchor.
- We find anchoring effects despite monetary incentives and feedback on performance.
- Learning effects are negligible with respect to anchoring in our repeated task.
- Compared to a neutral anchor, the social anchor substantially increases the bias.
- Social anchoring is a valid explanation for biased behavior in social contexts.

1. INTRODUCTION

The anchoring heuristic is one of the most thoroughly investigated behavioral biases. Following Tversky and Kahnemann's (1974) seminal paper, a considerable body of experimental literature has evolved, that assumes its "*robust and pervasive influence*" (Furnham and Boo 2011, p. 39). However, while anchoring has been investigated comprehensively for individualistic decisions, its social dimension has been neglected to date (Furnham and Boo, 2011). This shortcoming connects to the recent doubts on the universal prevalence of behavioral biases under economic conditions; an argument that has been put forth by List and Millimet (2008), Levitt and List (2007) and Loomes et al. (2003) who argue that monetary incentives connected with feedback can reduce behavioral anomalies through learning effects. Presenting experimental evidence on anchoring effects for willingness-to-pay/-accept, recent studies point to a lack of robustness under economic conditions (Fudenberg et al., 2012; Maniadis et al., 2014; Alevy et al., 2010 for a field experiment; Tufano, 2010; Simonson and Drolet, 2004). As economic transactions take place in social settings that foster learning effects through monetary incentives and the observation of other market participants in repeated tasks, doubts on the unconditional robustness of the anchoring bias seem reasonable.

We thus argue that experimental studies of "*social anchors*" are necessary to more accurately investigate actual anchoring-situations in market contexts. For an example of these social anchors, consider forecasters who anchor their predictions on the publicly available consensus values (Fujiwara et al., 2013; Campbell and Sharpe, 2009). All individual forecasts that constitute the respective consensus values are publicly observable, as is the most recent consensus forecast. Thus, the anchor values are constituted endogenously through the aggregation of prior decisions, while there are strong monetary incentives for unbiased predictions. We assume that this derivation of real-world anchors is applicable to a wide range of economic situations prone to anchoring effects. Endogenous anchors with an observable, social formation thus promise additional external validity in comparison to the classical external experimenter-provided anchors.

Consequently, we aim at establishing for the first time the behavioral impact of a social context on anchoring effects. Our basic expectation is that the implementation of a social anchor setting fosters a bias-reduction through monetary incentives, feedback and learning effects, which are the core elements of the market serving as a "*catalyst for rationality and filter for irrationality*" (List and Millimet, 2008, p.1). However, behavioral research on conformity-seeking behavior (see e.g. Klick and Parisi, 2008) may suggest that the social

derivation of anchor values may ultimately increase the individual adherence to anchor values compared to experimenter-provided external anchors. Our results thus serve at more closely determining whether market conditions have a debiasing effect or even aggravate anchoring effects through conformity pressure.

More specifically, besides adding evidence to the discussion on market forces and biases, we aim at commenting on the growing body of empirical studies in various economic settings that assume anchoring to be the driving force behind systematic distortions in the behavior observed. Recent examples of this trend include art and online auctions (Beggs and Graddy, 2009; Dodonova and Khoroshilov, 2004), real estate purchases (Bucchianeri and Minson, 2013) and sports betting (Johnson et al., 2009; McAlvanah and Moul, 2013). Another large strand of literature draws on prediction behavior with time series data drawn from financial forecasts (Fujiwara et al., 2013), earnings forecasts (Cen et al., 2013), macroeconomic forecasts (Bofinger and Schmidt, 2003; Campbell and Sharpe, 2009) and sales forecasting (Lawrence and O'Connor, 2000). While anchoring does seem like a plausible explanation for the empirical patterns in the respective studies, their experimental base remains inadequate by featuring the classical non-incentivized decisions, external experimenter-given anchors, neither feedback on performance, nor information on other participant's decisions, all of which run contrary to market conditions. For anchoring to hold as an interpretation regarding actual markets, laboratory validations are required that encompass the central features of the decision situations potentially prone to biased decisions.

To further the discussions in the two strands of literature presented, we implement a simple estimation task that allows us to measure the effect of a socially derived anchor while providing economic conditions, i.e. information on the other players' decisions, feedback for learning effects and strong monetary incentives for unbiased decisions. Unlike the classical anchoring studies, we implement a relatively simple rational strategy of taking unbiased decisions. Accordingly, if social anchors have an impact even when avoiding them is rather simple and profitable, we suggest that their actual influence is bound to increase in a more complex decision situation. To account for this notion, we run a second experiment with increased cognitive load. In both experiments, the anchor values result from the aggregated decisions of all participants and contain no additional task relevant information. We thus introduce a "social anchor", whereby the decisions of all other subjects and the resulting average value are displayed. The average value subsequently serves as the anchor for the following round. To qualify the relevance of the social anchor, we finally compare its impact

to results from Meub et al. (2013) who feature an identical experimental setting, but implement a classical external anchor.

In the following, we review the relevant literature to deduct our behavioral hypotheses. Traditional anchoring studies feature an externally given anchor and the additional question of whether participants expect the respective value to be higher or lower than the anchor in numerous variations (see Furnham and Boo, 2011 for a comprehensive review). Furthermore, a basic anchoring effect is shown by Wilson et al. (1996), who find anchoring even without the higher/lower question. Another result (e.g. by Epley and Gilovich, 2005) is that self-generated anchors also lead to robust anchoring effects. Critcher and Gilovich (2008) show how even incidental numbers in the subject's environment bias estimations. However, closest to the investigation of social anchoring is the experiment in Phillips and Menkhaus (2010). They show that an endogenous anchor, constituted by the average results of the respective last round, leads to anchoring effects on the willingness to pay and accept in an auction. They explain the ensuing deterioration of prices in their auction as resulting from the norm of starting a negotiation at the anchor, in this case the average price.

With regard to a socially constituted anchor, the aspect of behavioral conformity may affect results. The human "*meta-preference for conformity*" (Klick and Parisi, 2008, p. 1319) has individuals seeking to conform to the actions of others, ultimately to gain advantages through their affiliation to social groups. Conformity is well-documented in social psychology (see Cialdini and Goldstein, 2004 for a review) and has been applied to economics in various contexts; for instance, as a determinant for contributions to public goods (Carpenter, 2004; Giovinazzo and Naimzada, 2012), regarding coordination externalities in heterogeneous groups (Grajzl and Baniak, 2012), group creativity (Goncalo and Duguid, 2012) or auctions (Beraldo et al., 2013), with Dequech (2013) providing a comprehensive institutional perspective. As our anchor values are explicitly presented as the average prediction of a group, the subconscious drive to adapt to the observed behavior of the other members may enhance anchoring effects, although there is no monetary benefit to conformity.

Conversely, a rational strategy with monetary incentives for unbiased decisions may reduce anchoring. Although Chapman and Johnson (2002, p.125) state that "*incentives reduce anchoring very little if at all*" (referring to the studies of Tversky and Kahnemann, 1974; Wilson et al., 1996 and Epley and Gilovich, 2005), Wright and Anderson (1989) as well as Simmons et al. (2010) show that incentives reduce anchoring if subjects have task familiarity or are provided clues in terms of the direction of adjustment for their initial predictions. Meub et al. (2013) show that monetary incentives reduce anchoring to one-third of its strength when

compared to a non-incentivized setting. We argue that the ambiguous outcomes regarding the impact of incentives reflect the availability of a simple rational strategy in the respective experiments. Once given the realistic opportunity and incentives, subjects tend to act more rationally, which is one of the standard observations in economic experiments (see e.g. Smith and Walker, 1993; Rydval and Ortmann, 2004).

While learning effects in repeated tasks have not yet been investigated concerning their effect on anchoring, a number of studies have shown experts' susceptibility to anchoring. For instance, this has been investigated for car mechanics (Mussweiler et al., 2000), real estate agents (Northcraft and Neale, 1987) and legal experts (Englich and Mussweiler, 2001 and Englich et al., 2005; 2006). Accordingly, Furnham and Boo (2011) summarize that expertise fails to prevent anchoring. However, task specific knowledge has been shown to reduce anchoring by Wilson et al. (1996), as well as by Wright and Anderson (1989). The divergent results on task familiarity point to different processes that elicit anchoring effects (see Crusius et al., 2012). Thus, expert statements may be biased as anchor-consistent knowledge is activated in a cognitively effortful process, whereas in more simple tasks, anchors are used intuitively as a cue to the right answer (Wegener et al., 2001; 2010). Given that the decision situations investigated in empirical anchoring studies can be expected to feature non-intuitive settings, respective experimental studies need to implement cognitively effortful tasks to uphold external validity. Connected to this is the effect of cognitive load on subject's decision quality. Blankenship et al. (2008) show that a mental overload through time pressure and task complexity increases anchoring.

We contribute to the literature reviewed by furthering the knowledge on the effects of anchoring in a social context. This enables us to comment both on the robustness of anchoring under market conditions and on the interpretation of empirical studies that draw on anchoring. Our results show that a socially derived anchor does in fact trigger the anchoring bias, whereby higher cognitive load increases a subject's reliance on the anchor values. When compared to a neutral anchor in an otherwise identical setting, the social anchor has a stronger biasing effect. Thus, the social dimension increases anchoring, which we explain as resulting from conformity pressure. The comprehensive information on the derivation of the anchor and its factual uselessness for individual estimations elicits only weak learning effects. Overall, we state that under market conditions, anchoring effects are increased through implicit conformity pressure, which supports the validity of the empirical studies on anchoring.

The remainder of this paper is organized as follows. Section 2 presents the experimental design, section 3 introduces our behavioral hypotheses and results are presented in section 4, before section 5 concludes.

2. EXPERIMENTAL DESIGN

We report two experiments, both of which comprise a control and an anchor treatment. For all treatments, we implement an estimation task, whereby we ask participants to predict future values. These values result from various determinant values incorporated in a simple formula, which is common knowledge and remains constant during the experiment. One of the determinants is an unknown, uniformly distributed random variable. We thus implement an element of risk, which prevents participants from calculating future values exactly.

The first experiment (henceforth: *BASIC*) implements the formula $x_t = a_t + b_t - c_t + d_t$. x_t is the value participants are asked to predict, while a_t , b_t , c_t are the known determinants and d_t is the random variable that is uniformly distributed over the interval $[-25,25]$. The determinants for the anchor treatments (henceforth *anchor*) are randomly generated within each period on an individual level and shown to participants on the screen.¹ Accordingly, every subject has a different correct value to predict, which is common knowledge. At the end of every round, subjects are shown their individually correct value, the estimations of all other subjects and the average value resulting from these estimations. In the subsequent round, the previous average value is shown to participants as the anchor value on the screen. Note that the display of all estimations and the average prediction, i.e. the social anchor, does not have any additional informational value to the subjects. Participants are further asked whether the value in the current round will be higher or lower than the anchor value, which implements the standard experimental paradigm of anchoring (Tversky and Kahnemann, 1974). The control treatment has an identical design, albeit without the anchor values and the feedback screen.² In all treatments, subjects have one minute to calculate the values; there are fifteen rounds.

¹ We avoid negative values to be predicted and keep the calculations fairly simple. For the *anchor* treatment in *BASIC* determinants were drawn from uniform distributions with $a_t \in [50,150]$, $b_t \in [51,150]$, $c_t \in [0,75]$, $d_t \in [-25,25]$. For *COMPLEX*: $a_t \in [60,150]$, $b_t \in [0,50]$, $c_t \in [0,75]$, $d_t \in [0,10]$, $e_t \in [-25,25]$. Also, subjects are allowed to use the Windows calculator implemented in Ztree (Fischbacher, 2007).

² Note that the control treatments are drawn from a prior experiment, described in Meub et al. (2013). In this case, the determinants were identical for all participants. Given that participants could not observe the other player's estimations, there was no need to implement individual determinants.

The second experiment (henceforth: *COMPLEX*) implements the same setting as *BASIC*, yet introduces a higher cognitive load. Subjects are now asked to make their estimations in 30 seconds and use the formula $x_t = 2a_t - b_t - 0.5c_t + d_t^2 + e_t$, with e_t being the random variable that is again uniformly distributed over the interval $[-25, 25]$. As before, x_t is the value that participants are asked to predict, while a_t , b_t , c_t , d_t are the known determinants in round t . As before, the control treatment does not include the anchor values and the feedback screen.

The payoff in every round for all treatments is fifty cents minus the absolute difference between the respective estimation and the correct value, i.e. $\text{payoff}_{it} = (50\text{cent} - |x_{it} - y_{it}|)$, where y_{it} denotes the prediction of subject i in round t . However, the payoffs in every round could not become negative. Consequently, the rational strategy for all treatments is the calculation of the respective expected value of x_t using the formula and the given determinants. Thus, subjects minimize the expected absolute deviation of their prediction from the correct value, which in turn results in maximizing the payoff. They do so by ignoring the random variable, as its expected value is zero.³ The anchors shown to participants contain no additional information.

For both experiments, given the realizations of the respective random variable, following the rational strategy of predicting the expected values of x_t yields on average about 0.38€ per prediction. Since the anchor values are determined endogenously in *BASIC* and *COMPLEX*, potential gains by solely relying on the anchor value cannot be calculated ex-ante. Accordingly, we rely on the averages of the actual predictions and realizations of the determinants. For *BASIC* (*COMPLEX*), naïve anchoring would only lead to average gains of 0.19€ (0.13€) per period. Thus, there are quite strong monetary incentives for unbiased predictions.

Experiments were conducted in the Laboratory for Behavioral Economics at the University of Göttingen. Participants were recruited with ORSEE (Greiner, 2004) and were only allowed to participate in one session. Experiments were programmed using z-Tree (Fischbacher, 2007). There were six sessions for *anchor* in December 2012 with 58 (57) participants in *BASIC* (*COMPLEX*) and 3(4) sessions for *control* in June 2012 with 44 (35) participants. To achieve homogenous group sizes in *anchor*, we kept the number of subjects in a session close to constant, as we had 18/19/20(20/20/18) participants for *BASIC* (*COMPLEX*). All sessions

³ For example, consider for *BASIC* that $a_t=100$, $b_t=80$, $c_t=20$. Plugging in the values into the formula gives $x_t=100+100-40+d_t=160+d_t$. Since subjects now that all values within the interval of $[-25, 25]$ are equally likely realizations of the random variable d_t , they optimally assume d_t to be zero and predict 160 as the future value.

lasted around thirty minutes. Participants earned €5.89 on average. They were on average 22.7 years old, with 59% being female.

3. HYPOTHESES

Our setting in principal uncovers the anchor itself as useless and thus fosters rational decisions that maximize payoffs. Given, however, that both anchoring and conformity pressure have been shown to profoundly influence human decisions, we hypothesize that subjects are biased to the social anchors. Observation of other players is thus expected to only inefficiently reduce anchoring. Cognitive load is expected to increase the reliance on anchor values. We thus formulate hypothesis 1:

Hypothesis 1 (“*Rationality and anchoring bias*”). Subjects’ estimations are biased towards the social anchor.

Furthermore, we are interested in learning effects. We hypothesize that the comprehensive information on both the anchor and the other player’s decisions, as well as the repetition of the identical task, leads to the buildup of task-specific knowledge. As a number of anchoring studies claim that this does not prevent anchoring, we hypothesize that learning effects fail to reduce social anchoring.

Hypothesis 2 (“*Learning effects*”). The anchoring bias is not reduced by learning effects.

Lastly, we are interested in the magnitude of the anchoring effects triggered by social anchors. By comparing the effects to a treatment that features a neutral anchor drawn from Meub et al. (2013), we can estimate which anchor type more strongly affects decisions. As conformity pressure appears likely to influence subjects’ decisions, we hypothesize that the social anchor has a stronger impact.

Hypothesis 3 (“*neutral- versus social anchor*”). Subjects are more strongly biased towards a social anchor.

4. RESULTS

We present the experimental results in four subsections. First, we compare performance between treatments in terms of prediction accuracy. Second, we investigate if these differences point to an anchoring bias. Third, we test for learning effects. Fourth, we compare the results of the anchor treatments to a neutral, exogenous anchor in an otherwise similar setting to measure the relevance of the social anchor. This enables us to disentangle the basic underlying anchoring bias and additional effects potentially triggered by the social anchor.

4.1 OVERALL PERFORMANCE AND RATIONALITY

We analyze overall performance by considering subjects' prediction accuracy and rationality conditional on experiment and treatment. Recall that the profit-maximizing strategy of estimating the expected value is identical for all experiments and treatments, as the average prediction of the previous round does not contain any additional information. Moreover, only subjects in the treatment groups answer the higher/lower-question, which also does not affect rational behavior.⁴

Table 1 summarizes the mean absolute deviation of predictions from the expected values and the share of optimal estimations by treatments for all experiments.⁵ Predictions equal to the expected value are characterized as optimal.⁶

The prediction accuracy and the share of optimal predictions are higher for *BASIC*. More interestingly, there are significant differences between *anchor* and *control*. As a measure for overall prediction accuracy, the median absolute deviation clearly indicates that subjects'

⁴ In *BASIC (COMPLEX)* 73% (68%) of the higher/lower-questions were answered correctly.

⁵ The dataset contains 181 missing values (78 in the treatment groups and 103 in the control groups) when subjects did not enter a value in the respective round. Additionally, the dataset is corrected for subjects' predictions if the task was obviously misinterpreted. We assume this to be true if the estimation of subject i in period t (y_{it}) is smaller than 25 or negative ($y_{it} < 25$), i.e. subjects tried to estimate the random determinant and not the future value. Thus, 162 observations (8 in the treatment groups and 154 in the control groups; for *anchor (control)* 6(128) in *BASIC* and 2(26) in *COMPLEX*) were recoded as missing values. We assume that the higher number of predictions showing misunderstanding of the task in the control group is due to the lack of feedback or correction by observing others as in the treatment group. This leaves us with 2670 individual decisions.

⁶ The optimality criterion is eased for predictions that deviate by a maximum of 0.5 points in the treatment groups. This adjustment is necessary as the expected values were not integers in all rounds of *COMPLEX* and thus deviations of 0.5 only reflect the restriction for subjects having to round entered values to be integers. Accordingly, we round down absolute deviations by 0.5 points in such cases.

performance is worse when they are shown the social anchor (K-sample test; for *BASIC* $\chi^2(1)=25.5955$ with $p=0.000$, for *COMPLEX* $\chi^2(1)=13.9095$ with $p=0.000$). Considering the share of optimal predictions, we also find a significantly better performance for the control groups (Fisher’s exact; $p=0.000$ for both experiments). As *mean corrected* shows, the mean absolute deviation is somewhat misleading, given that results are driven by outliers in the control treatments, particularly in *COMPLEX*. Applying a Wilcoxon-Rank-Sum test on *mean corrected* points at significantly better performance in *control* (for *BASIC* $z=-6.441$, $p=0.000$; for *COMPLEX* $z=-5.744$, $p=0.000$).

Experiment	<i>BASIC</i>		<i>COMPLEX</i>		
	<i>control</i>	<i>anchor</i>	<i>control</i>	<i>anchor</i>	
absolute deviation	mean (std. dev.)	8.42 (21.27)	11.62 (23.71)	32.48 (102.04)	28.12 (42.62)
	mean corrected (std. dev.)	5.75 (9.47)	8.61 (13.47)	17.74 (29.73)	23.90 (31.29)
	median	0	4	10	11
	75%	10	13	25	35
	95%	30	50	100	109
share optimal	0.56	0.33	0.37	0.10	

Table 1: Treatment comparison subject to prediction accuracy.

Note: Values for *mean corrected* originate from treating values greater than the 97.5th percentile as outliers excluded from the calculation.

We conclude that the social anchor has a strong impact on overall prediction accuracy. These differences are economically relevant since the average absolute deviation, once corrected for outliers, increases by around 50% in *BASIC* and around 35% in *COMPLEX*. The share of optimal predictions is 23 percentage points higher in *control* of *BASIC* and 27 p.p. for *COMPLEX*. However, differences in performance do not necessarily indicate an anchoring bias.⁷ Therefore, we further investigate the distinct distribution of deviations from the expected values.

⁷ An alternative explanation might be seen in the representativeness bias (Kahnemann and Tversky, 1973). The distribution of predictions might reflect the distribution of the equation’s random determinant. Forecasters have been shown to display the tendency of replicating the distribution of a time series’ noise. Therefore, they incorporate the uncertainty rather than ignoring it for an optimal prediction (Harvey, 2007). Subjects in treatment

4.2 A SYSTEMATIC ANCHORING BIAS

If predictions in the treatment groups are systematically biased toward the averages of previous rounds, we interpret this as evidence in support of Hypothesis 1. The following graphs show the distribution of deviations from the expected values.

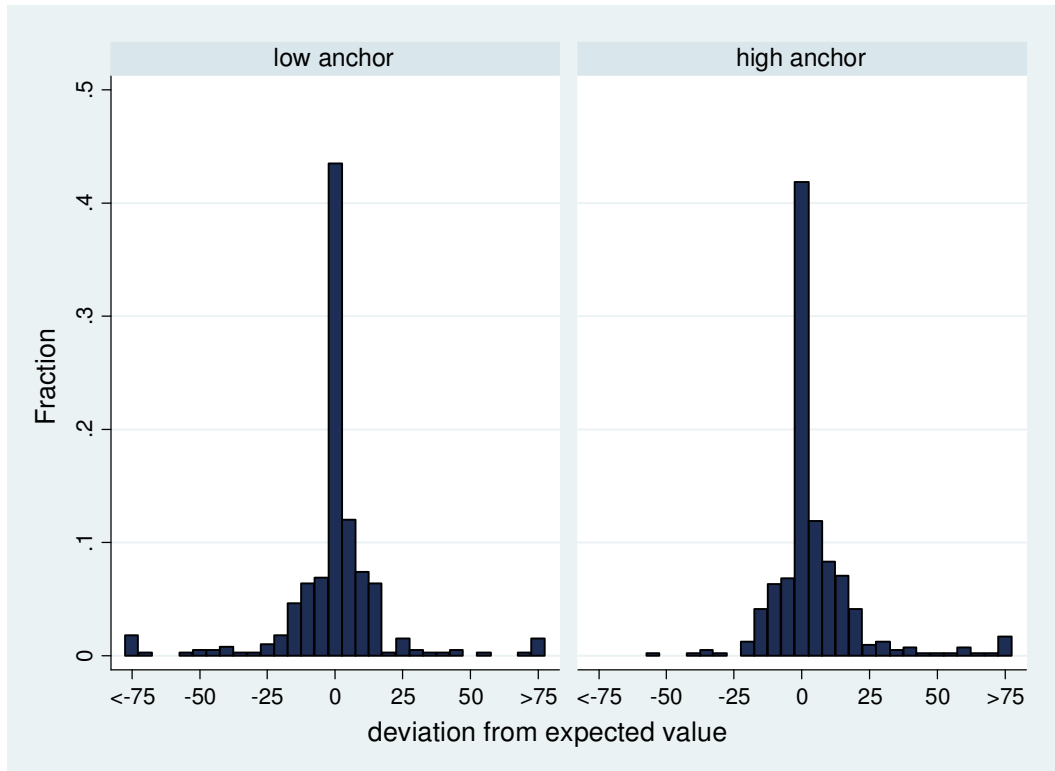


Figure 1: Distribution of predictions for *BASIC*

In *BASIC*, there are no apparent differences in the distribution of the deviations from the expected value for the treatment group with respect to the direction of the anchor value shown.

By contrast, in *COMPLEX*, we find the typical pattern resulting from a systemic anchoring bias: for low anchor values, more predictions are too low, i.e. smaller than the expected value. Considering the right hand side shows the same pattern for the case of high anchor values.

However, the distributions of deviations conditional on the anchor direction in the treatment groups significantly differ for both experiments (Two-sample Kolmogorov-Smirnov test; for *BASIC* corrected p-value=0.080, for *COMPLEX* corrected p-value=0.000).

groups might be more prone to the representativeness bias since they are also confronted with the noise in the predictions given by the other subjects.

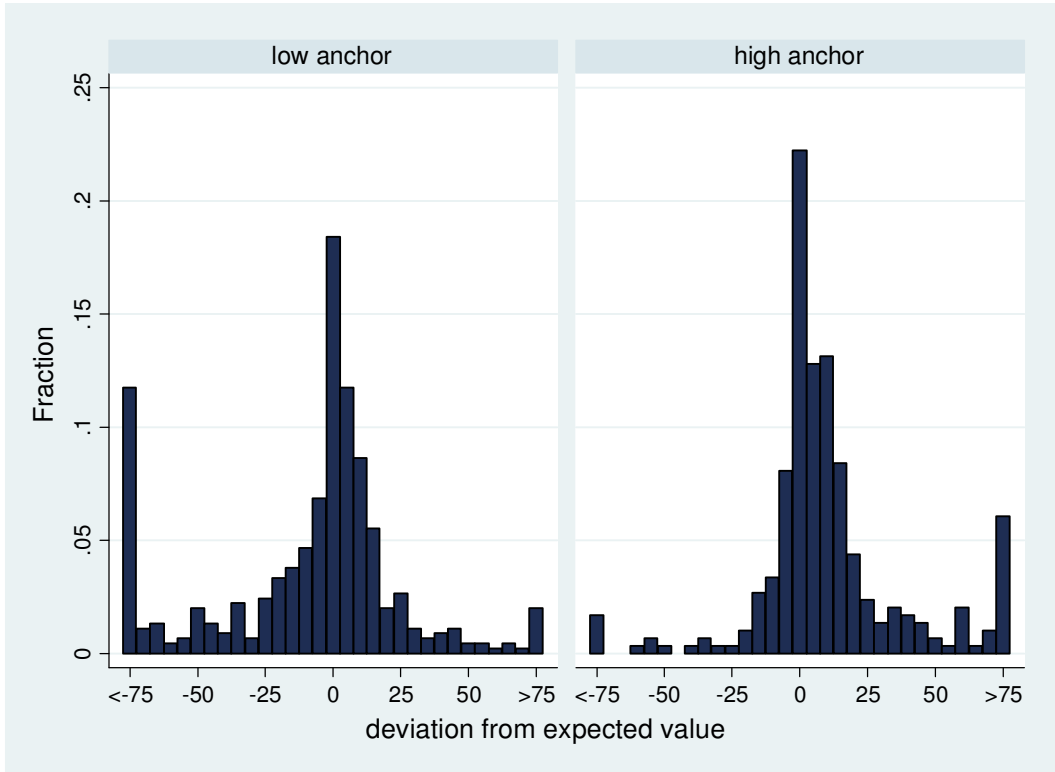


Figure 2: Distribution of predictions for *COMPLEX*

For a more profound analysis and quantification of the anchoring bias, we test for a specific anchoring pattern in the estimations by running a regression.

Equation (1) presents the model to explain the subjects' predictions. Let y_{it} denote the estimation of subject i at period t , and x_t the realized value at time t , with $E(x_t)$ giving its expected value.

$$y_{it} = \gamma_1 E(x_t) + \theta_1 [E(x_t) - \bar{y}_{t-1}] + u_{it} \quad (1)$$

In the given context, an optimal prediction of x_t can be explained by the expected value (*expected_value*) $E(x_t)$ only, i.e. ($\gamma_1=1$). However, we are interested in a potential bias caused by the anchor value, which is the average prediction of the previous round, denoted as \bar{y}_{t-1} . We include the term $\theta_1 [E(x_t) - \bar{y}_{t-1}]$ (*anchor_deviation*) to control for an anchoring bias, measuring the deviation of the average prediction of the previous round \bar{y}_{t-1} and the expected value in the current round $E(x_t)$. An unbiased estimation is given if $\theta_1=0$, whereas an estimation biased toward the anchor value is given if $\theta_1 < 0$.

In sum, information is used efficiently if a regression of Eq. (1) results in an estimation of γ_1 that is not significantly different from 1. At the same time, all other variables should show an insignificant effect on the values predicted. In this case, there would be no evidence supporting H1, but rather indicating that on average and *ceteris paribus* estimations are made optimally and are unbiased.

Additionally, we extend the model to allow for learning effects, an aspect discussed in subsection 4.3. Therefore, we introduce an interaction term for the anchoring bias picking up changes after the first five periods, i.e. for the last ten periods, as well as for the last five periods: the dummy variables to identify the periods are denoted as P^{10} and P^5 respectively. We can write the regression model as follows:

$$y_{it} = \gamma_1 E(x_t) + \theta_1 [E(x_t) - \bar{y}_{t-1}] + \theta_2 [(E(x_t) - \bar{y}_{t-1})P^{10}] + \theta_3 [(E(x_t) - \bar{y}_{t-1})P^5] + u_{it} \quad (2)$$

Table 2 provides the results of a pooled OLS regression on our unbalanced panel dataset of Eq. (1) and Eq. (2), applying robust Driscoll and Kraay standard errors.

Equation	(1)	(1)	(2)	(2)
Experiment	<i>BASIC</i>	<i>COMPLEX</i>	<i>BASIC</i>	<i>COMPLEX</i>
expected_value	1.0138*** (.0051)	1.0033*** (.0046)	1.0138*** (.0051)	1.0059*** (.0039)
anchor_deviation	-.0934*** (.0129)	-.2188*** (.0454)	-.1145*** (.0080)	-.2618** (.1048)
anchor_deviation*P10			.0261 (.0161)	.0358 (.1015)
anchor_deviation*P5			.0094 (.0296)	.0580*** (.0169)
F-Statistic ($\gamma_1=1$)	7.25**	0.50	7.34**	2.25
Prob. > F	(.0184)	(.4911)	(.0179)	(.1573)
F-Statistic ($\theta_2 = \theta_3 = 0$)			1.69	5.93**
Prob. > F			(.2229)	(.0148)
F-Statistic ($\theta_1 = \theta_2 = \theta_3 = 0$)			102.87 ***	110.76***
Prob. > F			(.0000)	(.0000)
Observations	802	748	802	748
No. of Groups	58	57	58	57

Table 2: OLS regression of Eq. (1) and Eq. (2) with estimations (y_{it}) as dependent variable.

Note: Robust Standard Errors in parentheses; for F-Statistics p-value in parentheses. (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

For both experiments, we find a negative and significant effect of the deviation in the anchor value ($\theta_1 < 0$), which has to be interpreted as an on average bias towards the average estimation of the previous period.⁸ For a decreasing (increasing) expected value in t compared

⁸ In the following, we describe two main tests for the robustness of our results. First, we control for the influence of the deviation of the previous round's estimation y_{it-1} from the expected value of the current round $E(x_t)$. This

to the average prediction in $t-1$, subjects in the treatment groups give significantly higher (lower) estimations. This has to be interpreted as a systematic inability to ignore the average estimation of the previous round. Additionally, subjects in *BASIC* fail to predict optimally on average, given that the marginal effect for the expected value ($\gamma_1 > 1$) indicates a general overestimation of the values to be predicted.

Besides the significance of the bias, its relevance has to be addressed. Based on the average absolute difference of the anchor values and the expected values of 37.5 points in *BASIC* (60.3 in *COMPLEX*), the estimated marginal effect of -0.093 (-0.219) amounts to a ceteris paribus bias of 3.4875 (13.2057) points on average. This corresponds to 2.1% (6.5%) of the average correct values. The cognitive load evidently has a strong influence on anchoring, as the magnitude of the bias is tripled for *COMPLEX*. As already indicated by Figure 1, there are only small effects caused by the anchor value in *BASIC*, although an estimated bias for each prediction of all individuals in each round of 2.1% has to be considered as economically relevant.

We conclude that the additional, albeit useless, information shown in the treatment groups creates a general noise in subjects' estimations. The social anchor values have an overall significant and relevant impact, especially when cognitive load is high. On average, subjects are unable to ignore the averages determined by the predictions of all players, as the rational strategy would suggest. Note that the anchoring bias is not only driven by subjects performing poorly who subsequently draw on these values, since the share of optimal predictions significantly deteriorates at the same time. Therefore, we interpret our results as presenting strong evidence in favor of H1.

control variable might be required due to the possible correlation of predictions made in consecutive rounds. Since the individual and the average estimation of previous rounds are positively correlated, this would lead to omitted variable bias. We find a significant marginal effect of the difference between last round's prediction and the current expected value. However, the effects are rather small (-0.0417 in Eq. (1) for *BASIC*, -0.0796 in Eq. (1) for *COMPLEX*, -0.0412 in Eq. (2) for *BASIC*, -0.0892 in Eq. (2) for *COMPLEX*). Most importantly, the anchor value's deviation and the interaction term in *COMPLEX* remain highly significant for all models, although marginal effects of the anchor deviation are slightly lower (-0.0414 in Eq. (1) for *BASIC*, -0.1533 in Eq. (1) for *COMPLEX*, -0.0646 in Eq. (2) for *BASIC*, -0.2452 in Eq. (2) for *COMPLEX*). Second, we test the robustness of our results with respect to outliers, as they might drive the results. Therefore, we exclude all predictions that deviate by more than three times the maximum value of the random determinant, i.e. if $y_{it} < [E(x_t) - 3 * 25]$ or $y_{it} > [E(x_t) + 3 * 25]$. Again, the anchor deviation is still significant for all models. Unsurprisingly, marginal effects of the anchor deviation are estimated to be somewhat smaller (-0.0456 in Eq. (1) for *BASIC*, -0.1005 in Eq. (1) for *COMPLEX*, -0.0806 in Eq. (2) for *BASIC*, -0.0896 in Eq. (2) for *COMPLEX*).

4.3 LEARNING EFFECTS

In H2, we hypothesized that learning effects should be absent as task-specific knowledge generally fails to prevent biased decisions. There is evidence in support of learning effects when considering the regression results of Eq. (2) for *COMPLEX*. For *BASIC*, an F-test on the interaction terms of the anchor deviation and the last ten and last five periods fails to reject the null of no-joint-significance (p-value=0.2229). However, for *COMPLEX*, the results clearly indicate a reduction of the anchoring bias as the game proceeds. Figure 3 presents the development of the share of optimal predictions over periods, which supports the notion that there are slight learning effects.

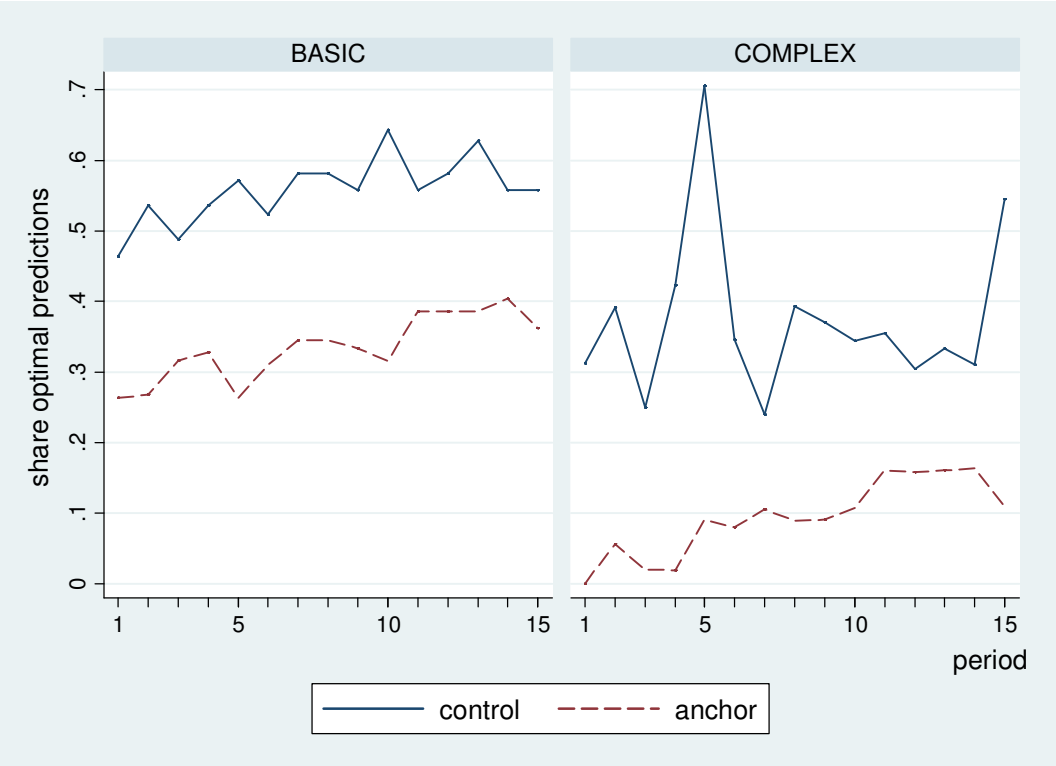


Figure 3: Share of optimal predictions over periods.

The graphs points to learning effects for both experiments, given that the share of optimal predictions increases gradually over time, although such evidence is weaker for *BASIC*.⁹ In the first period of *COMPLEX*, there are no optimal predictions in the treatment group, while the share of optimal predictions in the last five periods amounts to more than 10%.

⁹ The extremely high share of optimal predictions in period five of the control group in *COMPLEX* is due to the expected value being equal to 100. Thus, many subjects applying a rule of thumb hit the expected value, rather by accident than through a correct calculation.

In sum, we find mixed evidence regarding H2. There are some learning effects regarding the general understanding of the game. While there seem to be no learning effects in terms of the reduction of the anchoring bias for *BASIC*, performance improves over time for *COMPLEX*, as the anchoring bias tends to weaken. However, learning effects are rather weak as the share of rational predictions is strictly lower for all periods compared to the control treatments. Also, the magnitude of the marginal effects of the anchor interaction terms are quite small.

4.4 COMPARING THE SOCIAL TO A NEUTRAL ANCHOR

Lastly, we compare the impact of the social anchor to the results of a similar anchoring experiment, drawn from Meub et al. (2013). This study comprehends two analogous treatments implementing the same calculation tasks as in *BASIC* and *COMPLEX*, but does not feature a social anchor. Instead, subjects are merely displayed the correct value after each round, which subsequently becomes the anchor value for the ensuing round. There is no social observation, only feedback on the correct value after each round. All other factors are identical.

By comparing the two experiments, we aim to disentangle the impact of other behavioral influences towards the social anchor such as conformity pressure and a basic anchoring effect that is independent of the social context. Accordingly, Figure 4 shows the average absolute deviation over periods for the social and the neutral anchor.

As can be seen from the graphs, there are only small differences between the social and neutral anchor in *BASIC*, whereby subjects facing the social anchor perform slightly worse in the first five rounds. For *COMPLEX*, we find subjects in the neutral anchor treatment to perform better, with the exception of period 12.¹⁰ Overall, performance is better, since the average absolute deviations pooled for all periods is 7.9 (11.6) for the neutral anchor (social anchor) in *BASIC* and 17.5 (28.2) in *COMPLEX*. A Mann-Whitney test shows significant differences between the social and neutral anchor treatment for *BASIC* ($z=-2.374$, $p=0.0176$) and *COMPLEX* ($z=-5.680$, $p=0.0000$).

¹⁰ The somewhat extreme values of period five and twelve for the neutral anchor in *COMPLEX* can be ascribed to the experimental design. Since subjects in the neutral anchor treatment could not observe each other and were presented the same determinant values for their calculation and the same anchor values, specific characteristics in some rounds affect all estimations homogenously. Thus, there is more noise on average between periods, which cancel out in the social anchor treatment with individual determinant values. For example, in period five and twelve of *COMPLEX*, the determinant d is at its maximum and has to be squared to calculate the expected value, which causes high individual deviations from the expected value if subjects fail to do so.

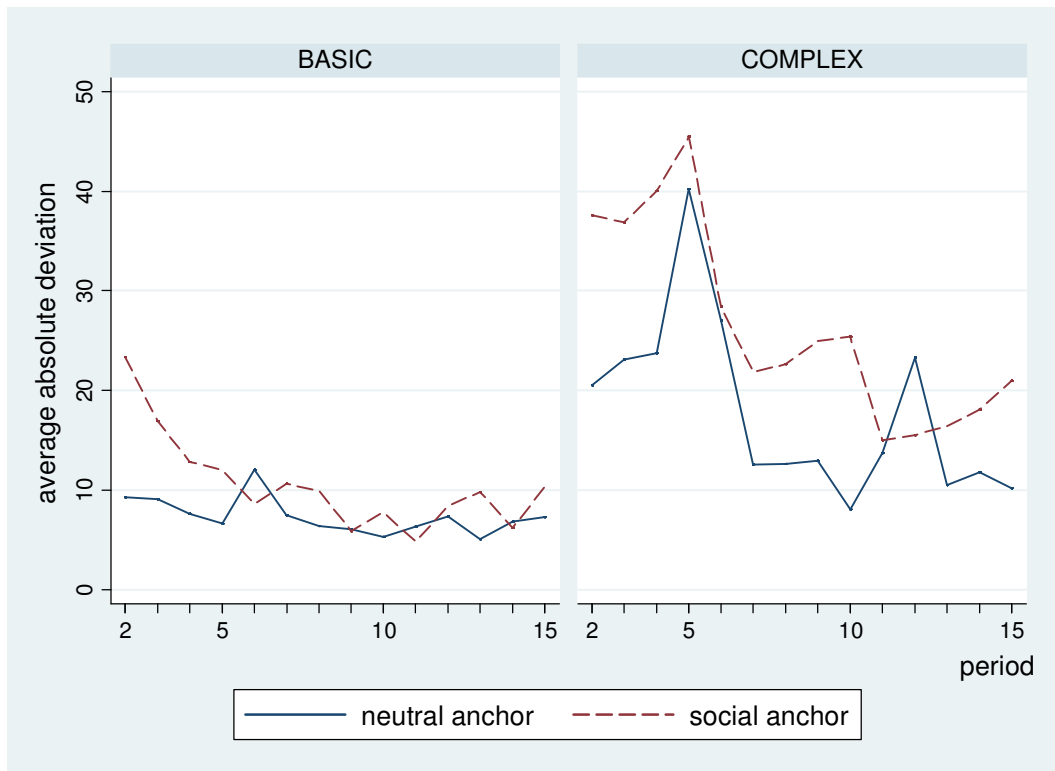


Figure 4: Average absolute deviations over periods by anchor setting.

Additionally, running a similar analysis for the neutral treatment as in subsection 4.3, Meub et al. (2013) show the same pattern of a systematic anchoring bias. For *BASIC*, the anchoring bias amounts to 0.94% of the average value to predict, compared to 2.1% for the social anchor. In *COMPLEX*, the average bias is 2.11% for the neutral treatment and 6.5% for the social anchor.

We conclude that subjects facing a social anchor are even more prone to the anchoring bias than those facing a neutral anchor. This result holds although the anchor is obviously useless for correct estimations in our setting, since it is common knowledge that every subject receives individual, random values for the determinants. Apparently, a social environment increases anchoring rather than reducing it through additional information on the anchor itself. We ascribe this effect to the observation of other players and hypothesize that it subconsciously activates subjects’ “*meta preference for conformity*” (Klick and Parisi, 2008, p. 1319). Thus, the average values lead subjects to conform to the perceived group norm. While our design gives evidence for the social anchoring effect, it does not provide unambiguous proof for an interpretation based on conformity pressure. However, we find evidence in support of H3, i.e. a stronger effect of social anchors.

5. CONCLUSION

In line with Furnham and Boo (2011), this study argues that research on the anchoring bias has neglected to consider its social dimension and focused on purely individualistic choices instead. This limits the external validity of the experimental anchoring studies, as actual markets feature extensive opportunities for learning through observation of the other economic agents. By implementing an observational framework with socially derived endogenous anchors, our setting more closely resembles the decisions faced by subjects in the markets commonly investigated in empirical studies, such as auctions or forecasting. Further, by implementing strong monetary incentives and feedback on performance, we are able to present arguments as to whether lab-implemented market forces may serve as a filter for irrational decisions (List and Millimet, 2008).

In spite of monetary incentives, a simple rational strategy and feedback on performance, we find a strong anchoring bias resulting from the socially derived anchor, which increases along with higher task complexity. There are only small learning effects in the case of high cognitive load. Thus, the obvious derivation of the anchor values through the prior decisions of other subjects does not succeed in eliminating the bias. Finally, the comparison to a neutral anchor shows that, overall, a social anchor leads to substantially stronger effects than a classical external one. We explain this as resulting from the implicit pressure of conforming to the average decisions of all other subjects, despite its factual irrelevance.

Our study thus does not support the notion that market conditions may generally serve as a remedy for behavioral biases. We argue that they may instead foster other influences, such as conformity pressure towards consensus values. Consequently, our results lend experimental support to the empirical studies that report biases towards social anchors in economic contexts. We suggest that their interpretation is valid, given that market conditions in our study not only fail to eliminate the bias, but rather increases it. To gain a more profound understanding of anchoring in social contexts, further experimental studies should more closely investigate the interdependencies of anchoring and other behavioral influences, such as conformity pressure.

ACKNOWLEDGEMENTS

We are grateful to Kilian Bizer and Markus Spiwoks for their helpful comments. Further, we thank participants of the *cege* and *GLOBE* doctoral research colloquium at Goettingen University. Tonia Sauerwald provided excellent student assistance in programming and conducting the experiment.

REFERENCES

- Alevy, J.E., Landry, C.E., List, J.A., 2011. Field Experiments on Anchoring of Economic Valuations. University of Alaska Anchorage, Department of Economics, Working Paper No. 2011-02.
- Beggs, A., Graddy, K., 2009. Anchoring effects: evidence from art auctions. *The American Economic Review* 99 (3), 1027-1039.
- Beraldo, S., Filoso, V., Marco, S., 2013. Endogenous Preferences and Conformity: Evidence from a Pilot Experiment. MPRA Paper No. 48539.
- 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 (6), 1465-1476.
- 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 (4), 437-449.
- Bucchianeri, G.W., Minson, J., 2013. A homeowner's dilemma: Anchoring in residential real estate transactions. *Journal of Economic Behavior & Organization* 89 (6), 76-92.
- 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 (2), 369-390.
- Carpenter, J.P., 2004. When in Rome: conformity and the provision of public goods. *The Journal of Socio-Economics*, 33 (4), 395-408.
- 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 (1), 47-76.

- Chapman, G.B., Johnson, E.J., 2002. Incorporating the irrelevant: Anchors in judgments of belief and value, in: Gilovich, T., Griffin, D., Kahneman, D. (Eds.), *The Psychology of intuitive Judgment: Heuristics and Biases*. New York: Cambridge University Press, 120-138.
- Cialdini, R.B., Goldstein, N.J., 2004. Social influence: Compliance and conformity. *Annual Review of Psychology* 55, 591-621.
- Critcher, C.R., Gilovich, T., 2008. Incidental environmental anchors. *Journal of Behavioral Decision Making* 21 (3), 241-251.
- Crusius, J., van Horen, F., Mussweiler, T., 2012. Why process matters: a social cognition perspective on economic behavior. *Journal of Economic Psychology* 33 (3), 677-685.
- Dequech, D., 2013. Economic institutions: explanations for conformity and room for deviation. *Journal of Institutional Economics* 9 (1), 81-108.
- Di Giovinazzo, V., Naimzada, A., 2012. ... Do as the Romans do. A model of conformity with the endogenous formation of preferences. *The Journal of Socio-Economics* 41 (5), 654-658.
- Dodonova, A., Khoroshilov, Y., 2004. Anchoring and transaction utility: evidence from online auctions. *Applied Economic Letters* 11 (5), 307-310.
- Englich, B., Mussweiler, T., 2001. Sentencing under uncertainty: anchoring effects in the courtroom. *Journal of Applied Social Psychology* 31 (7), 1535-1551.
- Englich, B., Mussweiler, T., Strack, F., 2005. The last word in court: a hidden disadvantage for the defense. *Law and Human Behavior* 29 (6), 705-722.
- Englich, B., Mussweiler, T., Strack, F., 2006. Playing dice with criminal sentences: the influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin* 32 (2), 188-200.
- 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 (3), 199-212.
- Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2), 171-178.

- Fudenberg, D., Levine, D.K., Maniadis, Z., 2012. On the robustness of anchoring effects in WTP and WTA experiments. *American Economic Journal: Microeconomics* 4 (2), 131-145.
- 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.
- Furnham, A., Boo, H.C., 2011. A literature review of the anchoring effect. *The Journal of Socio-Economics* 40 (1), 35-42.
- Goncalo, J.A., Duguid, M.M., 2012. Follow the crowd in a new direction: When conformity pressure facilitates group creativity (and when it does not). *Organizational Behavior and Human Decision Processes*, 118 (1), 14-23.
- Grajzl, P., Baniak, A., 2012. Mandating behavioral conformity in social groups with conformist members. *Journal of Economic Behavior & Organization*, 82 (2), 479-493.
- Greiner, B., 2004. An online recruitment system for economic experiments. *GWDG Berichte* 63, 79-93.
- Harvey, N., 2007. Use of heuristics: Insights from forecasting research. *Thinking & Reasoning* 13 (1), 5-24.
- 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.
- 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 (4), 410-434.
- Kahneman, D., Tversky, A., 1973. On the psychology of prediction. *Psychological Review* 80 (4), 237-251.
- Klick, J., Parisi, F., 2008. Social networks, self-denial, and median preferences: Conformity as an evolutionary strategy. *Journal of Socio-Economics* 37 (4), 1319-1327.
- Lawrence, M., O'Connor, M., 2000. Sales forecasting updates: how good are they in practice? *International Journal of Forecasting* 16 (3), 369-382.
- Levitt, S.D., List, J.A., 2007. What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives* 21 (2), 153-174.
- 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 (1), 1-53.

- Loomes, G., Starmer, C., Sugden, R., 2003. Do Anomalies Disappear in Repeated Markets? *The Economic Journal* 113 (486), 153-166.
- 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.
- 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.
- Meub, L., Proeger, T.E., Bizer, K., 2013. Anchoring: a valid explanation for biased forecasts when rational predictions are easily accessible and well incentivized? *Cege Discussion Papers* No. 166.
- Mussweiler, T., Strack, F., Pfeiffer, T., 2000. Overcoming the inevitable anchoring effect: considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin* 26 (9), 1142-1150.
- Northcraft, G.B., Neale, M.A., 1987. Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes* 39 (1), 84-97.
- Phillips, O.R., Menkhous, D.J., 2010. The culture of private negotiation: Endogenous price anchors in simple bilateral bargaining experiments. *Journal of Economic Behavior & Organization*, 76 (3), 705-715.
- Rydval, O., Ortmann, A., 2004. How financial incentives and cognitive abilities affect task performance in laboratory settings: an illustration. *Economics Letters* 85 (3), 315-320.
- 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 (6), 917-932.
- Simonson, I., Drolet, A., 2004. Anchoring Effects on Consumers' Willingness-to-Pay and Willingness-to-Accept. *Journal of Consumer Research* 31 (3), 681-90.
- Smith, V.L., Walker, J., 1993. Monetary rewards and decision cost in experimental economics. *Economic Inquiry* 31 (2), 245-261.
- Tufano, F., 2010. Are 'True' Preferences Revealed in Repeated Markets? An Experimental Demonstration of Context-Dependent Valuations. *Experimental Economics* 13 (1), 1-13.

- Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. *Science* 185 (4157), 1124-1131.
- Wegener, D.T., Petty, R.E., Blankenship, K.L., Detweiler-Bedell, B., 2010. Elaboration and numerical anchoring: implications of attitude theories for consumer judgment and decision making. *Journal of Consumer Psychology* 20 (1), 5-16.
- Wegener, D.T., Petty, R.E., Detweiler-Bedell, B.T., Jarvis, W., Blair G., 2001. Implications of attitude change theories for numerical anchoring: anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology* 37 (1), 62-69.
- 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 (4), 387-402.
- 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 (1), 68-82.

APPENDIX

Instructions for *BASIC* and *COMPLEX*. The differences between experiments and treatments are indicated in braces. The original instructions were in German and are available from the authors upon request.

The Game

In this game, you will estimate a value in each round. There are 15 rounds in which you will give your estimation. In each round, the correct value results from the determinants **A, B, C and D** {*COMPLEX: A, B, C, D and E*}. The determinants A, B and C {*COMPLEX: A, B, C and D*} will be displayed to you in each round. The determinant D {*COMPLEX: E*} varies arbitrarily between -25 and 25 in each round; you do not know its exact value.

The formula to calculate the value is:

$$\text{value} = \mathbf{A + B - C + D} \text{ \{COMPLEX: } 2*A - B - 0.5*C + D^2 + E\}}$$

This formula is valid for every round of the game. {*BASIC, COMPLEX - anchor treatments: As soon as all players have submitted their estimation at the end of each round, the estimations of all the other players will be displayed, as well as the resulting average estimation. Starting from the second round, you will also have to estimate whether the value will be higher or lower than the average estimation of the preceding round.*} In each round, you will have one minute {*COMPLEX: 30 seconds*} to enter your estimation and click on OK to confirm it.

Please note: If you do not enter a number within this minute {*COMPLEX: these 30 seconds*} and confirm it with OK, your payment in the corresponding round will be 0 Euros.

The Payment

Your payment is calculated according to the accuracy of your estimation with regard to the value. The payment is calculated as follows. You receive 50 cents in each round, with the difference between your estimation and the correct value being deducted in cents. It is not possible for your payment to become negative.

Example: value = 100
 your estimation = 75
 difference between your estimation and the value = 25
 your payment: 50ct. – 25 ct. = **25ct.**

The gains of each round are added together and paid to you after the end of the game. Furthermore, you will receive a basic payment of € 1.50.

Chapter IV

ARE GROUPS LESS BEHAVIORAL? THE CASE OF ANCHORING

with Till Proeger

Published:

Cege Discussion Papers, no. 188, Göttingen University, February 2014. Download: <http://www.uni-goettingen.de/de/60920.html>

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 estimates. In contrast to previous anchoring studies, we find groups to successfully reduce, albeit not eliminate, anchoring in factual knowledge tasks. For the other domains, groups and individuals are equally biased by external anchors. We thus suggest that group cooperation reduces biases prevalent on the individual level for predominantly intellectual tasks, yet fails to improve decision-making when judgmental aspects are involved.

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

JEL Classification: C91; C92; D8

HIGHLIGHTS

- We compare group and individual performance in a classical anchoring setting.
- We consider 3 task domains: factual knowledge, probability- and price-estimations.
- Both groups and individuals are systematically biased by external anchors.
- Groups succeed in reducing, yet not eliminating, anchoring in intellectual tasks.
- Group decision-making fails to reduce anchoring in judgmental tasks.

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 increasing 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 consider the case of the anchoring-and-adjustment heuristic, initially presented by Tversky and Kahneman (1974). Hence, we choose a prominently discussed heuristic, which is increasingly used as a stylized fact to explain specific behavioral patterns in quite diverse economic situations.² Forty years of psychological anchoring studies have found the anchoring bias to be fairly robust against experimental variations (cp. the literature

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; 2005); Levitt and List (2007); Cecchi and Bulte (2013).

² Recent examples include real estate price forecasts (Bucchianeri and Minson, 2013), 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).

review by Furnham and Boo, 2011). Contrasting this view, recent economic field experiments on anchoring in price valuations find only moderate effects (Simonson and Drolet, 2004; Bergman et al., 2010; Tufano, 2010; Alevy et al., 2011; Fudenberg et al., 2012; Manianidis et al., 2014.) These results support the notion of market conditions correcting irrational consequences 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 previously overlooked in experimental studies.

Consequently, 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 estimates and implement strong monetary incentives for unbiased decisions for groups and individuals alike. We argue that these different domains of decision-making cover a wide range of economically relevant situations that are potentially prone to irrationally anchored decisions. While our anchoring tasks 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.

A large body of literature has dealt with the causes, consequences and potential strategies of avoiding the anchoring bias, with Furnham and Boo (2011) providing a general overview. Following Tversky and Kahnemann's (1974) seminal paper, three broad strands of literature have evolved discussing theoretical foundations of the anchoring bias. As outlined by Epley and Gilovich (2010), the first wave of anchoring research focused on establishing anchoring as a robust judgmental distortion and ensuring that that it was in fact a non-rational bias. The second wave attempts to identify and model the exact psychological mechanisms that trigger and influence anchoring within the classical experimental paradigms, such as confirmatory hypothesis testing, numeric / magnitude priming and insufficient adjustment from given values. Epley and Gilovich (2010) provide an overview of the past literature, while Frederick and Mochon (2012) and Critcher et al. (2014) provide the latest investigations for this domain. A recent third wave of research is concerned with assessing the influence of varying contextual factors on the elicitation of anchoring effects, drawing on the attitude and

persuasion perspective in psychological research. Wegener et al. (2010a) suggest that a contextual perspective might point out sources of variability by investigating various moderating factors in decision situations that were not previously considered. He shows the merits of this perspective by investigating in detail whether thoughtful and non-thoughtful cognitive processes have equal consequences for anchored decisions. This more contextual approach suggested by Wegener et al. (2010a) is accepted to further the applicability of experimental research for real-world situations prone to anchoring effects. For the other contributions to this methodological discussion, we refer to Wegener et al. (2010a) and ensuing contributions by Russo (2010), Frederick et al. (2010), Epley and Gilovich (2010) and Wegener et al. (2010b). Similarly, Crusius et al. (2012) plead for more process orientation in research on biases and heuristics and thus outline potential consequences for experimental economics. Overall, while there is a rich literature on various theoretical aspects regarding anchoring, its theoretical ambivalence regarding the influence of monetary incentives and group decision-making precludes us from deriving distinct behavioral predictions.

This same issue applies to the broad literature on small groups in social and organizational psychology (Laughlin, 2011 provides a comprehensive review). In this field, the topic of bias reduction through group cooperation has been summed up by Kerr et al. (1996), as well as Kerr and Tindale (2004). While there are some instances reported in psychological studies where groups are less biased, such as the cases of the hindsight bias (Stahlberg et al., 1995) and the overconfidence bias (Sniezek and Henry, 1989), the overall heterogeneity of the results and experimental paradigms makes general predictions in terms of whether groups more effectively avoid behavioral biases difficult.

For the specific area of bias reduction in anchoring, studies in two domains report results on group behavior. Firstly, experiments on legal juries are regularly based on group decision-making. For instance, Hinsz and Indahl (1995) report that groups are equally biased by anchors provided to them during trials when deriving jury rulings. This is in line with the work on anchoring effects in individual legal judgment (see e.g. Englich et al., 2005; Englich and Soder, 2009). Secondly, Whyte and Sebenius (1997) find that for a non-incentivized negotiation task, groups acting as 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 equally biased.

Nonetheless, while these studies (Whyte and Sebenius, 1997 in particular) might provide some perspective on group behavior when confronted with external anchors, we argue that they hold limited relevance for economic group research for two distinct 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 for the particular case of anchoring, see Alevy et al., 2011). Thus, given that no unambiguous prediction on groups' biases can be deduced from the generalist discussions in psychology, we review the studies with individual players that are relevant for our anchor tasks and our primary additional feature, namely monetary incentives.

Tasks on factual knowledge, as in the seminal anchoring task by Tversky and Kahneman (1974), have been repeated numerous times and lead to robust and substantial anchoring effects, e.g. in the studies by Blankenship et al. (2008), McElroy and Dowd (2007), Epley and Gilovich (2005) and multiple others discussed in the review by Furnham and Boo (2011). The same holds true for probability estimations, as shown e.g. by Chapman and Johnson (1999) and Plous (1989) for the likelihood of specific US military interventions and the occurrence of a nuclear war. Price estimates and willingness to pay are covered in the studies by Sudgen et al. (2013), Adaval and Wyer (2011), Bateman et al. (2008), Critcher and Gilovich (2008), Nunes and Boatwright (2004), Simonson and Drolet (2004), Ariely et al. (2003) and Mussweiler et al. (2000) among others, with overall robust anchoring effects for individuals in varying experimental settings, even when goods are considered in which subjects have experience in consuming.

Regarding the role of monetary incentives as a behavioral influence, results are ambiguous. While an economic perspective on the positive effects of monetary incentives and performance is fairly straightforward (see e.g. Smith and Walker 1993), no consensus has evolved over its role for debiasing decisions in anchoring. Tversky and Kahnemann (1974), Wilson et al. (1996) and Epley and Gilovich (2005) offer prizes for the best decision and find no debiasing effects. In contrast, Wright and Anderson (1989), Simmons et al. (2010) and Meub et al. (2013) find subjects to be less biased when given monetary incentives under the condition that there is a realistic opportunity of achieving better solutions through increased cognitive effort.

The overall scope of anchoring effects is discussed by Van Exel et al. (2006), who summarize that anchoring effects are bound to increase when the solutions to tasks are ambiguous,

subjects are unfamiliar and personally uninvolved with the tasks and the anchors seem plausible and trustworthy. Thus, following psychological studies, anchoring on a purely individual level is characterized as “*exceptionally robust, pervasive and ubiquitous*” (Furnham and Boo, 2011, p. 41) for a broad variety of task domains and experimental parameters. Nevertheless, the empirical evidence for group performance and anchoring is limited, with only Whyte and Sebenius (1997) distinctly suggesting that groups tend to reproduce the individually biased decisions and psychological small group research presenting ambiguous results regarding biases and heuristics.

By contrast, economic small group research offers fairly clear predictions for group performance, yet provides no evidence in terms of biases and groups. In their reviews of the past ten years of economic group experiments, Charness and Sutter (2012) and Kugler et al. (2012) summarize that groups are overall more successful than individuals, bringing them closer to game-theoretical requirements for rational decision-making. In interactive settings, teamwork is shown to induce greater competitiveness. While individual decision-makers are self-restrained in pursuing selfish goals by altruistic norms, the establishing of in-groups renders selfish motives acceptable as a means of maximizing collective income. As groups foster competitiveness, they tend to expect more selfish behavior by competitors, which further drives non-altruistic decisions. Furthermore, it is shown that groups generally alleviate the cognitive limitations of individuals. Collaboration thus enables more rational decisions through the transfer of insight from cognitively superior individuals to the group. Group performance can then approach or in some cases even beat that of the group’s best member. The effectiveness of this mechanism crucially depends on the demonstrability of task solutions (Laughlin et al., 2002). Hence, groups consistently outperform individuals in intellectual tasks with a clear and demonstrable correct solution. The counterpart are judgmental tasks that have more ambiguous answers and no single correct answer that is easily demonstrable (Cox and Hayne 2006). Group performance depends on the respective task’s position on a continuum from intellectual to judgmental (Laughlin, 1980). Consequently, groups have the potential to mitigate individuals’ bounded rationality through the transfer of information and it can be assumed that groups circumvent anchored decisions through improved intra-group information availability. The positive effect of “more heads” on the overall cognitive performance might lead to the expectation that groups will be less biased by external anchors. Nevertheless, the active discussion of anchor values might as well foster the activation of anchor-consistent knowledge and even increase anchoring effects.

Accordingly, there are two contradictory notions to be derived from previous research. The bulk of psychological research on anchoring effects leads to the prediction that there are hardly any means of avoiding the bias, including monetary incentives and group decision-making. In contrast, following economic small group research, the cognitive superiority of groups would predict that groups successfully avoid external anchors. To account for these contradictory notions, we present an anchoring study comprising three different anchoring tasks that are compatible with previous experimental studies in psychology. 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 in economically relevant domains.

We find that groups are significantly less biased by the anchor for a factual knowledge task. For probability and price estimates, individual and group decisions are equally biased by the external anchor; accordingly, individual biases are perpetuated by group cooperation. It appears that a group's ability to reduce individual biases depends on the task characteristics. In the case of intellectual tasks that have a clearly defined correct solution, debiasing is effective. For rather judgmental tasks, groups approach the performance of average individuals. Overall, we suggest that groups are 'less behavioral' than individuals in certain domains, albeit not universally.

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

We keep in line with standard anchoring studies by implementing anchoring tasks from three domains that have been covered extensively for individuals, showing robust anchoring effects (Furnham and Boo, 2011). There are five questions for each domain, namely factual knowledge questions, probability estimates and price valuations. We digress from the classical anchoring paradigms by providing strong monetary incentives for unbiased behavior and implement team decisions.

There are two experiments: one featuring individual decisions and one with groups of three players. Groups are randomly matched at the beginning; they are required to find a unitary solution using the chat implemented in zTree (Fischbacher, 2007). In each period, groups and individuals have 90 seconds to enter their answers. Each experiment comprises two treatments: *calibration* and *anchor*. Thus, we follow the procedure outlined by Jacowitz and Kahneman (1995), which allows for quantifying the anchoring bias and enables statistical testing.

Calibration has participants simply enter their answer to the respective questions. The resulting answers are then used to derive the anchor values, whereby the low and the high anchors represent the 15th or the 85th percentile of the estimations' distribution, respectively.

In *anchor*, these values are shown to participants combined with the additional question of whether participants assume the respective answer to be higher or lower than this shown value, which corresponds to the seminal experimental paradigm for anchoring (Tversky and Kahnemann 1974). While the order of questions is held constant for both experiments, there are two strictly opposing orders of high and low anchors such that there are at least two questions for both anchor conditions in every domain.

It is to note that we do not test whether group or individual players give estimates that are closer to the anchors; rather, we investigate if and to what extent they deviate from the respective players' behavior in the absence of anchors. Consequently, the absolute anchor values have to be different for group and individual players, although anchors are the same in relative terms.

Questions of each domain are randomly ordered and subsequently assigned according to the position in the sequence, whereby we established five blocks, each comprising one question of every domain. Within these blocks, the domains are again randomly ordered.

Table 1 presents our 15 questions. The five factual knowledge questions featured topics related to the city of Goettingen and were chosen to ensure that 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

no.	Order	Domain	Question
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 Bafög (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.

Note that question specifics hold minor relevance in our setting, as we control for players' capabilities to answer the questions by running *calibration*. Consequently, the anchoring bias does not depend on question specific knowledge, given that it is measured conditional on the performance in *calibration*.

³ The pictures shown to participants are documented in appendix B. All respective brands were erased from the pictures to preclude subjects from being influenced by them. The items were sold on the internet, comparable to eBay auctions. Although smartphones were banned during the experiment, we additionally ensured that the correct answers to our questions were not easy to look up on the internet.

Following the 15 questions, we assess the 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 implement monetary incentives to establish that giving the most accurate estimates possible is the rational strategy. It was explained in the instructions that only five of the fifteen questions were rewarded monetarily, but it did not state to which of the 15 questions this applied. We chose the five factual knowledge questions as they have definite correct answers that are drawn from the University of Goettingen statistical office. Payoffs depended 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. Additionally in the group experiment, all participants received a basic payment of €2.5 to account for the increased duration of the sessions due to chat communication. Accordingly, the minimum payoff for individuals (groups) amounted to €5 (€7.5), while the maximum was €12.5 (€15.0). Understanding of the game and the payoff mechanism was ensured through control questions before the experiment.

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. The sessions lasted around 30 minutes for individuals and around 40 minutes for groups. We had 72 participants in the individual experiment and 64 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.⁴

⁴ We recorded demographic information, gender and fields of study of all participants in a questionnaire after the game, finding that there are no significant or systematic differences with respect to the distribution over experiments. 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

3. RESULTS

Before analyzing our results in depth, we first provide a descriptive overview of the anchoring bias in group and individual players. We therefore calculate the *anchor index* as described by Jacowitz and Kahnemann (1995). The anchor index 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.

An index value of 0 implies that there is no difference in median estimations between the low and high anchor condition, a value of 1 indicates a difference equal to the distance in the anchor values, while values greater than 1 correspond to a difference in medians greater than the distance of the respective anchors. The anchor index values are shown in Figure 1.

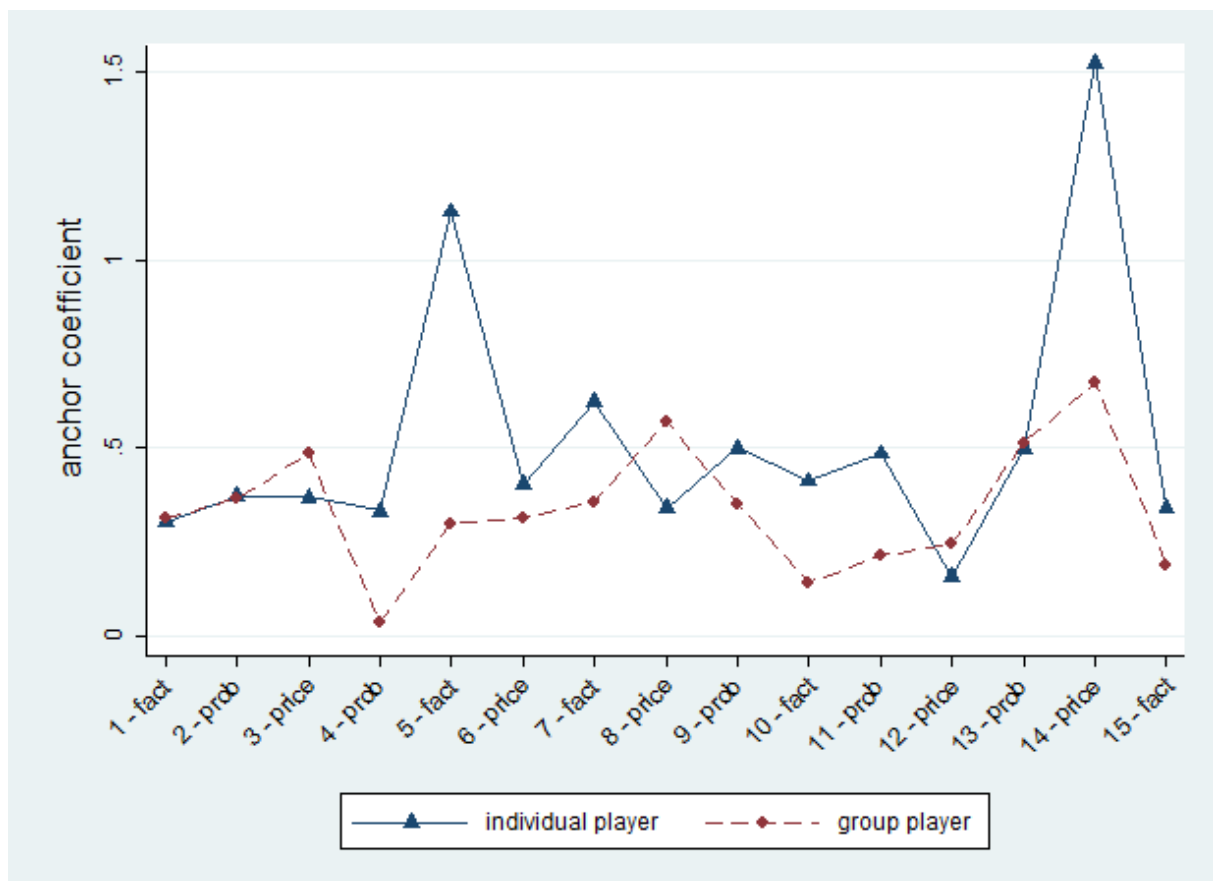


Figure 1: Anchor index for individual and group players. “fact” refers to factual knowledge questions, “prob” refers to probability- and “price” to price estimations.

(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 instructions were in German and are available from the authors upon request. A translation is provided in Appendix A.

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

While this descriptive analysis gives a first impression pointing to an anchoring bias for both individuals and groups, it is only based on the median predictions and might thus obliterate important differences. We therefore analyze our results in detail in the following subsections.

3.1 CALIBRATION TREATMENT AND ANCHOR DERIVATION

As mentioned above, we run a calibration treatment for both individual and group players. This enables us to derive 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 these calibration treatments. Jacowitz and Kahneman (1995) outline this method to allow for quantifying and actually testing the anchoring bias. Using this procedure, we can identify differences between individual and group players by analyzing their reaction to anchors conditional on their unbiased behavior as measured by the respective calibration treatment. Table 2 summarizes descriptive statistics for the calibration treatments.

The anchor values derived for group and individual players are quite similar. Question 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.⁵ Furthermore, the 85th percentile for groups in question 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 question 4. In factual knowledge questions, both tend to overestimate the correct answers, while there is no significant difference between individuals and groups in terms of estimation accuracy.⁶

⁵ E.g. for individuals players, the low anchor at question 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.

⁶ 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

Question	<i>Individuals (n=24)</i>					<i>Groups (n=24)</i>			
	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 2: Descriptive statistics for the calibration treatments⁷

As described, the distribution of estimations serves as a benchmark for behavior in the absence of anchors. We can thus range in responses by assigning point values to the estimations given in anchor treatments according to the corresponding distribution of the calibration treatments: 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 group's range of estimations, they transform to 0 or 100 points, respectively. This ordinal transformation procedure allows for pooling the results of varying questions. Thereby, point values smaller than 50 indicate a downward bias and values greater than 50 point to an upward bias.⁷

In order to comprise the bias in both conditions within one key figure, we define the *anchor effectiveness index*, by which we measure the average deviation from the calibration treatment's median. We therefore calculate deviations depending on the anchor condition according to the following procedure: in the high anchor condition, we subtract 50 from each

players, the mean is .912 and the median .595. This difference is not statistically significant (Wilcoxon rank-sum test, $z=.206$, $p=.8366$).

⁷ 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.

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 each other. 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. Consequently, we are able to make a straightforward comparison of individual and group players' susceptibility toward anchors.

Additionally, we 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 the calibration treatments, a share greater than 15% indicates that the anchoring bias moves some responses 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 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 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.

3.2 FACTUAL KNOWLEDGE

Table 3 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	
<i>Individuals</i>								
Q1	20	15 (37.5)	20 (62.5)	37.50 (20.23)	61.46 (26.01)	16.67 (0)	25 (0)	11.98
Q5	6206	4350 (36.11)	10000 (100)	41.90 (23.96)	86.81 (18.88)	4.17 (0)	62.5 (62.5)	22.45
Q7	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
Q10	18000	17500 (50)	21293 (83.33)	51.39 (29.41)	78.47 (20.55)	20.83 (0)	33.33 (16.67)	13.54
Q15	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
<i>Groups</i>								
Q1	18	13 (30)	18 (50)	38.42 (25.44)	46.56 (14.57)	15.79 (15.79)	0 (0)	4.71
Q5	7000	7000 (54.5)	8500 (72.23)	58.18 (23.94)	71.77 (19.61)	10 (0)	21.05 (15.79)	11.44
Q7	5000	4000 (52.63)	6500 (68.42)	46.26 (18.36)	66.32 (22.87)	0 (0)	20 (0)	10.19
Q10	18625	18000 (45)	19000 (60)	52.65 (25.93)	57.78 (25.04)	5.88 (5.88)	16.67 (0)	2.71
Q15	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 3: 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. If 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 question level in the low anchor condition, as indicated by median point values not smaller than 50.

Pooling the results for the five questions 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 $p=.0000$).

Nonetheless, there is a significant difference in point values between the high and low anchor condition (Wilcoxon signed-ranks test, $z=5.159$, $p=.0000$). Thus, the anchor values strongly bias individuals' estimates.

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 easily be seen by the higher deviation from the 50 points representing unbiased behavior.

It is quite obvious that groups are 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* supports the notion of group players being 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.⁸ 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 the same in the high and low anchor condition (Wilcoxon signed-ranks test, $z=.896$, $p=.3704$).

While there are more *outliers*, as defined above, in the low anchor condition for groups (Wilcoxon rank-sum test, $z=-1.848$, $p=.0645$), there are more 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,

⁸ 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.

the ordinal transformation rather weakens the apparent difference in the anchoring bias. In total, there are about twice as much *outlier* estimates for individual players.

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.

We argue that these 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 benign outcome if the anchor draws estimates toward the correct answers. In contrast, a tremendously negative effect on accuracy results when the anchor draws estimations in the wrong direction. Groups' reluctance to adjust estimates toward anchors renders their performance more robust.

Result 1: *Group cooperation reduces the anchoring bias for factual knowledge questions to an economically relevant extent.*

3.3 PROBABILITY ESTIMATES

Table 4 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	
<i>Individuals</i>								
Q2	52	45 (45.45)	71 (68.18)	45.45 (21.78)	67.80 (10.55)	8.33 (4.17)	4.17 (0)	11.18
Q4	98	97 (54.17)	100 (100)	49.83 (23.43)	73.61 (34.02)	8.33 (0)	58.33 (0)	11.89
Q9	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
Q11	75	60 (20.83)	77 (60.42)	32.81 (25.40)	55.73 (27.74)	33.33 (0)	20.83 (8.33)	11.46
Q13	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
<i>Groups</i>								
Q2	65	54 (45)	70 (60)	41.5 (27.53)	67.22 (28.81)	20 (20)	33.33 (22.22)	12.63
Q4	97	98 (62.5)	98 (62.5)	59.17 (18.37)	71.49 (25.92)	0 (0)	36.84 (0)	5.77
Q9	70	56 (29.17)	76 (62.5)	33.75 (23.33)	66.04 (22.71)	10 (0)	10 (5)	16.15
Q11	80	79 (54.17)	85 (79.17)	47.29 (29.75)	66.01 (29.14)	15 (5)	31.58 (0)	9.19
Q13	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 4: Descriptive statistics for probability estimates

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

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 $p=.0000$; for high anchors $p=.0000$). The point values for the high anchor condition significantly exceed the low ones (Wilcoxon signed-ranks test, $z=5.374$, $p=.0000$).

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 $p=.0000$) and significantly different transformed point values with respect to the anchor condition (Wilcoxon signed-ranks test, $z=4.362$, $p=.0000$).

For both anchor conditions, there is no difference between groups and individuals in deviations from the calibration group. 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 $z=-1.109$, $p=.2672$).

In all, 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.*

3.4 PRICE ESTIMATES

Table 5 presents the results for the five price estimations. For all questions, the median predictions 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.

Individual players' point values systematically deviate from 50 for both anchor conditions (Sign test, one-sided, for low anchors $p=.0129$; for high anchors $p=.0000$). Point values in the high anchor condition are again significantly greater than in the low one (Wilcoxon signed-ranks test, $z=5.641$, $p=.0000$).

For groups, we find the same pattern, given that point values are significantly different from 50 (Sign test, one-sided, for low anchors $p=.0000$; for high anchors $p=.0000$) and there is a difference in point values between the high and low anchor condition (Wilcoxon signed-ranks test, $z=5.162$, $p=.0000$).

	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	
<i>Individuals</i>								
Q3	155	150 (50)	304 (79.17)	49.13 (25.30)	77.26 (17.16)	8.33 (8.33)	12.5 (8.33)	14.06
Q6	120	100 (41.67)	150 (66.67)	39.41 (23.41)	64.76 (19.19)	25 (0)	12.5 (4.17)	12.67
Q8	40	35 (45.83)	50 (66.67)	44.10 (26.58)	69.97 (20.37)	29.17 (4.17)	29.17 (0)	12.93
Q12	184	168 (47.92)	199 (62.5)	49.65 (25.12)	59.20 (23.47)	16.67 (0)	8.33 (0)	4.77
Q14	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
<i>Groups</i>								
Q3	252	165 (29.17)	500 (79.17)	38.13 (15.72)	79.86 (10.23)	5 (0)	11.11 (5.56)	20.39
Q6	115	88 (35.42)	130 (66.67)	36.67 (19.38)	64.69 (19.81)	15 (5)	10.53 (0)	13.40
Q8	42	30 (25)	50 (75)	36.46 (24.78)	65.00 (20.43)	50 (0)	10 (0)	14.27
Q12	175	161 (45.83)	200 (70.83)	47.71 (26.88)	68.75 (21.01)	20 (0)	15 (0)	10.52
Q14	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 5: Descriptive statistics for price estimates

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 $z=.848$, $p=.3967$). These findings are consequently reflected by 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 estimation task.

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

3.5 PERCEPTION OF THE ANCHOR VALUES

As an additional feature to our design, we asked all participants in treatment groups to rate how relevant they perceived the anchor values, 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. 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, $z=-.358$, $p=.7207$). This gives us some 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.*

In sum, we cannot conclude that groups 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 or price estimations.

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 biases persistent on an individual level. This would in fact reduce the relevance of heuristics and biases in actual market contexts, given that the majority of decisions in economic domains are taken by groups. However, no study in experimental economics to date has made a direct comparison between individual and group performance concerning 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.

Although groups are biased by the anchor, they are in fact able to reduce anchoring in the domain of factual knowledge. In contrast, for probability estimation and price estimation, groups are equally biased by the external anchors as individual players. Thus, stating that groups are generally less affected by behavioral biases might be overly optimistic. 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 intellectual 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 intellectual reasoning, group performance tends to approach the level of average individuals. Apparently, in rather intellectual 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 or price estimates, 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.

ACKNOWLEDGEMENTS

We are grateful to Kilian Bizer for helpful comments and to Tonia Sauerwald for excellent student assistance in programming the experiment.

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 (2), 355-365.
- Alevy, J.E., Craig Landry, C.E., List, J., 2011. Field Experiments on Anchoring of Economic Valuations. University of Alaska Anchorage, Department of Economics Working Paper, 2011-02.
- Ariely, D., Loewenstein, G., Prelec, D., 2003. "Coherent arbitrariness": Stable demand curves without stable preferences. *Quarterly Journal of Economics* 118 (1), 73-105.
- 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 (2), 127-41.
- Bergman, O., Ellingsen, T., Johannesson, M., Svensson, C., 2010. Anchoring and cognitive ability. *Economics Letters* 107 (1), 66-68.
- 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 (6), 1465-1476.
- 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 (4), 437-449.
- Bucchianeri, G.W., Minson, J., 2013. A homeowner's dilemma: Anchoring in residential real estate transactions. *Journal of Economic Behavior & Organization* 89 (6), 76-92.
- 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 (2), 369-390.

- 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.
- 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 (1), 47-76.
- Chapman, G.B., Johnson, E.J., 1999. Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes* 79 (2), 115-153.
- Charness, G., Sutter, M., 2012. Groups make better Self-Interested Decisions. *Journal of Economic Perspectives* 26 (3), 157-176.
- Cox, J., Hayne, S., 2006. Barking up the right tree: are small groups rational agents?. *Experimental Economics* 9 (3), 209-222.
- Critcher, C.R., Gilovich, T., 2008. Incidental environmental anchors. *Journal of Behavioral Decision Making* 21 (3), 241-251.
- Critcher, C.R., Rosenzweig, E.L., Gilovich, T., 2014. Beyond Anchoring: Contextual Influence on Adjustment. Unpublished Manuscript.
<http://claytoncritcher.squarespace.com/papers/>, last access: 03/07/14.
- Crusius, J., van Horen, F., Mussweiler, T., 2012. Why process matters: a social cognition perspective on economic behavior. *Journal of Economic Psychology* 33 (3), 677-685.
- Englich, B., Mussweiler, T., Strack, F., 2005. The last word in court: a hidden disadvantage for the defense. *Law and Human Behavior* 29 (6), 705-722.
- Englich, B., Soder, K., 2009. Moody experts: how mood and expertise influence judgmental anchoring. *Judgment and Decision Making* 4 (1), 41-50.
- 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 (3), 199-212.
- Epley, N., Gilovich, T., 2010. Anchoring unbound. *Journal of Consumer Psychology* 20 (1), 20-24.
- Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2), 171-178.

- Frederick, S., Kahneman, D., Mochon, D., 2010. Elaborating a simpler theory of anchoring. *Journal of Consumer Psychology* 20 (1), 17-19.
- Frederick, S.W., Mochon, D., 2012. A scale distortion theory of anchoring. *Journal of Experimental Psychology: General* 141 (1), 124-133.
- Fudenberg, D., Levine, D.K., Maniadis, Z., 2012. On the robustness of anchoring effects in WTP and WTA experiments. *American Economic Journal: Microeconomics* 4 (2), 131-145.
- 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.
- Furnham, A., Boo, H.C., 2011. A literature review of the anchoring effect. *The Journal of Socio-Economics* 40 (1), 35-42.
- Greiner, B., 2004. An online recruitment system for economic experiments. *GWDG Berichte* 63, 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.
- 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.
- Jacowitz, K.E., Kahneman, D., 1995. Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin* 21 (11), 1161-1166.
- 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 (4), 410-434.
- Kerr, N.L., MacCoun, R.J., Kramer, G.P., 1996. Bias in judgment: Comparing individuals and groups. *Psychological Review* 103 (4), 687-719.
- Kerr, N.L., Tindale, R.S., 2004. Group performance and decision making. *Annual Review of Psychology* 55, 623-655.
- 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.

- Laughlin, P.R., 1980. Social combination processes of cooperative problem-solving groups on verbal intellectual tasks, in: Fishbein, M. (Ed.), *Progress in Social Psychology*. Hillsdale NJ: Erlbaum, 127-155.
- Laughlin, P.R., 2011. *Group Problem Solving*. Princeton and Oxford: Princeton University Press.
- 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.
- 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.
- List, J.A., 2003. Does Market Experience Eliminate Market Anomalies?. *The Quarterly Journal of Economics* 118, 41-71.
- List, J.A., 2004a. Testing neoclassical competitive theory in multilateral decentralized markets. *Journal of Political Economy* 112 (5), 1131-1156.
- List, J.A., 2004b. Neoclassical theory versus prospect theory: Evidence from the marketplace. *Econometrica* 72 (2), 615-625.
- List, J.A., 2005. The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. National Bureau of Economic Research, Working Paper No. 11616.
- 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 (1), 1-53.
- 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.
- 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.
- 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., Bizer, K., 2013. Anchoring: a valid explanation for biased forecasts when rational predictions are easily accessible and well incentivized? CEGE Discussion Paper No. 166.
- Mussweiler, T., Strack, F., Pfeiffer, T., 2000. Overcoming the inevitable anchoring effect: considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin* 26 (9), 1142-1150.
- Nunes, J.C., Boatwright, P., 2004. Incidental prices and their effect on willingness to pay. *Journal of Marketing Research* 41 (4), 457-466.
- Plous, S., 1989. Thinking the unthinkable: the effects of anchoring on likelihood estimates of nuclear war. *Journal of Applied Social Psychology* 19 (1), 67-91.
- Russo, J.E., 2010. Understanding the effect of a numerical anchor. *Journal of Consumer Psychology* 20 (1), 25-27.
- 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 (6), 917-932.
- Simonson, I., Drolet, A., 2004. Anchoring Effects on Consumers' Willingness-to-Pay and Willingness-to-Accept. *Journal of Consumer Research* 31 (3), 681-90.
- Smith, V.L., Walker, J., 1993. Monetary rewards and decision cost in experimental economics. *Economic Inquiry* 31 (2), 245-261.
- Snizek, J.A., Henry, R.A., 1989. Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes* 43 (1), 1-28.
- 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.
- Sugden, R., Zheng, J., Zizzo, D.J., 2013. Not all anchors are created equal. *Journal of Economic Psychology* 39, 21-31.
- Tufano, F., 2010. Are 'True' Preferences Revealed in Repeated Markets? An Experimental Demonstration of Context-Dependent Valuations. *Experimental Economics* 13 (1), 1-13.
- Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. *Science* 185 (4157), 1124-1131.

- Van Exel, N., Brouwer, W., van den Berg, B., Koopmanschap, M., 2006. With a little help from an anchor. Discussion and evidence of anchoring effects in contingent valuation. *Journal of Socio-Economics* 35 (5), 836-853.
- 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 (1), 5-6.
- 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.
- 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.
- 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 (4), 387-402.
- 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 (1), 68-82.

APPENDIX A

Instructions: The differing information for the group experiment is 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 round, 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 <i>{best group-answer}</i>	25 ECU
Forth to sixth best answer <i>{second best group-answer}</i>	20 ECU
Seventh to ninth best answer <i>{third best group-answer}</i>	15 ECU
All other answers <i>{group-answer}</i>	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 estimates.

i) question no. 3



ii) question no. 6



iii) question no. 8



iv) question no. 12



v) question no. 14



Chapter V

OVERCONFIDENCE AS A SOCIAL BIAS: EXPERIMENTAL EVIDENCE

with Till Proeger

Published:

Economics Letters, 122(2), 203-207.

DOI: <http://dx.doi.org/10.1016/j.econlet.2013.11.027>

ABSTRACT

The overconfidence bias is discussed extensively in economic studies, yet fails to hold experimentally once monetary incentives and feedback are implemented. We consider overconfidence as a social bias. For a simple real effort task, we show that, individually, economic conditions effectively prevent overconfidence. By contrast, the introduction of a very basic, purely observational social setting fosters overconfident self-assessments. Additionally, observing others' actions effectively eliminates underconfidence compared to the individual setting.

Keywords: laboratory experiment; overconfidence bias; real effort; self-assessment; social interaction

JEL Classification: C91; D03; D83

HIGHLIGHTS

- We compare overconfidence for an individual and a social, observational experiment.
- We find realistic confidence levels in the individual setting on average.
- Introducing the observation of other subjects considerably increases overconfidence.
- The social setting reduces underconfidence compared to the individual setting.
- Our results suggest that overconfidence is a social rather than an individual bias.

Chapter VI

STRATEGIC COORDINATION IN FORECASTING: AN EXPERIMENTAL STUDY

with Kilian Bizer, Till Proeger and Markus Spiwoks

Published:

Cege Discussion Papers, no. 195, Göttingen University, March 2014.

Download: <http://www.uni-goettingen.de/de/60920.html>

ABSTRACT

Reputational herding has been considered as a driving force behind economic and financial forecasts clustered around consensus values. Strategic coordination can consequently explain poor performances of prediction markets as resulting from the distinct incentives that forecasters face. While this notion has been considered theoretically and empirically, the underlying behavioral working mechanisms have not yet been described. We thus put forth an exploratory experiment on the emergence and robustness of coordination in a forecasting setting implementing contradictory incentives for accurate forecasts and coordination. Forecasts are shown to be inaccurate and biased toward current values. This in turn has subjects aiming at coordination benefits. Predominantly, coordination is achieved through the risk-dominant equilibrium as the game proceeds. Once established, coordination is fairly stable and adds to overall welfare. Our results support the assumption of rational herding as a driving force for predictions of poor accuracy that are systematically biased towards focal points.

Keywords: coordination; incentives; laboratory experiment; reputational herding; sunspot equilibrium

JEL classification: C90; D03; D83; G17

HIGHLIGHTS

- We present a novel experimental approach to strategic coordination in forecasting.
- We implement contradictory incentives for prediction accuracy and herding.
- While accuracy is weak, coordination evolves and proves fairly stable.
- Coordination is predominantly achieved through the risk dominant equilibrium.
- Our results add to theories of biased forecasts due to reputational herding.

1. INTRODUCTION

Economic actors regularly coordinate their decisions in a broad range of settings, within which a common course of action yields higher payoffs. A widely discussed area of coordinated actions can be found on markets for economic and financial forecasting. It has been convincingly argued that prediction markets are particularly prone to rational herding, which manifests in a bias towards consensus forecasts in individual predictions. By conforming to the majority of other agents, individuals benefit from concealing their inability to provide accurate forecasts. Building on John Maynard Keynes' writings on financial markets, Scharfstein and Stein (1990) suggest in their seminal paper that forecasters generate a "sharing-the-blame" effect by strategically coordinating their predictions. Accordingly, analysts rationally decide against the truthful reporting of private information, rather adapting to a common consensus to be perceived as being more capable. In this way, the blame for false predictions can be avoided by pointing to the communities' overall inability of correctly anticipating actual developments.

The systematic, strategic overweighing of publicly available information has subsequently been modelled theoretically and investigated empirically in a considerable number of studies. Most recently, Ottaviani and Sorensen (2006b) have formalized the conception in a *theory of reputational cheap talk*, whereby forecasters aim at improving their reputation by pretending to principals to be well informed. This leads to a strong bias towards the overall consensus forecasts and consequently to inaccurate predictions due to the strategic non-disclosure of private information, the ultimate result being that "*paradoxically, the desire of analysts to be perceived as good forecasters turns them into poor forecasters*" (Ottaviani and Sorensen, 2006b, p.443). This concept has been tested empirically, yielding ambiguous results. Most prominently, Hong et al. (2000) suggest that particularly less experienced analysts seek safety in consensus forecasts, while similarly, Lamont (2002) and Clement and Tse (2005) find forecasts to be clustered around consensus values. By contrast, building upon models of anti-herding, studies such as those by Batchelor and Dua (1990), Zitzewitz (2001a; b) and Bernhardt et al. (2006) present evidence in favor of strategic differentiation in prediction markets.

While these results are contradictory, a compromising interpretation has been put forth by Marinovic et al. (2010), who argue that the lack of strong competition and ex-post evaluations of forecasting accuracy leads prediction markets to quickly deteriorate to pure beauty contests. Thereby, herding on consensus values becomes predominant as analysts are

primarily rewarded for their ability to anticipate the expectations and choices of all other participants. Forecasters then cease their predictive efforts and resort to merely guessing the actions of others, thus striving for coordination. Therefore, the distinct structure of incentives might explain why herding on consensus values is more or less pronounced in prediction markets for different economic or financial key figures.

Under these conditions, forecasting that maximizes individual reputation can be interpreted as a coordination problem in which players minimize deviations from the average prediction by agreeing upon a common focal point. While this notion has been discussed extensively in previous studies, no behavioral study to date has explored strategic coordination in a forecasting setting. Studying strategically biased forecasts experimentally could provide a useful empirical insight into the coordination patterns of independent agents in a forecasting environment. Given that underlying mechanisms of coordination are inaccessible from a purely theoretical perspective or solely through an evaluation of time series data, a controlled laboratory environment might provide additional evidence. Consequently, we run an exploratory experimental study to provide behavioral insight into analysts' coordinative behavior in a forecasting setting.

We build our investigation on two distinct strands of literature in experimental economics. Firstly, we use recent experiments on forecasting as our working horse, whereby payoff-maximization demands that subjects correctly interpret fundamentals and make accurate predictions. These settings (see e.g. Becker et al., 2008) primarily focus on showing subjects' inability to rationally process graphical or statistical information, which is interpreted as a refutation of the rational expectation hypothesis. Secondly, we implement a coordination mechanism as an alternative mode of generating payoffs. This approach follows a large body of experimental studies that consider rationality in coordination games. Based on Duffy and Fisher's (2005) seminal paper, recent studies have considered the effectiveness of random focal points, so-called sunspots¹, in fostering coordination between subjects. Connecting these two experimental approaches, we are able to study if and how subjects coordinate in forecasting games.

While the basic structure of our setting somewhat resembles the experimental approach to reputational herding in Bedke et al. (2009), it differs by implementing an actual forecasting

¹ Duffy and Fisher (2005) explain the term "sunspot" as originating from William S. Jevons, who assumed solar activity to be a considerable influence on economic activity. In the modern meaning, it is referred to as a random variable that is unrelated to the fundamentals yet nevertheless fosters coordination.

task and an endogenous coordination mechanism. Subjects are asked to perform judgmental forecasting of future values based on the graphical and numerical depiction of several determinants. Due to the complexity and interaction of the underlying fundamentals determining the values to be forecast, accurate predictions – albeit the first best solution – are effectively hard to achieve. Frustration with the task might prompt subjects to increasingly consider the alternative strategy, i.e. coordination on a common focal point, which also leads to a payoff in each period. While theoretically all values within the range of possible forecasts could be used to coordinate, we implement a unique payoff and risk dominant equilibrium at the respective rounds' current value. Uniqueness of this particular efficient coordination regime is achieved by implementing costs for forecasts that deviate from the focal point. In societies of eight, all players are also randomly matched to groups of two, who are able to communicate via chat. All fundamentals, prior predictions and earnings of other players are common knowledge.

Within this framework, we investigate whether subjects succeed in agreeing on focal points, even when the task of forecasting future values is conflicting to coordination. We can thus explore four particular aspects. Firstly, we aim at analyzing forecasting accuracy and the predictive quality of the rational expectation hypothesis. Secondly, we assess whether subjects in a forecasting setting succeed in coordinating on external focal points. Thirdly, we investigate the robustness of coordination once it is established and finally, we discuss the welfare implications of our results.

Overall, we find that subjects fail to make forecasts in line with the rational expectation hypothesis. Accordingly, forecast accuracy is poor, even when subjects do not aim at coordination. However, coordination on the sunspot equilibrium constituted by current values increases strongly over periods and is fairly successful overall. The coordination is found to be efficient and quite robust, i.e. once subjects achieve coordination on the sunspot at minimal costs, they tend not to deviate from this strategy. Along with dismal performance in forecasting yet increasing cooperation, subjects' effort levels and confidence in the accuracy of their predictions decreases. It can thus be assumed that frustration with the forecasting task fosters coordination. While total welfare is considerably lower than the maximum, for a number of subjects it approaches the benchmark representing efficient coordination as the game proceeds.

The remainder of this paper is organized as follows. Section 2 reviews the literature related to our study. Section 3 presents the game and our experimental design. The results are presented in section 4, before section 5 offers a conclusion.

2. RELATED LITERATURE

The forecasting behavior of financial and economic analysts has been discussed extensively from various scientific angles.² While the sheer volume of studies precludes a comprehensive review, broad trends upon which our study builds can be outlined.

Starting in the early-1990s, two closely related strands of literature have evolved that apply the concept of analyst's rational herding. Primarily, based on the seminal contributions by Welch (1992), Bikhchandani et al. (1992) and Banerjee (1992), the concept of informational cascades has become widely discussed. These authors implement a sequential decision-making regime in which the disregarding of private information in the face of potentially conflicting public information becomes rational in certain situations. Bayesian updating has agents rationally interpreting the weights of private and public information and deciding accordingly. Thus, when aggregate public information becomes overwhelming, a cascade situation develops and the single private information is rationally discarded by all following agents in the decision sequence (Graham, 1999). The private information of these players is not revealed by their decisions and information aggregation becomes insufficient. Starting with the seminal contribution by Anderson and Holt (1997), experimental studies have documented subjects' failure to rationally make use of private and public information. Thereby, subjects display a tendency to cling to private information that systematically contradicts rational behavior derived from perfect Bayesian updating (Weizsäcker, 2010)

² A generalist theoretical perspective on economic aspects of forecasting is given in Elliot and Timmermann's (2008) literature review. The authors encompass the large number of empirical and theoretical studies within several discussions on forecasting. More specifically and closer to our investigation is Devenov and Welch (1996), who offer an overview of theoretical explanations as to why herding among agents in financial markets might occur. Individual forecasting behavior is a prominent topic in experimental psychology, particularly from the perspective of heuristics and biases, leading to non-optimal results in many forecasting settings. Harvey (2007) offers a comprehensive introduction to the literature concerned with individual forecasting behavior. More generally, Assad (2012) gives an overview of the literature on experimental finance and includes aspects of forecasting. Stekler (2007) differentiates between various components regarding the process of macroeconomic forecasting and reviews the respective studies. Ramnath et al. (2008) provide an applied institutional perspective, presenting a taxonomy on the role of financial analysts in capital markets. Similarly, v. Campenhout and Verhestraeten (2010) more specifically review contributions on herding behavior among financial analysts. Another perspective of applied forecasting regarding corporate analysts is given by Lawrence et al. (2006), who review the central research of the past 25 years in judgmental forecasting.

offers a meta-study). Rational herding thus occurs less often than it is predicted by theory due to subject's reluctance to discard private information.

Reputational herding is closely related to discussions of herding evoked by informational cascades. However, an additional dimension of incentives is considered, which rewards imitating the choices of previous agents. Consequently, agents gain positive reputational externalities from the fact that their decisions resemble those of a reference group. Ottaviani and Sorensen (2000) discuss the similarities between both concepts and propose that an agency-based reputational model might yield superior explanatory power for a wide range of economic contexts. The seminal model for reputational herding goes back to Scharfstein and Stein (1990), who formalize Keynes' remarks on conformity in financial markets.³ Other influential models of reputational herding have been contributed by Hirshleifer (1993), Trueman (1994), Prendergast and Stole (1996) and Holmstrom (1999), who point at different sources, dimensions and consequences of reputational pressure. Froot and Scharfstein (1992) more closely discuss the beauty contest-like structure of reputational concern, which evolves due to short time horizons. Most recently, Bar-Isaac (2012) suggests a model on the interplay of career concerns, transparency and the resulting incentive for the acquisition of expertise.

Two contradictory basic notions can be identified within the theoretical discussion on the parameters of reputational herding. The first such notion emphasizes individual gains from conforming to a publicly observable consensus value. Accordingly, conformity and coordination are seen as the dominant strategy in forecasting situations. By contrast, proponents of strategic differentiation, i.e. anti-herding, point to potential reputational advantages of making forecasts that significantly deviate from consensus values. By appearing to be bold, analysts can gain competitive advantages, signaling to have access to superior private information.

Supporting the first notion, Ehrbeck and Waldmann (1996) present a model favoring the reputational gains of biased forecasts, within which agents aim at mimicking the most capable

³ The original, widely quoted citation being: *"it is the long-term investor, he who most promotes the public interest, who will in practice come in for most criticism, wherever investment funds are managed by committees or boards or banks. For it is in the essence of his behavior that he should be eccentric, unconventional, and rash in the eyes of average opinion. If he is successful, that will only confirm the general belief in his rashness; and if in the short-run he is unsuccessful, which is very likely, he will not receive much mercy. Worldly wisdom teaches that it is better for reputation to fail conventionally than to succeed unconventionally."* (Keynes 1936, p. 157-158)

forecasters, instead of merely minimizing prediction errors. However, their assumptions of a strategic bias towards common values fail to hold empirically. Accordingly, they suggest a behavioral explanation of biased forecasts. In three related models, Ottaviani and Sorensen (2006a;b;c) argue that reputational concerns regularly fail to create appropriate incentives for the truthful reporting of private information. Thus, the role of strong incentives to herd on the mean forecasts is emphasized, particularly for the case when markets are unaware of analysts' strategic herding. In turn, if the market knows about these strategic incentives, forecasts are made honestly and necessarily deviate from the consensus equilibrium due to the full revelation of private information. Moreover, if markets for forecasts include a competition for accuracy with pre-specified rules in a winner-takes-all setting, agents' predictions become fully differentiated. Ottaviani and Sorensen (2006c) more closely discuss the conditions that preclude the realization of equilibria for truth telling. In these equilibria, relative reputational concerns that would otherwise lead to a differentiation of forecasts have no impact and competition fails to alter the resulting equilibrium.

In contrast to the models predicting a strong overall bias towards common information, numerous reputational models suggest anti-herding in forecasting. Most prominently, Trueman (1994), Zwiebel (1995), Prendergast and Stole (1996), Avery and Chevalier (1999), as well as Zitzewitz (2001b) show how reputational competition might effectively lead to differentiated forecasts. Laster et al. (1999) present a model with common information of all forecasters and thus a common expectation of future values. However, due to the structure of the demand side, forecasts cover the future value's probability distribution. Furthermore, forecasters more dependent on publicity will deviate more strongly from consensus values. Effinger and Polborn (2001) similarly point to positive reputational effects of not following predecessors, while Laux and Probst (2004) again show heterogeneity in forecasts accruing from strategic behavior. However, this only holds when past performance is perfectly evaluated by the markets. Otherwise, analysts cease their effort to become better informed and revert to herding. Kim and Zapatero (2009) argue that relative performance evaluation is the primary driver of analysts' herding. They further propose that herding is most likely to occur when the market comprises strong penalties for underperforming analysts.

Theoretical models have effectively mapped the various potential influences of reputational concerns on forecasters, with both herding and anti-herding as dominant strategies, the according equilibria crucially depending on the market structure faced by analysts.

Complementary to the theoretical work, empirical studies have investigated forecasting time series, discussing whether and under what conditions herding around common public information or signals, most prominently consensus values⁴, occurs in forecasting. Graham (1999) tests a sequential model of analysts' predictions, finding that herding behavior by contributors to investment newsletters depends on their former reputation. In laboratory studies, Cote and Sanders (1997), as well as Ackert et al. (2008) find empirical evidence for Trueman's (1994) model, as subjects with weak private information herd on the predictions made by others who are endowed with stronger private information. Welch (2000) provides empirical support for consensus herding consistent with models in which analysts coordinate based upon common information. Gallo et al. (2002), Bofinger and Schmidt (2003), Spiwoks (2004), Spiwoks et al. (2005; 2008; 2010), Spiwoks and Hein (2007), Ager et al. (2009), Jegadeesh and Kim (2009), as well as Gubaydullina et al. (2011) also find empirical evidence for analysts herding around consensus values. Bizer et al. (2013) in a recent classroom experiment with inexperienced participants also find a pattern of forecasts strongly biased towards current values of the respective time series. While Clement and Tse (2005) strengthen this notion, they point to the simultaneous occurrence of bold forecasts incorporating analysts' private information. Consequently, these bold forecasts have more informational value than those resulting from herding, although such forecasts only constitute a small fraction of their data. Batchelor (2007) adds to these results by showing an overall bias towards the consensus and a resulting weak forecasting accuracy overall. However, some forecasters consistently aim at signaling ability by making distinctively optimistic or pessimistic predictions in comparison to the consensus.

A similarly large body of literature has evolved aiming at supporting the theoretical perspective of strategic differentiation in forecasting. Batchelor and Dua (1990) show that analysts systematically deviate from average forecasts in an optimistic manner, suggesting that differentiation is responsible. In a follow-up study, Batchelor and Dua (1992) find forecasters to be conservative in the sense that they cling to their past predictions. However, they are shown not to be consensus-seeking. In two separate studies, Zitzewitz (2001a; b) reports equity analysts' earnings forecasts and thereby presents strong evidence of anti-herding in form of an exaggeration of differences with respect to the consensus. Furthermore,

⁴ Zarnowitz and Lambros (1987) provide an extensive discussion on potential definitions of consensus and uncertainty in economic forecasting. Gregory et al. (2001) provide a more recent contribution concerning how a consensus should be defined, as well as providing additional empirical insights.

implementing the notion that forecasters can extract valuable information from consensus forecasts, Bernhardt et al. (2006) as well as Chen and Jiang (2006) present evidence in favor of anti-herding in forecasting. Most recently, Doern and Weisser (2011) investigate macroeconomic forecasts and find a strong dispersion of forecasts, leading them to conclude that predictions overall are informational efficient and unbiased.

Another strand of literature approaches the role of reputation in forecasting by analyzing which analysts consistently stick to consensus values or make deviating predictions on a personal level. Stickel (1992) shows that analysts with a high reputation who made it on an “all star” list tend to rely less on consensus values. Cooper et al. (2001) support this point of view, showing that earnings forecasters herd by following superior lead analysts. As superior analysts publish their forecasts earlier, timeliness is suggested as a good measure of performance. Consequently, herding on consensus values is not universal, but rather depends on the expertise of individual forecasters. Analyzing individual career paths, Hong and Kubik (2003) show that accurate, as well as generally optimistic forecasters are more likely to have favorable careers. Clarke and Subramanian (2006) investigate the connection of employment risk and deviations from consensus values, showing that analysts with both very high and very low employment risk are most likely to deviate from consensus values. Leone and Wu (2007) measure the validity of analyst rankings and argue that they correctly represent their performance in terms of accuracy. Another prominent discussion deals with experience of analysts and their tendency to herd on consensus values. Chevalier and Ellison (1999), Hong, et al. (2000), Lamont (2002), and Clement and Tse (2005) report results consistent with theoretical work by Avery and Chevalier (1999) and Holmstrom (1999), thus indicating that older analysts herd less compared to their young counterparts. By contrast, Prendergast and Stole (1996) and Graham (1999) theoretically show that older analysts and managers should herd more. These assumptions are supported empirically by Stark (1997), Graham (1999), Li (2002) and Boyson (2010) who suggest that increased experience does not lead to more extreme deviations from consensus values.

While there are numerous theoretical and empirical contributions on reputational herding in prediction markets, to the best of our knowledge no experimental study to date has attempted to reconcile forecasting and strategic coordination in a single experimental setting. We thus review the respective experimental research.

Primarily, there is a large number of empirical studies on forecasting from the perspective of experimental economics, operational research and experimental psychology, with Leitner and

Wildburger (2011) providing a comprehensive review. While several aspects are considered in these fields, reconnecting the findings to the more general discussions on consensus-seeking behavior in economic forecasting remains difficult. We argue that the most accessible finding in this regard is the ubiquitous rejection of the rational expectation hypothesis. Unsurprisingly, this finding is replicated in a number of experimental studies using graphical depiction of indicators that constitute the future value to be predicted.⁵ Thus, subjects' inability to adequately process fundamentals and reconstruct the underlying models can be interpreted as asserting the notion of reputational herding by incapable forecasters.

Secondarily, with regard to coordination behavior, the experimental studies on “sunspot” equilibria (SSE) are closest to our study and can be used to derive our behavioral predictions. The first experiment testing the rationality of subjects in SSE-coordination games was put forth by Marimon et al. (1993). In a simple market setting, colored squares serve as a coordination mechanism, yet no coordination on SSE resulted. In the seminal contribution on SSE, Duffy and Fisher (2005) are the first to show that extrinsic, irrelevant information fosters coordination. In their setting, salient sunspots without connection to the fundamentals influence subjects' decisions and foster coordination. Crawford et al. (2008) show that sunspots' effectiveness is limited when they are not payoff dominant. Similarly, Bosch and Vried (2013) report that the impact of focal points strongly depends on the respective payoff and time that subjects have to consider coordination choices. This result is discussed by Agranov and Schotter (2012), who find that detrimental effects of payoff asymmetries can be neutralized by implementing communication among subjects. Alternatively, Roos et al. (2010) consider SSE as a reason for coordination failure, which might lead to welfare losses, as subjects are dragged away from the payoff-dominant equilibrium by the sunspot. Fehr et al. (2011) show that the efficiency of sunspot coordination depends crucially on the precision of public information, private signals and their respective correlation. Accordingly, extrinsic information might also lead to miscoordination, which is interpreted as evidence for the considerable fragility of sunspot-coordination. Further applications of this concept are provided by Arifovic et al. (2012), who show sunspot coordination in macroeconomic forecasting, as well as Shurchkow (2013), with coordination framed as speculation attacks. Finally, Bardsley et al. (2009) comprehensively discuss the working mechanisms of coordination by independent subjects, yet find contradictory evidence for the two potential

⁵ See Becker et al. (2005; 2007; 2008; 2009) for different variants of this experimental design. While parameters are changed, the basic conclusion of non-optimal forecasts, as measured by the benchmark of rational expectations, is replicated.

explanations of team reasoning and cognitive hierarchy. Consequently, the exact behavioral mechanisms for successful coordination have not been determined to date. In a closely related experimental strand of literature, external recommendations for coordinated play are considered. Duffy et al. (2013) provide an overview, concluding that experiments so far have shown that recommendations to play a particular equilibrium are only followed imperfectly and particularly less if the recommended equilibrium is payoff-dominated by some other equilibrium of the game. We interpret these findings as pointing to the fragility of coordination on focal equilibria, given that even experimenter-given advice is only followed irregularly. Therefore, coordination on endogenously generated equilibria in forecasting might not emerge robustly.

A third strand of literature from experimental economics relevant to our study is the superior performance of groups in achieving coordination. While teamwork itself has been shown to increase overall rationality across experimental studies (see e.g. Kugler et al., 2012), coordination in particular is similarly facilitated by the cooperation of players; for instance, as in Feri et al.'s (2010) study of coordination games. These findings hold relevance in our setting, which comprises both teamwork and coordination, suggesting that coordination might be facilitated by a two-player team setting, as implemented in this study.

3. EXPERIMENTAL DESIGN AND PROCEDURE

We implement a forecasting task that is built on the experimental design by Becker et al. (2007). Additionally we introduce incentives for coordination, extending the basic setting, which primarily encourages accurate predictions. The coordination element is based on the experimental literature on sunspot equilibria (Duffy and Fisher, 2005). Under fully rational behavior, the welfare maximizing outcome is defined by all subjects making optimal forecasts and thereby automatically coordinating on the expected future values. However, this is an unlikely outcome given the high complexity of the forecasting task. Ultimately, the decision problem for all subjects comes down to choosing between the strategies of making accurate forecasts or attempting to coordinate on a common focal point. In the following, we present the specifics of the game to clarify the dual game structure.⁶

3.1 GAME SETUP

Let us first consider the forecasting task, before explaining the coordination aspect of the game and finally describing the respective payoff regime.

Forecasting task

Participants are asked to make 21 forecasts on the development of a time series x_t , which are the realizations of the simple linear equation:

$$x_t = \theta_1 a_t + \theta_2 b_t + \theta_3 c_t + \theta_4 d_t + \varepsilon_t$$

with $\theta_1 = 11$, $\theta_2 = -6$, $\theta_3 = -4$, $\theta_4 = -0.3$ and ε_t being a normally distributed random variable with $\varepsilon_t \sim N(0,15) \forall t \in T$. These weighting factors and the realizations of the random variable are unknown to participants. Weightings are chosen such that values for a_t drive x_t positively to a similar extent as b_t and c_t combined negatively. The influence of d_t is rather small, which simplifies the task, once this has been recognized by subjects. However, participants receive values for a_t, b_t, c_t, d_t in every period, henceforth denominated as determinants. All values are rounded integers. Expected future values can subsequently be derived without taking ε_t into account, given that:

$$E(x_t) = E(\theta_1 a_t + \theta_2 b_t + \theta_3 c_t + \theta_4 d_t) + E(\varepsilon_t) = \theta_1 a_t + \theta_2 b_t + \theta_3 c_t + \theta_4 d_t$$

⁶ Instructions including a screenshot of the game play are attached in appendix A and B. Original instructions were in German and can be obtained from the authors upon request.

Overall, there are $T = 30$ periods. The forecast horizon h is five periods. Let \tilde{x}_t denote the forecast at period t , which aims at x_{t+h} . This means in period $t = 5$, subjects make a forecast $\tilde{x}_{t=5}$ for $x_{t=10}$ and so on. Subjects observe determinant values and the values of x_t of the first five periods before making forecasts. Accordingly, the first forecast is given in period 5, and the last one in period 25. Thus, while subjects did not know the correct formula, they had a certain insight into the past development of the time series and could reflect upon its derivation.

Additionally, the correct value x_t is shown in each period t ; this has subjects wait five periods to assess the accuracy of their predictions. Consequently, the first feedback is given in period 10.

Figure 1 summarizes the determinant values, while Figure 2 shows the resulting time series of x_t and respective expected values for $x_t (= x_t - \varepsilon_t)$ over the course of the game.

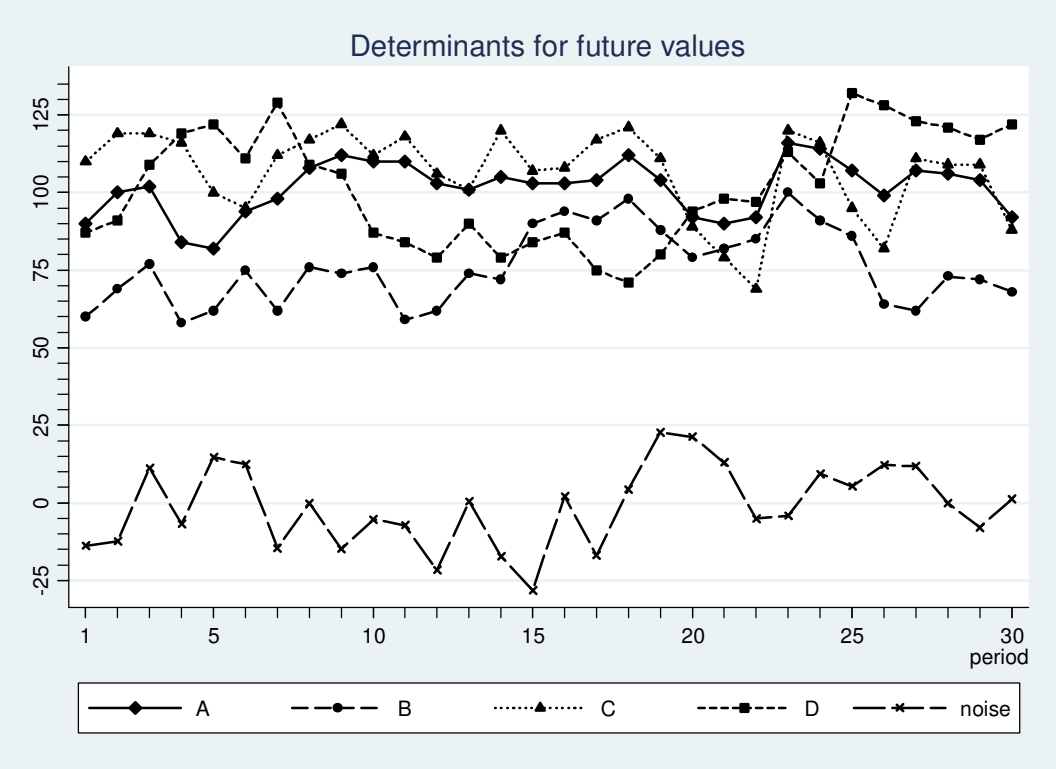


Figure 1: Time series' for determinants

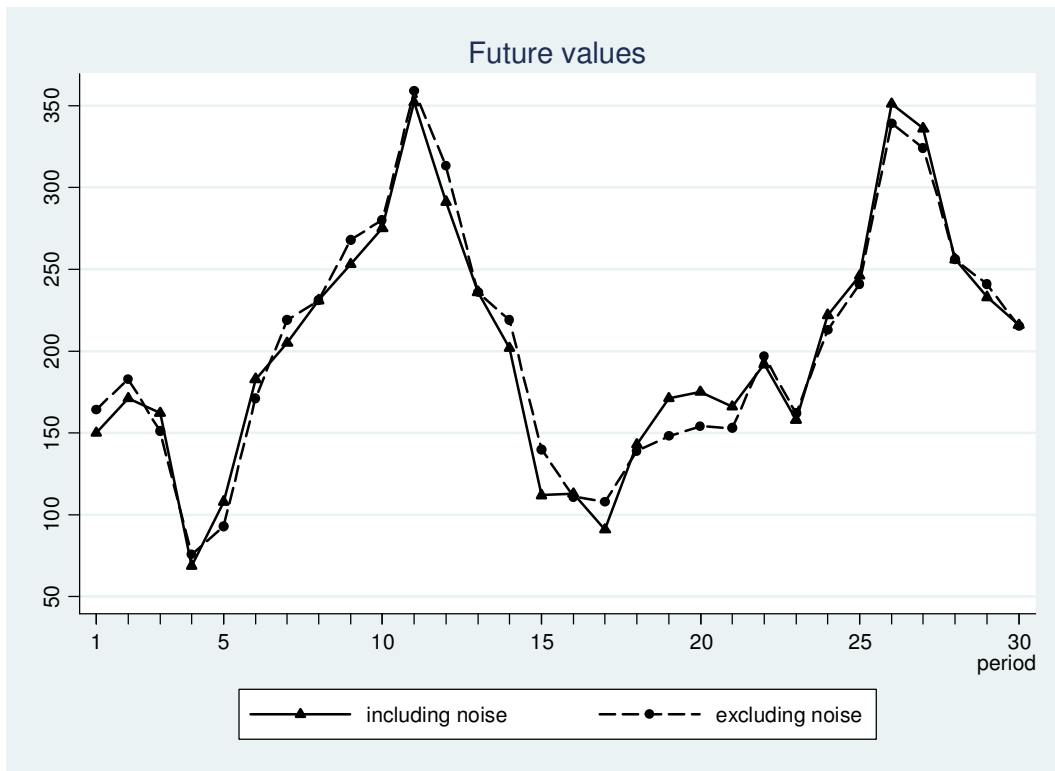


Figure 2: Time series to be forecast

Along with every forecast, subjects are asked to report a self-assessment on three questions. The first question refers to the subject’s effort when making the prediction, the second question prompts subjects to assess the accuracy of the respective forecast and the third is concerned with subjects’ confidence about the accuracy of the forecast given. All three questions have to be answered by noting the respective assessment on a scale from 1, representing “very low”, to 7, reflecting “very high”. Answering the three obligatory questions is not rewarded by an additional payoff.

Coordination task

Subjects are matched to societies of eight players. Each society consists of four teams with two members. Within a team, members are allowed to communicate using the chat implemented in zTree (Fischbacher, 2007). We thus simulate analyst communities that exchange information and potentially influence each other’s decisions. However, the two group members are not obliged to give unanimous forecasts. All subjects in a society receive information about all other players’ forecasts in the previous periods, as well as the respective payments achieved (costs and bonuses). This structure allows participants to coordinate by observing all others’ decisions, whereby coordination can be fostered explicitly through the chat and implicitly through the forecast itself.

Payoffs

Our payoff regime incorporates three distinct aspects, as described below.

Firstly, we implement *costs of forecasts* equal to 1Cent times the percentage point deviation from the respective current value. According to Brown (1998), subjects are more likely to make more accurate forecasts when threatened by losses, rather than continuous positive gains. Thus, we endow each subject with €8.00, while the payoff for each period can become negative. Let C_t denote the costs of forecast denominated in € in period t and \tilde{x}_t the forecast aiming at x_{t+h} ; accordingly, we can write:

$$C_t(\tilde{x}_t, x_t) = \left| \frac{\tilde{x}_t - x_t}{x_t} \right|$$

Building on Brennscheidt (1993), we implement a punishment of extreme predictions. More importantly, these costs of forecasts ensure the existence of a unique, risk dominant sunspot equilibrium for coordination, as discussed in the subsequent section.

Secondly, each participant is paid an *accuracy bonus* of €2.5 for every accurate prediction, whereby we define a prediction as accurate if $0.95x_{t+h} \leq \tilde{x}_t \leq 1.05x_{t+h}$, i.e. if the interval $M_t = [0.95x_{t+h}, 1.05x_{t+h}]$ contains \tilde{x}_t . Thus, an accurate prediction deviates from the correct value by 5% at most. Let $y_{i,t}^a$ denote the function that determines the accuracy bonus payment for each player $i = 1, \dots, 8$; therefore, we can formulate:

$$y_{i,t}^a(\tilde{x}_{i,t}) = \begin{cases} 1, & \text{if } \tilde{x}_{i,t} \in M_t \\ 0, & \text{else} \end{cases} \quad \forall 9 < t < 26$$

Thirdly, there is a *coordination bonus* of €1 in each period, for which at least five of the eight subjects have to give forecasts in a specific interval. The range of the interval is derived from the five forecasts that are closest in relative terms in each period. The average of the highest and the lowest of these five forecasts builds the midpoint of the interval, with the lower and the upper bound again defined by a 5% deviation.

Let \tilde{x}_t^r denote the prediction in period t that is assigned rank r . Ranks are assigned to predictions sorted by their size, where $r = 1, \dots, 8$. X_t^j denotes the j^{th} vector in period t , comprising five forecasts. Thereby, vector $j = 1$ includes values from the smallest forecast ($\tilde{x}_t^{r=1}$) to the fifth smallest forecast ($\tilde{x}_t^{r=5}$). The same procedure applies to $j = 2, 3, 4$.

Therefore, we define:

$$X_t^j = (\tilde{x}_t^{r=j}, \tilde{x}_t^{r=j+1}, \tilde{x}_t^{r=j+2}, \tilde{x}_t^{r=j+3}, \tilde{x}_t^{r=j+4}) \forall j = 1, \dots, 4$$

$X_t^{j^*}$ denotes the vector with the smallest relative deviation of the upper bound from the lower bound. Let $A(X_t^j)$ denote a function that gives these relative deviations; hence:

$$A(X_t^{j^*}) = (\tilde{x}_t^{r=j^*+4} / \tilde{x}_t^{r=j^*}) \leq A(X_t^j) = (\tilde{x}_t^{r=j+4} / \tilde{x}_t^{r=j}) \forall j^* \neq j$$

The interval for deriving the coordination bonus can then be written as:

$$H_t = [0,95(\tilde{x}_t^{r=j^*} + \frac{(\tilde{x}_t^{r=j^*+4} - \tilde{x}_t^{r=j^*})}{2}), 1,05(\tilde{x}_t^{r=j^*} + \frac{(\tilde{x}_t^{r=j^*+4} - \tilde{x}_t^{r=j^*})}{2})]$$

If at least five subjects make forecasts lying within this interval, these subjects are rewarded the coordination bonus. Let $y_{i,t}^c$ denote the function to determine the coordination bonus for each subject and period; therefore:

$$y_{i,t}^c(\tilde{x}_{i,t}, \tilde{x}_{k \neq i,t}) = \begin{cases} 1, & \text{if } \tilde{x}_{i,t} \in H_t, \sum_{i=1}^8 g_{i,t} \geq 5 \\ 0, & \text{else} \end{cases} \quad \forall i, t$$

$$\text{with } g_{i,t}(\tilde{x}_t^i) = \begin{cases} 1, & \text{if } \tilde{x}_{i,t} \in H_t \\ 0 & \forall i, t \end{cases}$$

The comprehensive payoff function comprising the three components for each player and period is then given by:

$$Y_{i,t}(\tilde{x}_{i,t}, \tilde{x}_{k \neq i,t}) = y_{i,t}^a(\tilde{x}_{i,t}) + y_{i,t}^c(\tilde{x}_{i,t}, \tilde{x}_{k \neq i,t}) - C_{i,t}(\tilde{x}_{i,t}, x_t)$$

In short, subjects can receive payments by making accurate forecasts and coordinating, while they have to bear costs when deviating from each rounds' naïve forecast.

3.2 PROPERTIES OF THE GAME

This section develops the criteria for rational behavior and formulates expectations concerning subjects' behavior.

Optimal forecasts

The optimal strategy unrestricted by cognitive limitations is given by forecasts equal to the expected future values. Note that the accuracy bonus is not guaranteed by an optimal forecast

due to the realizations of the random variable ε_t . However, given the actual time series of x_t forecasts equal to expected future values pay the accuracy bonus in 14 of the 21 forecasting periods. The average costs for these forecasts amount to €0.52. Thus, in the case of optimal forecasting, an average net payment of €1.15 is realized.

In this case, whether additional payments for coordination are realized depends on other players' behavior. Consider the situation in which at least five players of a society give optimal forecasts. Each of these subjects then earns the *cooperation bonus*, in addition to the *accuracy bonus*. Accordingly, the average net payment amounts to €2.15 per period.

The coordination regime

Given the forecasting task's complexity, it is reasonable to assume that subjects will fail to make optimal forecasts. In this case, they are enabled to establish a coordination regime, which pays the *coordination bonus*. Note that all forecasting values inducing costs smaller than €1 are Nash-equilibria for coordination. However, there is a strictly risk dominant equilibrium in coordination, which is given by the current values. Thereby, naïve forecasts can realize the coordination bonus without incurring any costs. This can be interpreted as a sunspot equilibrium, given that current values hold very little additional information for forecasting future values.

In sum, effective coordination is possible in many ways, e.g. by forecasting a simple trend or choosing a series of salient numbers like 200,300,200 etc. However, there is only one possibility of efficient coordination given by forecasting the current values. While inefficient coordination pays the *coordination bonus* of €1 minus the costs of deviations from current values, efficient coordination gives an average net payment equal to the *coordination bonus*. Nevertheless, the average net payment for efficient coordination is still lower than for making accurate forecasts without having four other players predicting optimally.

Minimizing Costs

A third potential outcome is intertwined with the attempt to establish a coordination regime. Risk-averse subjects might rely on the strategy of naïve forecasting to minimize costs, whereby an efficient coordination might evolve. In the absence of coordination, this strategy pays €0 in each period and subjects would end up with their initial endowment of €8.

Expected behavior

Having outlined three basic strategies players can choose to pursue, it cannot be unambiguously defined which strategy ultimately defines rational behavior, given that it depends on each subject's cognitive ability and the expectation toward the other subjects' behavior within a society.⁷ Moreover, risk preferences might play a crucial role, especially since potentially negative outcomes in each period are likely to trigger loss aversion.

Overall, there is a well-defined and easily accessible risk dominant equilibrium in coordination, given by the strategy of naïve forecasting. Thereby, current values can be interpreted as sunspots facilitating coordination. We are interested in whether societies are able to establish this efficient equilibrium, i.e. a coordination regime at minimal costs, and, if so, how fast, comprehensive and stable coordination proves to be. Furthermore, we look into the dynamics of coordination and identify whether the successful cooperation of others prompts subjects to join in, dismissing their attempts at making accurate forecasts.

3.3 PROCEDURE

The experiment was conducted in the Laboratory for Behavioral Economics at the University of Göttingen. Participants were recruited using ORSEE (Greiner, 2004) and were only allowed to participate in one session. Experiments were programmed using z-Tree (Fischbacher, 2007). There were 25 sessions featuring 25 societies in December 2011 and January 2012, which gives us 200 participants in total. Prior to each session, understanding of instructions was made sure by running control questions. The sessions lasted around ninety minutes with participants earning €14.50 on average. Participants were on average 23 years old, 55% were female and 38% students of economics or business administration.

⁷ To analyze incentives and expectations in detail, one might consider a level- k approach as presented by Nagel (1995). However, a level-k based approach to individual behavior is beyond the exploratory scope of this paper.

4. RESULTS

We analyze our results with respect to prediction accuracy, the effectiveness of coordination and overall welfare. Furthermore, we are interested in subjects' self-assessment concerning their effort to make predictions, as well as the confidence in their respective accuracy.

4.1 ACCURACY OF PREDICTIONS

We begin by analyzing the accuracy of predictions. Therefore, we consider the *prediction-accuracy matrix*, as suggested by Andres and Spiwoks (1999). Accordingly, forecasts are classified within four categories set up by two dimensions, i.e. by the combination of *Theil's new inequality coefficient* (Theil, 1975) and the *TOTA-coefficient*. These two measures assess forecasting accuracy by a straightforward comparison of actual and naïve forecasts. We are thus able to reveal subjects' attempts at coordination on the risk dominant equilibrium by measuring their forecasting performance.

Theil's new inequality coefficient

To classify predictions within the *prediction-accuracy matrix*, we primarily need to calculate *Theil's new inequality coefficient*, which allows for an implicit comparison of actual and naïve forecasts. It is defined as the square root of the ratio resulting when dividing the mean of squared prediction errors by the actual mean of changes in the objective variable. Again, let T denote the total number of forecasts, h the forecast horizon, x_t the actual value in period t and \tilde{x}_t the forecast for period t . The coefficient is then defined as follows:

$$U = \frac{\sqrt{\frac{1}{T-h} \sum_{t=h+1}^T (P_t - A_t)^2}}{\sqrt{\frac{1}{T-h} \sum_{t=h+1}^T (A_t)^2}} \quad \text{with} \quad P_t = \frac{\tilde{x}_t - x_{t-h}}{x_{t-h}} \quad \text{and} \quad A_t = \frac{x_t - x_{t-h}}{x_{t-h}}$$

For exact predictions, U becomes 0. Predictions worse than the naïve one are indicated by $U > 1$ and better predictions by $U < 1$.

Pooled over all subjects, U 's average amounts to 1.045, whereby 33% of prediction series show a value smaller than 1, while only 2.5% achieve a value smaller than 0.9. This clearly indicates weak prediction accuracy in general.

Furthermore, we are able to decompose the prediction error by rearranging the coefficient equation. We can write:

$$\frac{\frac{1}{T-h} \sum_{t=h+1}^T (P_t - A_t)^2}{\frac{1}{T-h} \sum_{t=h+1}^T (P_t - A_t)^2} = \frac{(\bar{P} - \bar{A})^2 + (s_P - s_A)^2 + 2(1 - r_{PA})s_P s_A}{\frac{1}{T-h} \sum_{t=h+1}^T (P_t - A_t)^2} = ME + VE + RE = 1$$

$$\text{with } ME = \frac{(\bar{P} - \bar{A})^2}{\frac{1}{T-h} \sum_{t=h+1}^T (P_t - A_t)^2} \quad VE = \frac{(s_P - s_A)^2}{\frac{1}{T-h} \sum_{t=h+1}^T (P_t - A_t)^2} \quad RE = \frac{2(1 - r_{PA})s_P s_A}{\frac{1}{T-h} \sum_{t=h+1}^T (P_t - A_t)^2}$$

\bar{P} , \bar{A} denote the averages of P_t and A_t respectively. s_P and s_A denote the standard deviation of P_t and A_t ; r_{PA} gives the respective correlation coefficient. Thus, ME describes the error with respect to the mean of future values, i.e. it measures the systematic error in predictions regarding the level of future values. A systematic error in forecasts with respect to the variance of future values is captured by VE. Accordingly, RE measures the random, non-systematic component of prediction errors. Andres and Spiwoks (1999) point out that the quality of predictions crucially depends on their ability to correctly assess the variance and levels of future values. If forecasts fail to capture the two aspects, this indicates the inability to forecast the general development of the objective time series. Better forecasts are indicated by prediction errors that are driven by the random component.

Across all forecast series, we find the median share of the random component to be 51%, while the variance component is about 32% and the miscalculation of the future values' level is 16%. Hence, only half of the total prediction error can be ascribed to the random component.

TOTA-coefficient

In order to more appropriately assess prediction accuracy, we additionally calculate the second dimension of the *prediction-accuracy matrix*: the *Topically Oriented Trend Adjustment (TOTA)-coefficient* (see Andres and Spiwoks, 1999; Bofinger and Schmidt, 2003). This coefficient represents the ratio of forecasts and correct values correlation, as well as the correlation of forecasts with current values at the time the forecast was made.

For a forecast that is orientated at future values, we can write in short:⁸

$$TOTA - coefficient = \frac{R_{forecats, future values}^2}{R_{forecasts, current values}^2} = \frac{R_{\hat{x}_t, x_{t+h}}^2}{R_{\hat{x}_t, x_t}^2} > 1$$

Where R^2 is the coefficient of determination, i.e. the squared correlation in a simple linear OLS regression without an intercept when regressing forecasts on future or current values. A $TOTA-coefficient < 1$ indicates topically oriented trend adjustment, i.e. forecasts more heavily rely on current values than future values. Essentially, topically oriented trend adjustment represents overestimation in case of a downward trend in the course of the objective value and an underestimation if there is an upward trend of future values.

Table 1 summarizes the results for the $TOTA-coefficient$ pooled for all subjects.

	Mean	std. dev.	min	max
<i>TOTA – coefficient</i>	0.6	0.02	0.49	0.73
$R_{forecats, future values}^2$	0.18	0.012	0.12	0.23
$R_{forecasts, current values}^2$	0.30	0.02	0.19	0.32

Table 1: The $TOTA-coefficient$

Overall, the correlation of forecasts and future values is weaker than the correlation between forecasts and current values. Given that no subjects achieve a $TOTA-coefficient$ greater than 1, all subject are prone to a topically trend adjustment by considerably relying on current values, which points to a coordination or cost minimizing strategy. The results are very homogenous, as can be seen by rather low standard deviations, as well as the minimum and maximum values of the coefficient.

Prediction-accuracy matrix

The *prediction accuracy matrix* allows us to categorize forecasts according to the $TOTA-coefficient$ and a selected goodness of fit measure, using *Theil's new inequality coefficient* (Andres and Spiwoks, 1999). The first category captures *quasi-naïve forecasts*, which can be characterized by the existence of topically oriented trend adjustment and a lower accuracy, which could have been achieved by naïve forecasts. Forecasts of the second category, denominated as *false forecasts*, similarly show low accuracy yet no topically oriented trend adjustment. The third category, *direction forecasts*, comprises forecasts showing topically oriented trend adjustment yet a higher accuracy than naïve forecasts. Finally, the fourth

⁸ For the comprehensive derivation of the $TOTA-coefficient$, we refer to Andres and Spiwoks (1999).

category, denominated as *optimal forecasts*, is achieved if forecast are better than the naïve forecasts and do not show topically oriented trend adjustment.

Figure 3 shows the distribution of mean individual forecasts over the four categories described.

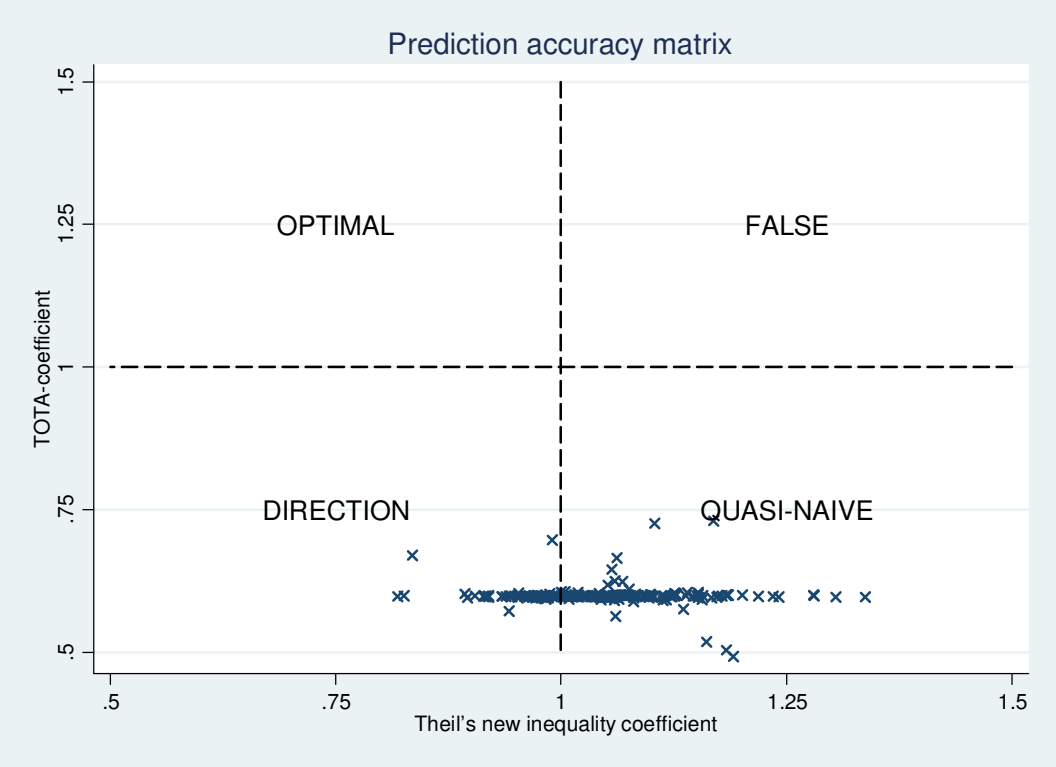


Figure 3: Prediction accuracy matrix for all subjects.

As can clearly be seen, the vast majority of forecasts fall within the category of *quasi-naïve forecasts* (67%), whereas only 33% are *direction forecasts*. All of the subjects forecast series are prone to the topically oriented trend adjustment. We can thus formulate our primary result.

RESULT 1: *Subjects fail to make accurate forecasts and rely considerably on current values.*

4.2 COORDINATION

We have shown that subjects are incapable of making accurate predictions, i.e. predictions that are equal to the expected future values. However, this might not reflect irrational behavior, given that participants could have chosen the alternative strategy of establishing a coordination regime. The effectiveness and stability of coordination is analyzed in the following, where we turn to the society level of aggregation to analyze our results.

Figures 4 to 6 offer representations of different aspects of coordination for each society, whereby we distinguish between effective and efficient coordination. The former merely

demands subjects to successfully establish a coordination regime, whereas the latter requires the coordination to minimize costs, i.e. using the respective naïve forecasts.

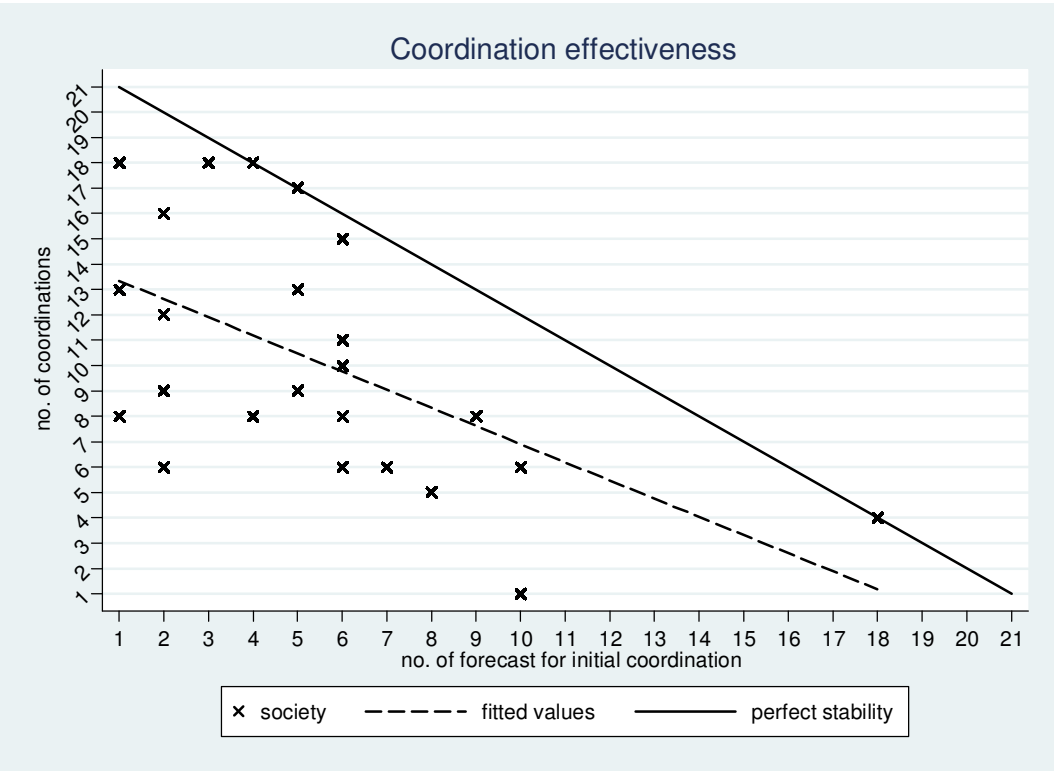


Figure 4: Coordination effectiveness.

Note: one society never managed to coordinate and is therefore not included. The number of coordination does not discriminate with respect to the number of subjects coordinating if at least the threshold of five players is met.

Figure 4 graphs the total number of successful coordination in relation to the initial period of successful coordination as an intuitive measure of effectiveness. As can be seen, the occurrence of a coordination regime within the first periods determines the subsequent course of coordination to some extent. Although there is a negative correlation of the initial period of coordination and the number of periods with successful coordination, only three societies manage to establish a perfectly stable regime, as indicated on the line presenting perfect stability. Thus societies achieving a successful coordination relatively early are likely to coordinate more often, but only few are able to coordinate in all subsequent periods. Although not perfectly stable, once established coordination regimes prove rather robust throughout the rest of the game. There is only one society that never coordinates.

Moreover, it can be expected that a successfully established coordination regime attracts the remaining subjects within a society. Figure 5 gives a more detailed perspective on the

dynamics of coordination by showing the average share of subjects coordinating within a society. Additionally, the share of societies featuring coordination in the respective period is graphed. Lastly, we present the share of subjects playing the risk-dominant and cost-minimizing coordination equilibrium. We define a subject as playing this strategy if her respective forecast does not deviate by more than 5% from the naïve forecast.

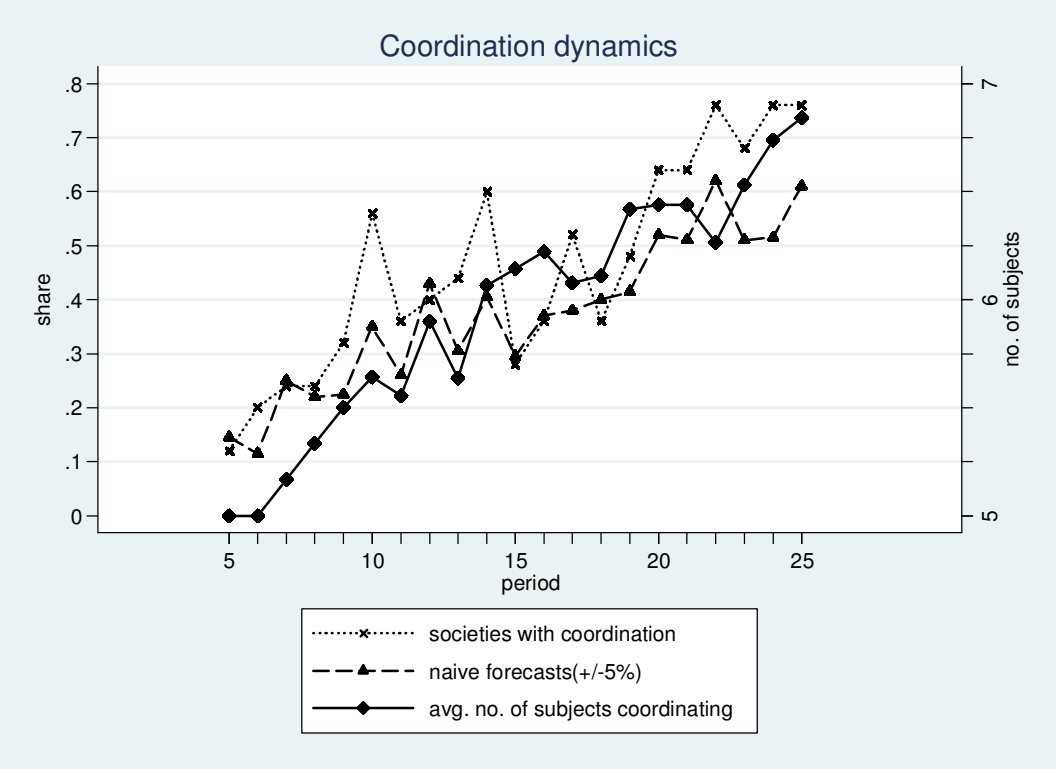


Figure 5: Coordination dynamics.

Note: share of coordinated societies with respect to all societies, naïve forecasts with respect to all forecasts and number of subjects coordinating.

Obviously, the share of societies achieving coordination is increasing over periods, which is driven by the rising share of subjects playing the risk dominant coordination equilibrium, i.e. make naïve forecasts. Also, the size of the coordination groups within societies increases throughout the game, which supports the notion of successful coordination attracting subjects that otherwise aim at accurate forecasts. However, only 22.63% of periods with successful coordination comprise all eight subjects of a society. In most cases (38.27%), only the minimum threshold of five players for the realization of the coordination bonus is reached. The coordination group is formed by six players in 32.1% of the times, and by seven players in 7%. Thus, a considerable number of subjects do not adjust to the majority within the

society following a coordination regime. They thereby forego considerably high payments by sticking to the strategy of attempting to forecast accurately.⁹

Considering efficiency of coordination, evidently naïve forecast are the predominant mechanism of coordination as the presentation of costs of forecasts in Figure 6 shows.

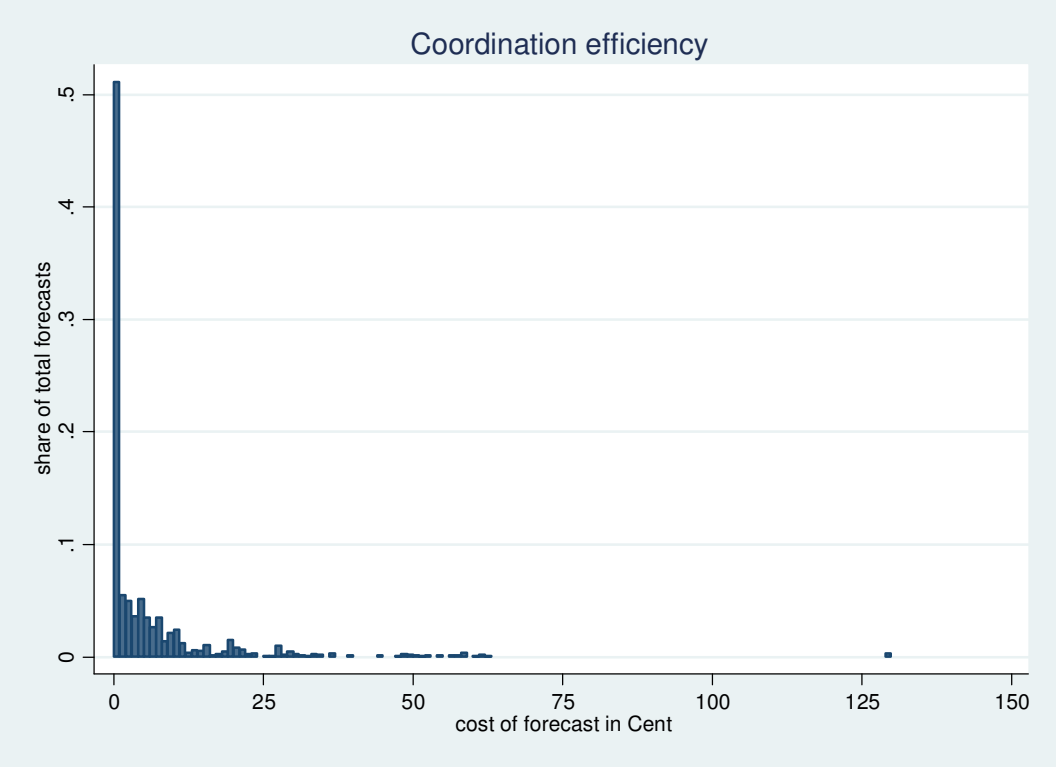


Figure 6: Coordination efficiency.

Note: only costs of forecasts yielding the coordination bonus are considered.

It can be seen that coordination is very efficient as the risk and payoff dominant equilibrium is used for a coordination regime most often, i.e. player coordinate on current values. More than 50% of forecast yielding the coordination bonus are exactly equal to current values, and the vast majority is not deviating more than 10% resulting in costs of less than 10 Cent. Coordination effectiveness is clearly fostered by the salient risk dominant equilibrium of making forecasts equal to current values. In none of the situations featuring successful coordination players manage to coordinate on the expected future values, abstaining from the cases where current values and future expected values happen to be similar. Thus the optimal

⁹ Another explanation might be seen in subjects' incapability of understanding the coordination mechanism. However, since they are able to communicate with one group member and there is full information about all other players' payments and forecasts this seems less likely.

solution of coordination on the expected future values is not achieved. We can formulate our second main result accordingly.

RESULT 2: Overall, coordination increases as the game proceeds and attracts subjects within societies, but attempts at forecasting are not universally discarded in favor of the coordination regime. However, subjects coordinate efficiently once they abstain from attempting to forecast accurately.

4.3 SELF-ASSESSMENT: EFFORT AND CONFIDENCE

Note that forecasts show poor quality (see Figure 3) even in the very first periods in which coordination is not established. Thus, it seems reasonable to assume that subjects fail to make correct forecasts and subsequently turn to a coordination strategy as indicated by the dynamics of coordination. This perspective is supported by the reported self-assessments.

As described above, subjects answered three questions in each forecasting period. These questions refer to self-assessments regarding subjects’ effort in forecasting, as well as their confidence in forecasting accuracy. Figure 7 provides an overview of the respective results.

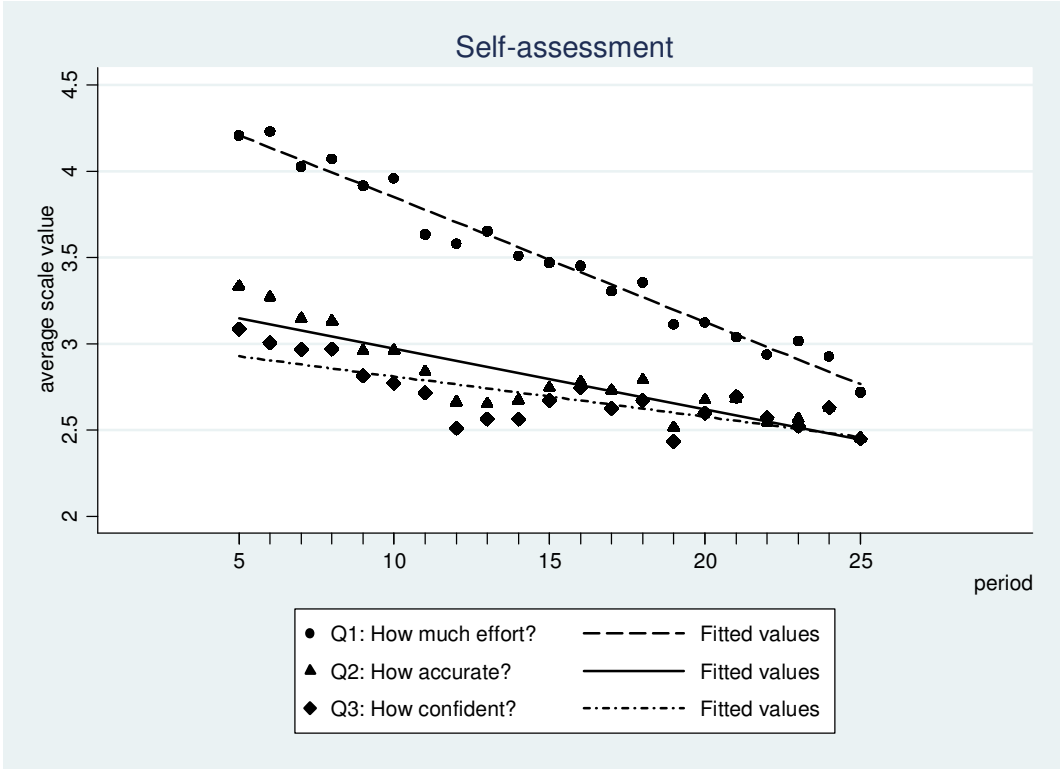


Figure 7: Overview of additional forecasting self-assessments

Not surprisingly, given increasing coordination, reported effort and confidence levels deteriorate as the game proceeds. Subjects’ effort is quite high for the first five periods, i.e.

before subjects receive feedback, as indicated by an average reported effort level of 4.1. In periods 6-15, the average deteriorates to 3.4 and ultimately to 2.9 for the last ten predictions, for which 50.2% of subjects report a value smaller than three. Thus, if a stable coordination regime is established, the effort for each additional forecast is rather low.

In line with this finding, there is a trend towards more pessimistic self-assessments with respect to the accuracy of forecasts, which is reflected in decreasing reported levels for the second and third question. For both questions, the median is equal to one in the last 10 periods, which is the lowest possible self-assessment. However, given that reported confidence levels start off lower than the effort levels, the downward trend is weaker.

In sum, there is more successful coordination as the game proceeds, which lowers the effort for making predictions, as well as the confidence in the accuracy of forecasts. Although there might be some bias in the self-assessments as they were not incentivized, the results are consistent with the findings presented in subsection 3.1 and 3.2.

***RESULT 3:** As effective coordination increases, subjects' forecasting efforts deteriorate. Accordingly, reported confidence levels decrease, given that subjects aim at coordination rather than accurate forecasts.*

4.4 WELFARE ANALYSIS

For a comprehensive welfare analysis, we use the structure of actual payoffs. Beforehand, we define four benchmarks to evaluate the payoffs achieved.

For the first-best strategy, i.e. deriving the correct formula from the first five periods and predicting the expected future value, an average payoff of €1.15 can be achieved as derived in section 3.2. We consider this case as our first benchmark. Optimally, at least four other subjects also forecast the expected values, which pays an additional €1 as a coordination bonus in each period. This gives us our second benchmark of €2.15. The third benchmark is formed by an efficient coordination at minimal costs for forecasts, i.e. by coordination on the risk dominant equilibrium. Thus, in each period, subjects would receive €1. The last benchmark is given by the cost minimizing strategy, which is realized if subjects give strictly naïve forecasts without an effective coordination. Given that costs of forecast are zero and there are no bonuses, the respective payoff amounts to €0. Recall that subjects were given a budget of €8 and payoffs for each period could become negative. On average, subjects earned

€14.50, showing that they yielded a positive amount on average, but less than the 29€ accruing from efficient coordination in all periods.

Figure 8 summarizes the average costs, bonuses and total payoffs for all periods, as well as future values.

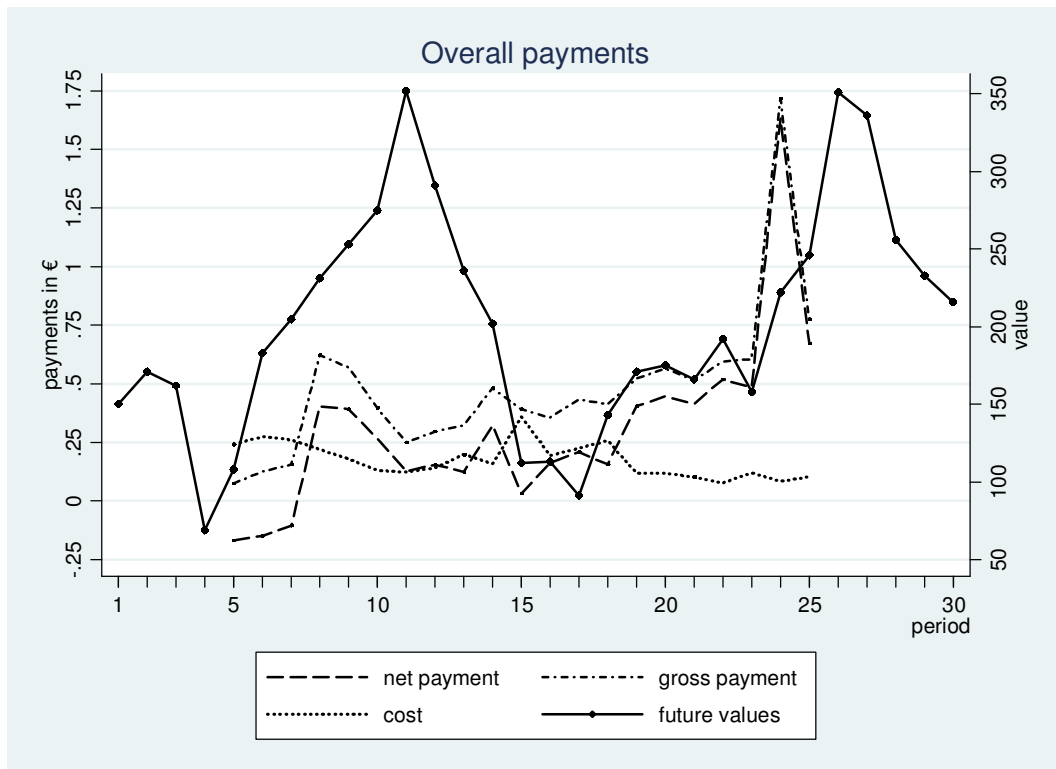


Figure 8: Overview of payoffs, costs of forecasts and future values.

It can be seen that subjects overall start off worse than the lowest benchmark given at €0. Subjects in the first periods unsuccessfully try to make accurate forecasts, thus bearing substantial costs. As the game proceeds, they increasingly coordinate effectively, which increases average payoffs. Net payments tend to go down for relatively low current values as attempts for accurate forecasts induce higher costs since relative deviation from the naïve forecast increases. However, subjects on average never reach the benchmark of efficient coordination, which would yield €1 on average. Although costs tend to decrease during the game, narrowing the gap between gross and net payments, efficient coordination is not realized ubiquitously.

There are some salient events in our time series of actual values to be forecast that warrant further discussion to explain the pattern of payoffs. Within the first periods, payments tend to increase until period 8, at which point they reach a local maximum, before subsequently decreasing for several periods. This can be explained by some correct forecasts by “accident”

in period 8 and 9 (37 of 200 in period 8; 28 of 200 in period 9). For period 8, the future value to be forecast amounts to 236, which is close to the current value of 231. In period 9, the future value is 202, which was correctly predicted by some subjects relying on salient numbers: in this case, 200. Additionally, there is an extraordinary high number of payments for group coordination, given that many subjects oriented their forecasts at the value of 200 (91 subjects received the coordination bonus). Moreover, many subjects gained the extra €2.5 in period 24 merely by sticking to the naïve forecast.

As shown in Figure 9, payoffs on the society level are very heterogeneous, which can be explained by the varying effectiveness and dynamics of coordination. Thus, some societies even approach average payoffs amounting to the benchmark of efficient coordination, while others achieve very low average payoffs due to unsuccessful coordination and weak forecasting accuracy.

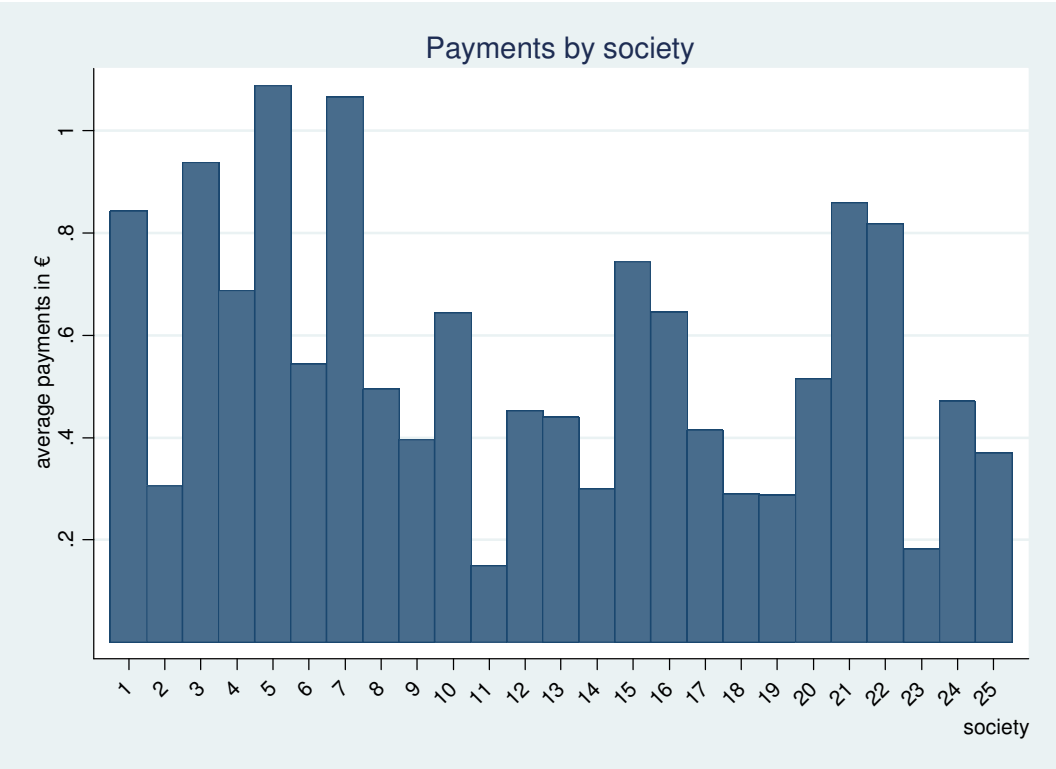


Figure 9: Average (net) payoffs per period on the society level

Overall, the welfare analysis shows that subjects fail to obtain the payoffs associated with perfectly rational forecasting. Also, coordination is not achieved universally, thus limiting the respective payoffs. However, on the society level, results are quite heterogeneous with overall welfare depending on subjects’ ability to learn to play the coordination equilibrium. We can formulate our forth core result.

RESULT 4: *Subjects fail to achieve maximal payoffs through rational forecasting. Coordination is not perfectly efficient, given that it takes some time for subjects to establish a coordination regime, which subsequently fails to be perfectly stable. Coordination considerably increases payoffs, yet not to the full extent possible. In addition, individual payments crucially depend on the society-specific dynamics.*

5. CONCLUSION

While the notion that analysts' "desire [...] to be perceived as good forecasters turns them into poor forecasters" (Ottaviani and Sorensen, 2006a, p.443) has been described theoretically and investigated empirically, the behavioral mechanisms of reputational herding remain largely unexplored. An empirical study in a controlled laboratory setting might thus reveal the behavioral processes involved in the interplay of forecasting accuracy and coordination among forecasters. In this paper, we aimed to experimentally investigate subjects' ability to coordinate on sunspot equilibria in a forecasting environment. Our study connects two separate strands of literature, one concerned with the efficiency of judgmental forecasting and the other discussing coordination on external focal points. Determining the occurrence, speed and stability of coordination on external signals adds to the understanding of reputational herding in forecasting settings from a behavioral perspective.

We provide results on subjects' behavior regarding the interplay of forecasting and coordination. While strong incentives for accuracy lead to initial efforts for achieving correct predictions, overall prediction accuracy is considerably low. As in previous experiments, we can reject the hypothesis of rational expectations for our forecasting task. Frustration with the task leads subjects to start playing the efficient coordination equilibrium, to which a majority of subjects conforms within a few periods. Sunspots are predominantly used to establish a simple coordination regime, while only few attempts to coordinate are made by relying on other salient numbers. Coordination on sunspot equilibria proves to be fairly robust once it is established. Overall welfare approaches but rarely reaches the benchmark payoff of efficient coordination. However, given subjects' inability to make correct forecasts, they forego significant payments rewarded for accurate predictions.

Presenting these results, we have illustrated a forecasting setting with conflicting incentives for accurate forecasts and coordination to show the effectiveness of payoff dominant SSE in fostering reputational herding. This provides additional empirical evidence for reputational

herding through external signals, as outlined by Spiwoks (2004). Accordingly, forecasters evade the responsibility of making accurate forecasts that correctly reveal their private information and coordinate on consensus values instead. This effectively conceals their inability to make correct predictions. Analysts who fail to coordinate and continue to pursue correct predictions in our setting gain comparably low payoffs, which in a real-world setting would translate to the loss of reputation. Accordingly, in a more dynamic setting, only forecasters capable of coordination would succeed. Those attempting to forecast honestly would ultimately lose their occupation. On a more general level, our results emphasize that once prediction markets are structurally incapable of providing incentives for forecasts that reveal the analyst's insight about the future state, it can be assumed that they will quickly resort to herding on extrinsic focal points. While this assessment might not be applicable to all prediction markets, the empirical results suggest that numerous financial and economic predictions can be characterized by herding on consensus values. Therefore, further experimental studies should focus on investigating parameters that succeed in reducing herding in forecasting by introducing incentives for the revelation of private information and thus fostering anti-herding.

ACKNOWLEDGEMENTS

The authors would like to thank Zulia Gubaydullina, Johannes Scheier and Stefan Schulz-Hardt for their very helpful comments on this work. We also thank participants of the GLOBE doctoral research seminar hosted by Claudia Keser. Joana Voigt and Dennis Kotte provided excellent student assistance in programming and conducting the experiment.

REFERENCES

- Ackert, L.F., Church, B.K., Ely, K., 2008. Biases in Individual Forecasts: Experimental Evidence. *The Journal of Behavioral Finance* 9 (2), 53-61.
- Ager, P., Kappler, M., Osterloh, S., 2009. The accuracy and efficiency of the Consensus Forecasts: A further application and extension of the pooled approach. *International Journal of Forecasting* 25 (1), 167-181.
- Agranov, M., Schotter, A., 2012. Ignorance Is Bliss: An Experimental Study of the Use of Ambiguity and Vagueness in the Coordination Games with Asymmetric Payoffs. *American Economic Journal: Microeconomics* 4 (2), 77-103.
- Anderson, L.R., Holt, C.A., 1997. Information cascades in the laboratory. *The American Economic Review* 87 (5), 847-862.
- Andres, P., Spiwoks, M., 1999. Prognosequalitätsmatrix – Ein methodologischer Beitrag zur Beurteilung der Güte von Kapitalmarktprognosen. *Jahrbücher für Nationalökonomie und Statistik* 219 (5-6), 513-542.
- Arifovic, J., Evans, G., Kostyshyna, O., 2013. Are sunspots learnable? An experimental investigation in a simple general-equilibrium model. Bank of Canada, Working Paper No. 2013-14.
- Asaad, C.T., 2012. Experimental finance: a cross-disciplinary exploration linking expectations and behaviours. *International Journal of Behavioural Accounting and Finance* 3 (3), 244-269.
- Avery, C.N., Chevalier, J.A., 1999. Herding over the career. *Economics Letters* 63 (3), 327-333.
- Banerjee, A.V., 1992. A simple model of herd behavior. *The Quarterly Journal of Economics* 107 (3), 797-817.
- Bardsley, N., Mehta, J., Starmer, C., Sugden, R., 2010. Explaining Focal Points: Cognitive Hierarchy Theory versus Team Reasoning. *The Economic Journal* 120 (543), 40-79.
- Bar-Isaac, H., 2008. Transparency, career concerns, and incentives for acquiring expertise. *The BE Journal of Theoretical Economics* 12 (1), 1-15.
- Batchelor, R., 2007. Bias in macroeconomic forecasts. *International Journal of Forecasting* 23 (2), 189-203.

- Batchelor, R., Dua, P., 1992. Conservatism and consensus-seeking among economic forecasters. *Journal of Forecasting* 11 (2), 169-181.
- Batchelor, R.A., Dua, P., 1990. Product differentiation in the economic forecasting industry. *International Journal of Forecasting* 6 (3), 311-316.
- Becker, O., Leitner, J., Leopold-Wildburger, U., 2005. Modelling judgmental forecasts under tabular and graphical data presentation formats, in: Schmidt, U., Traub, S. (Eds.), *Advances in Public Economics: Utility, Choice and Welfare* New York: Springer, 255-266.
- Becker, O., Leitner, J., Leopold-Wildburger, U., 2007. Heuristic modeling of expectation formation in a complex experimental information environment. *European Journal of Operational Research* 176 (2), 975-985.
- Becker, O., Leitner, J., Leopold-Wildburger, U., 2008. Modeling expectation formation involving several sources of information. *German Economic Review* 9 (1), 96-112.
- Becker, O., Leitner, J., Leopold-Wildburger, U., 2009. Expectation formation and regime switches. *Experimental Economics* 12 (3), 350-364.
- Bedke, N., Bizer, K., Spiwoкс, M., 2009. Gregarious Analysts - Experimental Evidence for Reputational Herding. *Journal of Money, Investment and Banking* 12, 26-36.
- Bernhardt, D., Campello, M., Kutsoati, E., 2006. Who herds?. *Journal of Financial Economics* 80 (3), 657-675.
- Bikhchandani, S., Hirshleifer, D., Welch, I., 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* 100 (5), 992-1026.
- Bizer, K., Scheier, J., Spiwoкс, M., 2013. Planspiel Kapitalmarktprognose. Ein empirischer Vergleich der Prognosekompetenz von Amateuren und Experten. *Sofia-Studien* 13-2.
- 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 (4), 437-449.
- Bosch-Domènech, A., Vriend, N.J., 2013. On the role of non-equilibrium focal points as coordination devices. *Journal of Economic Behavior & Organization* 94, 52-67.
- Boyson, N.M., 2010. Implicit incentives and reputational herding by hedge fund managers. *Journal of Empirical Finance* 17 (3), 283-299.

- Brennscheidt, G., 1993. Predictive Behavior – An Experimental Study. Lecture Notes in Economics and Mathematical Systems 403. Berlin: Springer.
- Brown, P.M., 1998. Experimental evidence on the importance of competing for profits on forecasting accuracy. *Journal of Economic Behavior & Organization* 33 (2), 259-269.
- Chen, Q., Jiang, W., 2006. Analysts' weighting of private and public information. *Review of Financial Studies* 19 (1), 319-355.
- Chevalier, J., Ellison, G., 1999. Career concerns of mutual fund managers. *The Quarterly Journal of Economics* 114 (2), 389-432.
- Clarke, J., Subramanian, A., 2006. Dynamic forecasting behavior by analysts: Theory and evidence. *Journal of Financial Economics* 80 (1), 81-113.
- Clement, M.B., Tse, S.Y., 2005. Financial analyst characteristics and herding behavior in forecasting. *The Journal of Finance* 60 (1), 307-341.
- Cooper, R.A., Day, T.E., Lewis, C.M., 2001. Following the leader: a study of individual analysts' earnings forecasts. *Journal of Financial Economics* 61 (3), 383-416.
- Cote, J., Sanders, D., 1997. Herding behavior: Explanations and implications. *Behavioral Research in Accounting* 9, 20-45.
- Crawford, V.P., Gneezy, U., Rottenstreich, Y., 2008. The power of focal points is limited: even minute payoff asymmetry may yield large coordination failures. *The American Economic Review* 98 (4), 1443-1458.
- Devenov, A. Welch, I., 1996. Rational herding in financial economics. *European Economic Review* 40, 603-615.
- Dovern, J., Weisser, J., 2011. Accuracy, unbiasedness and efficiency of professional macroeconomic forecasts: An empirical comparison for the G7. *International Journal of Forecasting* 27 (2), 452-465.
- Duffy, J., Fisher, E.O.N., 2005. Sunspots in the Laboratory. *The American Economic Review* 95 (3), 510-529.
- Duffy, J., Lai, E. K., Lim, W., 2013. Language and Coordination: An Experimental Study. *Genesis - The Journal of Genetics and Development* 11, 6-7.
- Effinger, M.R., Polborn, M.K., 2001. Herding and anti-herding: A model of reputational differentiation. *European Economic Review* 45 (3), 385-403.

- Ehrbeck, T., Waldmann, R., 1996. Why are professional forecasters biased? Agency versus behavioral explanations. *The Quarterly Journal of Economics* 111 (1), 21-40.
- Elliott, G., Timmermann, A., 2008. Economic forecasting. *Journal of Economic Literature* 46 (1), 3-56.
- Fehr, D., Heinemann, F., Llorente-Saguer, A., 2011. The Power of Sunspots: An Experimental Analysis. SFB 649 Discussion Paper No.070.
- Feri, F., Irlenbusch, B., Sutter, M., 2010. Efficiency gains from team-based coordination - Large-scale experimental evidence. *The American Economic Review* 100 (4), 1892-1912.
- Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2), 171-178.
- Froot, K.A., Scharfstein, D.S., Stein, J.C., 1992. Herd on the Street: Informational Inefficiencies in a Market with Short-Term Speculation. *The Journal of Finance* 47 (4), 1461-1484.
- Gallo, G.M., Granger, C.W., Jeon, Y., 2002. Copycats and Common Swings: The Impact of the Use of Forecasts in Information Sets. IMF staff Papers No 49 (1).
- Graham, J.R., 1999. Herding among investment newsletters: Theory and evidence. *The Journal of Finance* 54 (1), 237-268.
- Gregory, A.W., Smith, G.W., Yetman, J., 2001. Testing for forecast consensus. *Journal of Business & Economic Statistics* 19 (1), 34-43.
- Greiner, B., 2004. An online recruitment system for economic experiments. *GWDG Berichte* 63, 79-93.
- Gubaydullina, Z., Hein, O., Spiwoks, M., 2011. The Status Quo Bias of Bond Market Analysts. *Journal of Applied Finance & Banking* 1 (1), 31-51.
- Harvey, N., 2007. Use of heuristics: Insights from forecasting research. *Thinking & Reasoning* 13 (1), 5-24.
- Hirshleifer, D., 1993. Managerial reputation and corporate investment decisions. *Financial Management* 22 (2), 145-145.
- Holmstrom, B., 1999. Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies* 66 (1), 169-182.

- Hong, H., Kubik, J.D., 2003. Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts. *The Journal of Finance* 58 (1), 313-351.
- Hong, H., Kubik, J.D., Solomon, A., 2000. Security analysts' career concerns and herding of earnings forecasts. *The Rand Journal of Economics* 31 (1), 121-144.
- Jegadeesh, N., Kim, W., 2010. Do analysts herd? An analysis of recommendations and market reactions. *Review of Financial Studies* 23 (2), 901-937.
- Keynes, J.M., 1936. *The General Theory of Employment, Interest and Money*, London: Macmillan.
- Kim, M. S., Zapatero, F., 2009. Rational bias and herding in analysts' recommendations. *Marshall Research Paper Series, Working Paper FBE 08-10*.
- Kugler, T., Kausel, E.E., Kocher, M.G., 2012. Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley Interdisciplinary Reviews: Cognitive Science* 3 (4), 471-482.
- Lamont, O.A., 2002. Macroeconomic forecasts and microeconomic forecasters. *Journal of Economic Behavior & Organization* 48 (3), 265-280.
- Laster, D., Bennett, P., Geoum, I.S., 1999. Rational bias in macroeconomic forecasts. *The Quarterly Journal of Economics* 114 (1), 293-318.
- Laux, C., Probst, D.A., 2004. One signal, two opinions: strategic heterogeneity of analysts' forecasts. *Journal of Economic Behavior & Organization* 55 (1), 45-66.
- Lawrence, M., Goodwin, P., O'Connor, M., Önköl, D., 2006. Judgmental forecasting: A review of progress over the last 25years. *International Journal of Forecasting* 22 (3), 493-518.
- Leitner, J., Leopold-Wildburger, U., 2011. Experiments on forecasting behavior with several sources of information—A review of the literature. *European Journal of Operational Research* 213 (3), 459-469.
- Leone, A., Wu, J., 2007. What does it take to become a superstar? Evidence from institutional investor rankings of financial analysts. *University of Rochester working paper*.
- Li, X., 2002. Performance, herding, and career concerns of individual financial analysts. *EFA 2002 Berlin Meetings Discussion Paper*.
- Marimon, R., Spear, S.E., Sunder, S., 1993. Expectationally driven market volatility: an experimental study. *Journal of Economic Theory* 61 (1), 74-103.

- Marinovic, I., Ottaviani, M., Sorensen, P.N., 2011. Modeling Idea Markets: Between Beauty Contests and Prediction Markets, in: Williams, L.V. (Ed.), Prediction Markets: Theory and Applications. London: Routledge, 4-17.
- Nagel, R., 1995. Unraveling in Guessing Games: An Experimental Study. *The American Economic Review* 85 (5), 1313-1326.
- Ottaviani, M., Sorensen, P., 2000. Herd behavior and investment: Comment. *The American Economic Review* 90 (3), 695-704.
- Ottaviani, M., Sorensen, P.N., 2006a. Professional advice. *Journal of Economic Theory* 126 (1), 120-142.
- Ottaviani, M., Sorensen, P.N., 2006b. The strategy of professional forecasting. *Journal of Financial Economics* 81 (2), 441-466.
- Ottaviani, M., Sorensen, P.N., 2006c. Reputational cheap talk. *The Rand Journal of Economics* 37 (1), 155-175.
- Prendergast, C., Stole, L., 1996. Impetuous youngsters and jaded old-timers: Acquiring a reputation for learning. *Journal of Political Economy* 104 (6), 1105-1134.
- Ramnath, S., Rock, S., Shane, P., 2008. The financial analyst forecasting literature: A taxonomy with suggestions for further research. *International Journal of Forecasting* 24 (1), 34-75.
- Roos, M.W., Beugnot, J., Gürgüc, Z., Ovlisen, F.R., 2010. Coordination Failure Caused by Sunspots. *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010. Session: Experimental Game Theory No. F13-V2*.
- Scharfstein, D.S., Stein, J.C., 1990. Herd behavior and investment. *The American Economic Review* 80 (3), 465-479.
- Shurchkov, O., 2013. Sunspot Coordination in Dynamic Global Games. Wellesley College Working Paper.
- Spiwoks, M., 2004. External triggered herding bei Rentenmarkt-Analysten. *Financial Markets and Portfolio Management* 18 (1), 58-83.
- Spiwoks, M., Bedke, N., Hein, O., 2008. Forecasting the past: the case of US interest rate forecasts. *Financial Markets and Portfolio Management* 22 (4), 357-379.
- Spiwoks, M., Bedke, N., Hein, O., 2010. Topically Orientated Trend Adjustment and Autocorrelation of the Residuals – An Empirical Investigation of the Forecasting

- Behavior of Bond Market Analysts in Germany between 1989 and 2007. *Journal of Money, Investment and Banking* 14, 16-35.
- Spiwojs, M., Bizer, K., Hein, O., 2005. Anchoring Near the Lighthouse: Bond Market Analysts' Behavior Coordination by External Signal. Wolfsburg Working Papers No. 05-02.
- Spiwojs, M., Hein, O., 2007. Die Währungs-, Anleihen- und Aktienmarktprognosen des Zentrums für Europäische Wirtschaftsforschung. *AStA Wirtschafts- und Sozialstatistisches Archiv* 1 (1), 43-52.
- Stark, T., 1997. Macroeconomic forecasts and microeconomic forecasters in the survey of professional forecasters. Federal Reserve Bank of Philadelphia Working Paper No. 97-10.
- Stekler, H.O., 2007. The future of macroeconomic forecasting: Understanding the forecasting process. *International Journal of Forecasting* 23 (2), 237-248.
- Stickel, S.E., 1992. Reputation and performance among security analysts. *The Journal of Finance* 47 (5), 1811-1836.
- Theil, H., 1975. *Applied Economic Forecasting* (3. Ed.). Amsterdam: North-Holland.
- Trueman, B., 1994. Analyst forecasts and herding behavior. *Review of Financial Studies* 7 (1), 97-124.
- Van Campenhout, G., Verhestraeten, J.F., 2010. Herding Behavior among Financial Analysts: a Literature Review. HUB Research Paper No. 39.
- Weizsäcker, G., 2010. Do we follow others when we should? A simple test of rational expectations. *The American Economic Review* 100 (5), 2340-2360.
- Welch, I., 1992. Sequential sales, learning, and cascades. *The Journal of Finance* 47 (2), 695-732.
- Welch, I., 2000. Herding among security analysts. *Journal of Financial Economics* 58 (3), 369-396.
- Zarnowitz, V., Lambros, L.A., 1992. Consensus and uncertainty in economic prediction, in: Zarnowitz, V. (Ed.), *Business Cycles: Theory, History, Indicators, and Forecasting*. Chicago: University of Chicago Press, 492-518.
- Zitzewitz, E., 2001a. Measuring herding and exaggeration by equity analysts and other opinion sellers. Graduate School of Business Stanford University Working Paper.

Zitzewitz, E., 2001b. Opinion-producing agents: career concerns and exaggeration. Graduate School of Business Stanford University Working Paper.

Zwiebel, J., 1995. Corporate conservatism and relative compensation. *Journal of Political Economy* 103 (1), 1-25.

APPENDIX A

Instructions for the experiment. The original instructions were in German and are available from the authors upon request.

General Information

In this game, you and other participants will make predictions for the development of a price. This price results from the development of different determinants. You know the four most important determinants (called A-D). Additionally, you know all prices up to the point of prediction. Based on this information, all participants make a prediction in every period. Your payoff depends on your own and your co-players' predictions. You may communicate via chat with another participant, who is randomly assigned to you at the beginning of the game. The chat messages and all other data are recorded in an anonymous form. Every participant has the same instructions and information.

The course of the game (see figure 1)

The game has 30 periods. In every period, you will make a prediction for the price in five periods. You will make your first prediction in period 5. In every period, you will see three consecutive screens:

The “Information and Chatscreen” (see: figure 2)

What you see: Development of prices up to the current period

Overview of the forecasts of all other participants

Development of the determinants (A-D)

Overview of the payoffs of all other participants

Additionally, you can use the chatbox to communicate with your group member and a calculator, which is active with a click on the respective symbol.

What you do: You make a prediction for the price in five rounds. To become accustomed with the game, you have 5 minutes to make your prediction in the first round, 3 minutes for the next five periods and 2 minutes for the rest of the game. You will see the time you have left for your prediction on the upper right side of your screen. You can reach the next screen by clicking “to the input screen”.

The “Input Screen”

What you see: You will see an input box and three questions about your prediction.

What you do: You enter your forecast within 30 seconds, answer the three questions and click “end period”. **Please note the time constraint:** If you allow the 30 seconds to pass without

clicking “end period”, the game proceeds automatically. Your prediction for this period would be zero.

The “Payoff Screen”

What you see: Your current and overall payoff will be shown to you for 15 seconds.

What you do: No input is required.

Your Payoff

You start the game with a starting payment of 8€. You gain payoffs depending on your predictions, but also have to bear prediction costs. Thus, your payoffs for a single period can become negative. The payoff for all periods will be summed up and paid to you after the game. For the payoff, your prediction-costs, prediction accuracy and forecasts of other participants will be considered as follows.

Prediction costs: You will bear prediction costs in every round. They are estimated by the deviation of your prediction from the current price. You will be deducted 1Cent per percentage point deviation.

Prediction accuracy: All prices are the result of a fixed formula and the four determinants in the respective round, which determines 95% of the future price. 5% of the price is random. If your prediction is accurate, you will receive 2,50€ as an accuracy bonus. Accurate means that your prediction does not deviate by more than 5% from the actual price. You will learn if your prediction was correct in the period for which your prediction was made. For example, in period 10, you will learn whether your prediction in round 5 for period 10 was correct.

Example: *The price in the current period is 100. Your prediction is 120; thus, you have immediate prediction-costs of 20. 5 periods later, you learn that the actual price was 125. Your prediction (120) does not deviate by more than 5% from the actual price. Thus, you receive a 2,50€ accuracy bonus for your prediction from 5 periods earlier.*

If you do not predict accurately, you will not receive an accuracy bonus and have to bear prediction costs.

The predictions of other participants

If the predictions of 5 (or more) participants do not deviate by more than 5% from a common price, each of the 5 participants receives 1€ as “group-bonus”.

Example: The current price is 100; 5 participants predict a price of 120 (+/-5%). The actual price, 5 periods later, is 150. The prediction was thus inaccurate. However, each of the 5 participants receives a 1€ group bonus and has prediction costs of -0,20€.

Overview:

There are four situations in every period that lead to different payoffs.

1. Your prediction is inaccurate and you and at least 4 other participants have not predicted a common price.

Your Payoff: You earn 0€ minus your prediction-costs.

(Payoff = 0€ accuracy bonus + 0€ group-bonus – prediction costs)

2. Your prediction is inaccurate, but you and at least 4 other participants have predicted a common price (+/- 5%).

Your Payoff: You earn 1€ minus your prediction costs.

(Payoff = 0€ accuracy bonus + 1€ group-bonus – prediction costs)

3. Your prediction is accurate and you and at least 4 other participants have not predicted a common price.

Your Payoff: You earn 2,50€ minus your prediction costs.

(Payoff = 2,50€ accuracy bonus + 0€ group-bonus – prediction costs)

4. Your prediction is accurate and you and at least 4 other participants have predicted a common price (+/- 5%).

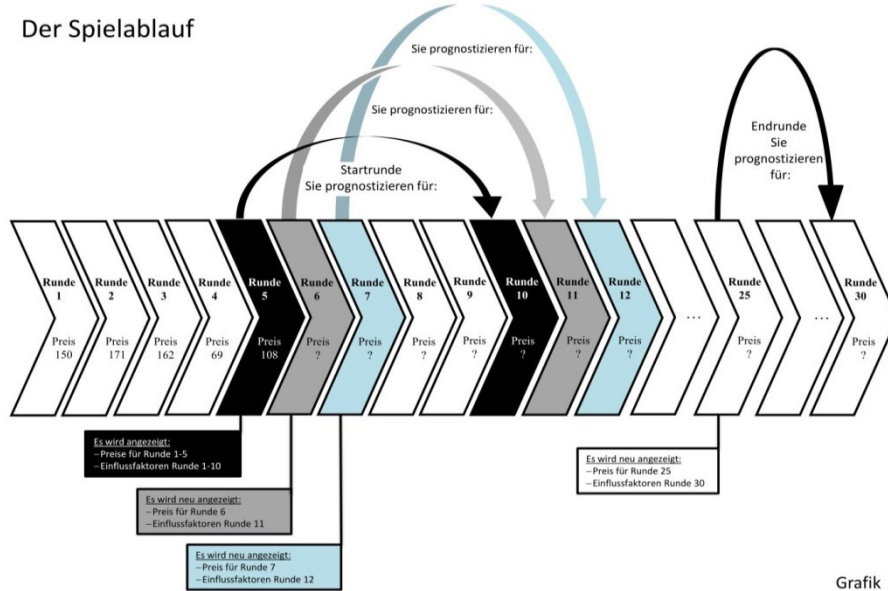
Your Payoff: You earn 3,50€ minus your prediction costs.

(Payoff = 2,50€ accuracy bonus + 1€ group-bonus – prediction costs)

We now ask you to answer some control questions on your computers. The game will start as soon as all participants have answered these questions correctly.

APPENDIX B

Additional explanatory graphics that were given to participants, along with the instructions that describe the structure of the game and the main screen.



Grafik 1

Der Informations- und Eingabebildschirm

The screenshot shows the game interface with the following callouts:

- a) Preise:** Hier sehen Sie die Preise in allen bisherigen Runden. (Points to the 'Bisherige Prognosen und Preise' table)
- b) Prognosen:** Hier sehen Sie die Prognosen aller Spieler in allen bisherigen Runden. (Points to the 'Bisherige Prognosen und Preise' table)
- c) Einflussfaktoren:** Hier sehen Sie die Einflussfaktoren; die aktuelle Runde hat eine Beschriftung. (Points to the 'Einflussfaktoren bis Runde 10' chart)
- d) Auszahlungen:** Hier sehen Sie alle Auszahlungen der Spieler in dieser Runde (links) und das Gesamtergebnis. (Points to the 'Auszahlungen (in EURO)' table)
- e) Chatbox:** Hier können sie mit Ihrem Mitspieler schreiben. (Points to the 'Kommunikationssystem für Gruppe 1' area)

Grafik 2

Chapter VII

A COMPARISON OF ENDOGENOUS AND EXOGENOUS TIMING IN A SOCIAL LEARNING EXPERIMENT

with Till Proeger and Hendrik Hüning

Published:

Cege Discussion Papers, no. 167, Göttingen University, August 2013.

Download: <http://www.uni-goettingen.de/de/60920.html>

ABSTRACT

This paper experimentally investigates social learning in a two-agent prediction game with both exogenous and endogenous ordering of decisions and a continuous action space. Given that individuals regularly fail to apply rational timing, we refrain from implementing optimal timing of decisions conditional on signal strength. This always renders it optimal to outwait the other player regardless of private signals and induces a gamble on the optimal timing and action. In this setting, we compare exogenous and endogenous ordering in terms of informational efficiency, strategic delay and social welfare. We find that more efficient observational learning leads to more accurate predictions in the endogenous treatments and increases informational efficiency compared to the benchmark exogenous treatment. Overall, subjects act sensitively to waiting costs, with higher costs fostering earlier decisions that reduce informational efficiency. For a simple implementation of waiting costs, subjects more successfully internalize information externalities by adjusting their timing according to signal strength. Simultaneous decisions in endogenous ordering avoid observational learning and compensate the higher degree of rational decisions. Overall, endogenous timing has no net effect on social welfare, as gains in accuracy are fully compensated by waiting costs. Our results hold relevance for social learning environments characterized by a continuous action space and the endogenous timing of decisions.

Keywords: endogenous timing; information externalities; laboratory experiment; social learning; strategic delay

JEL classification: C91, D82, D83

HIGHLIGHTS

- We compare social learning for endogenous and exogenous ordering with continuous action spaces.
- Endogenous ordering fosters more efficient observational learning.
- However, higher waiting cost and simultaneous decisions cancel out potential welfare gains.
- We find only few situations of excessive delay or war of attrition and no learning effects.
- Overall, social welfare is equal for exogenous and endogenous ordering of predictions.

1. INTRODUCTION

Studies on social learning emphasize the dismal effects of herding in information markets. Following the seminal papers by Banerjee (1992) and Bikhchandani et al. (1992), a number of studies show how rational subjects do not just follow private information but rather use public information constituted by prior decisions, thus frequently eliciting informational cascades. With individuals “following the crowd” (Fahr and Irlenbusch, 2011), private information is left unrevealed. Although subjects are on average reluctant to rationally follow cascades in experiments (Weizsäcker, 2010), socially non-optimal aggregation of information represents the core result of experimental studies following the seminal paper by Anderson and Holt (1997).¹

Recent studies by SgROI (2003), Ziegelmeyer et al. (2005), Çelen and Hyndman (2012) and, most recently, Ivanov et al. (2013) have furthered the analysis of social learning by allowing for endogenous ordering of decisions.² They point to fairly efficient observational learning, as well as deviations from rational timing that result in informational inefficiency. However, none of these studies allow for a quantification of the effect on informational efficiency and overall welfare, which requires comparison to a benchmark setting with exogenous decision order. The contribution of this paper is to investigate the degree to which information is used efficiently in a game of social learning with exogenous compared to endogenous timing of decisions. We further add to the discussions on non-optimal information aggregation by quantifying the net welfare effect of introducing endogenous rather than exogenous decision orders. To this end, we implement a two-player prediction game based on the theoretical model by Gul and Lundholm (1995). We compare a benchmark treatment of an exogenously fixed decision order with three treatments of endogenous ordering. To quantify information efficiency, we introduce continuous action spaces rather than binary action sets as used in

¹ Other studies using Anderson and Holt’s (1997) urn experiment include Willinger and Ziegelmeyer (1998), Anderson (2001), Hung and Plott (2001), Oberhammer and Stiehler (2003), Nöth and Weber (2003), Kübler and Weizsäcker (2004), Cipriani and Guarino (2005), Drehmann et al. (2005), Alevy et al. (2007), Goeree et al. (2007), Ziegelmeyer et al. (2008), Dominitz and Hung (2009), and Fahr and Irlenbusch (2011) for group players. Çelen and Kariv (2004) use the basic frame, yet implement continuous rather than binary signals.

² Many situations such as investment, market entry or forecasting are better characterized by endogenous ordering of choices. For instance, consider financial analysts forecasting future values of an economic variable. Analysts with little confidence in their private information may wait and observe other forecasts, as they are able to choose the point in time of their forecast. Since other analysts’ forecasts might reflect valuable information, analysts acting later tend to adjust their forecasts using the previous ones. Overall, information efficiency thus potentially improves. See Gul and Lundholm (1995) for an elaboration of these examples.

previous studies. Continuous action spaces allow for a concise analysis of social welfare resulting from the tradeoff between costs of delay and increased informational efficiency.

Our study constitutes an extension to the studies on observational learning that consider agents sequentially making binary choices in a fixed order, building on the seminal papers by Banerjee (1992) and Bikhchandani et al. (1992). Private information informs both agents imperfectly about the better alternative. Agents observe all preceding decisions. In the Nash Equilibrium (NE), subsequent agents might rationally discard their private information and information is aggregated inefficiently. The binary action set precludes the perfect transmission of information (Bikhchandani et al., 1998).³ Chamley and Gale (1994) extend the models for the endogenous ordering of decisions and waiting cost. There is strategic timing, given that prior decisions are public and have informational value; however, delaying a decision leads to waiting cost. In the NE, information is revealed imperfectly as there is either excessive delay (“war of attrition”) or no investments.⁴ Most relevant for our investigation of informational efficiency is the model of Gul and Lundholm (1995), who consider two agents predicting a value that is the sum of their distinct private information in continuous time. Private information is the realization of a uniformly distributed random variable, and thus the action set is continuous. The strength of the private signal is inversely related to waiting cost. This determines the optimal time of decision as both agents face a trade-off between accuracy of their prediction and delay costs. Individual predictions become public information. The resulting equilibria depend on the agents’ strategies. Firstly, in a unique symmetric NE, both agents act sensitively to their private signal, according to the trade-off between accuracy and delay costs. Due to the inverse relationship of private signals and delay costs, the timing of decisions reveals information about the signal strength, which improves the agents’ predictions. Secondly, in an asymmetric equilibrium, the first agent waits indefinitely for the other prediction regardless of her own signal. Given that excessive

³ The finiteness of the action set is explained by Bikhchandani et al. (1998) who state that informational cascades are likely to be most important for decision situations with “an element of discreteness or finiteness” (p.159) and that individuals tend to sort actions in discrete categories, even when they are actually continuous.

⁴ There are a number of further models for endogenous ordering of choices, with Chamley (2004) providing an overview. Closest to our investigation is Zhang (1997), who extends the basic model by informing agents about the precision of private information that is correlated with the true state. Agents with more precise information face higher waiting cost and thus act first. Zhang shows that for any given precision informational cascades will always occur in equilibrium. The equilibrium is inefficient due to excessive delay and imperfect revelation of private information. Frisell (2003) in turn introduces pay-off externalities. Strategic delay is reduced, as the advantage of being well informed decreases the stronger the pay-off externalities. For a sufficiently negative pay-off externality, the worst-informed agent acts first.

waiting is uninformative for the second agent, she predicts immediately and the first players' decision ensues. As both agents are insensitive to their signals, the result is similar to an exogenous decision sequence. In the symmetric NE, no informational cascades occur, but predictions of both agents are clustered due to two effects. The first mover anticipates that the other agent's signal is lower as she has not yet acted; the second agent in turn infers a higher signal of the first mover from her earlier prediction. The continuous action set allows for a perfect transmission and revelation of information.⁵ However, due to waiting cost, the sum of agents' expected utility, i.e. overall welfare, is lower compared to exogenous ordering.

SgROI (2003) presented the first experimental study implementing endogenous timing, adding non-informative signals to the seminal urn game by Anderson and Holt (1997). Facing constant waiting cost, subjects have 15 periods to pick an urn, and face a trade-off between waiting cost and potentially better predictions through the observation of prior decisions. Subjects receiving informative signals optimally decide in the first period to avoid waiting cost, while subjects receiving non-informative signals rationally decide immediately afterwards, using public information. As subjects' ordering works fairly well in SgROI (2003), overall informational efficiency should be close to optimal. Nonetheless, normal and reversed cascades continue to occur, and thus no perfect revelation of information is achieved.

In Ziegelmeyer et al. (2005), two subjects receive an integer signal as a realization of a random variable. Both subjects are asked to assess whether the sum of both signals is either positive or negative and face constant waiting cost. Both subjects are able to anticipate the strength of the other's signal depending on the respective period of decision, which should lead to information efficiency. However, subjects deviate from rational behavior by acting too early according to their signals, which in turn reduces delay costs. The authors interpret this as an internalization of informational externalities to reduce welfare-damaging delay.

In Çelen and Hyndman (2012), two subjects make a binary choice between an investment with fixed pay-off and a risky alternative with an unknown pay-off, and have 3 periods to take a decision. Decisions for the non-risky investment are reversible, while the choice of the risky option is not. Additionally, subjects receive private information on the actual payoff of the risky alternative. Subjects delay their decision in order to gather additional information, particularly when their signal does not favor the risky investment. Excessive waiting is partly explained by risk aversion when the accuracy of the private signal is low.

⁵ What drives this result is the continuous action set. As emphasized by Lee (1993), since the continuous action set allows perfect transmission of private information, informational cascades become fully informational revealing and asymptotically converge to the optimal decision.

Based on the model by Levin and Peck (2008), Ivanov et al. (2013) ask subjects to decide in discrete time whether to invest, not to invest or wait and decide later. Once a subject takes a decision, it becomes public information. Contrary to previous experiments, subjects receive two kinds of private information: a private signal about the return and a private signal concerning the cost of investment. Subjects generally use information correctly, yet deviate from rational timing.

Following Ivanov et al. (2013), it is well established that individuals fail to apply rational timing. Therefore, we refrain from investigating deviations from optimal timing of decision and take a different angle on the timing of choices. While we implement a two-player game with a simple rational strategy for predictions in any given situation, we adjust costs of delay and accuracy rewards so there is no NE for the timing of decisions. Being second mover is always preferred unconditional on signal strength, yet one could end up bearing waiting cost and still be first mover as the time horizon is finite and the other player might act symmetrically. Accordingly, the timing of a choice does not perfectly convey private information and simultaneous decisions that preclude observational learning are possible. We thus implement a multi-dimensional decision situation that resembles a gamble on gathering additional information by strategically delaying decisions. We argue that this resembles real world decisions and renders our comparison of fixed-order efficiency and endogenous ordering more interesting in terms of external validity compared to studies that investigate deviations from rational timing.

Implementing this concept, we find that endogenous timing on average increases the degree of rationality of predictions and thus their accuracy when compared to an exogenous setting. This increase is smaller when higher waiting cost are implemented, which leads to earlier and often simultaneous decisions that preclude observational learning. Therefore, subjects are sensitive to changes in waiting cost. In addition to observational learning, first movers correctly infer signal strength from the waiting time of the co-player when waiting cost are designed in a simple way, which adds to the overall increase in prediction accuracy for the endogenous setting. However, the gains in informational efficiency are compensated by waiting cost, resulting in no positive net welfare effect of endogenous timing. Our results suggest that there are no positive welfare effects of introducing endogenous rather than exogenous ordering, yet improvements in overall informational efficiency. We find neither excessive waiting, i.e. waiting when no additional information can be obtained, nor accuracy maximizing behavior, i.e. waiting to always become second mover. In turn, there are many

subjects minimizing waiting cost by making their prediction in the very first period, thereby passing on the opportunity of observational learning.

The remainder of this paper is structured as follows. Section 2 derives our theoretical framework, while section 3 contains our experimental design. Section 4 presents the results and section 5 concludes.

2. THEORETICAL FRAMEWORK

To structure our analyses, we present a basic framework of rational predictions. It applies to both the benchmark experiment with a randomly fixed decision order (*Exp1*) and to *Exp2*, which implements endogenous ordering of choices comprising three treatments (*high cost*, *low cost*, *signal dependent*), thus varying the implementation of waiting cost. Subsequently, we present considerations on the individual timing of decisions.

In both experiments, two players $i = 1, 2$ are randomly matched and participate in seven repetitions ($r = 1, 2, \dots, 7$ denoted as “projects”) of a non-cooperative game. We denominate the respective other player as the co-player. Both players are asked to predict the value of a project W in discrete time periods $t=1, \dots, T$. Both players receive private information $m_{i,r}$, which are independent realizations of a uniformly distributed random variable $M \in [1, 100]$. $W \in [2, 200]$ is the sum of private information.⁶ W 's realizations are denoted as $w_r = m_{1,r} + m_{2,r}$. Second movers can observe prior predictions. Following every project, the actual value of W , the two predictions and respective payoffs are shown to players.

We denote $z_{i,r}$ as the prediction of subject i in project r . Players are rewarded according to the absolute accuracy $d_{i,r} = |w_r - z_{i,r}|$ of their prediction. To make the payoffs more accessible to participants, we define fixed payoff intervals. Players receive 2000 ECU for a deviation $d_i \leq 5$, 1600 ECU for $6 \leq d_i \leq 10$, 1200 ECU for $11 \leq d_i \leq 15$, 800 ECU for $16 \leq d_i \leq 20$, 400 ECU for $21 \leq d_i \leq 25$ and 0 ECU for $d_i > 25$. There are no pay-off externalities.

For both experiments, a rational prediction is deducted as follows.

⁶ This basic structure is used in the model of Gul and Lundholm (1995) and was experimentally established by Ziegelmeyer (2005). Çelen and Kariv (2004) implement a similar structure of continuous signals and discrete action spaces into the seminal Anderson and Holt (1997) urn experiment.

2.1 PREDICTIONS

Let $t \in \{1,2\}$ denote the position in the decision order, i.e. $t = 1$ identifies the first mover. Since the first mover is uninformed about the other player's signal m_{-i} , the optimal prediction $z^*_{t=1}$ equals the sum of the private information m_i and the expected value of m_{-i} :

$$(1) \quad z^*_{t=1} = m_i + E(m_{-i})$$

In the exogenous ordering case the expected value of the second mover's signal $E(m_{-i})$ is equal to $E(M)=50.5$ (henceforth 50.5 is rounded to 50). The same is true for the endogenous case, when the timing of decisions does not depend on signal strength. In this case first movers cannot use the timing of the second mover as valuable information.

Subsequently, the second mover can perfectly infer the first mover's private information m_i . Subtracting the expected value $E(m_{-i})$ from the observed prediction $z^*_{t=1}$ yields the private information of the first mover $m_i = [z^*_{t=1} - E(m_{-i})]$. Thus, the optimal prediction of the second mover is given by:

$$(2) \quad z^*_{t=2} = m_1 + m_2 = w$$

The private information of the first mover is thus perfectly transmitted to the second mover.

Essentially, a rational first mover expects the co-player's signal to be 50 in the case of exogenous ordering or endogenous ordering without anticipation. Adding 50 points to her private signal gives the optimal prediction, which yields on average an absolute deviation d_i of 25 points and an average payoff of 400ECU. The second mover is aware of this strategy, and thus derives the first mover's signal by subtracting 50 from her prediction. This eliminates the first mover's deviation, and consequently earns the second mover 2000ECU for a correct prediction. Applying these rules fully describes rational behavior in the exogenous case. However, in the endogenous case players have to choose when to act, which makes the definition of optimal behavior more complex.

2.2 TIMING

The endogenous game (*Exp2*) is a non-cooperative waiting game. Every project comprise five successive periods in which players decide to predict or wait. Once a prediction is made, its value is shown to the respective co-player in the next period. A project ends once both predictions are made or the five periods elapse. Let $t_i \in \{1,2,3,4,5\}$ now denote the chosen decision period of individual i in the respective project r .

We implement a trade-off between an early prediction with low waiting cost and a delayed, yet potentially more accurate prediction with increased waiting cost for all treatments of

Exp2. Waiting cost are presented to subjects as the reduction of a time bonus $B_{i,r}$ to avoid triggering loss aversion. Universally, we implement $B_{i,r}(t_{i,r},x_r)=[(12-2t_{i,r})x_r/\alpha]$, where x_r is equal to w_r for the *low cost*, and *high cost* treatments and equal to $m_{i,r}$ for the *signal dependent* treatment. For *low cost* and *signal dependent* the weighting factor α is 2, while for *high cost*, α is 1. In all periods, subjects are informed about the time bonus for the respective round.⁷ For all treatments, waiting cost and signal strength are positively correlated in a linear way, and thus marginal waiting cost are constant. Compared to *high cost*, *low cost* reduces the weight of the time bonus by fifty percent in contrast to the accuracy bonus; the *signal dependent* treatment eliminates uncertainty concerning the exact value of waiting cost whereby the weight of the time bonus is between the two other treatments. Therefore, the treatments of *Exp2* allow checking for sensitivity towards the level and specific implementation of waiting cost.

Following the seminal experiment by SgROI (2003), the experimental studies on endogenous timing implement a threshold of signal strength, making it optimal to decide in a specific period. This enables a comparison of optimal and actual timing of decisions. In contrast, we implement a combination of signal strength and waiting cost that leads to an incentive to outwait the co-player for any given signal strength in any treatment. Consider for *high cost*, the highest possible private signal of $m_i = 100$. The maximum waiting cost is incurred when the decision is delayed to the last period ($t_i = 5$), whereby the player foregoes a time bonus of 1200 ECU. Given that the average premium for the second mover is 1600 ECU, as shown above, players always have an incentive to become second movers. This would necessarily lead to a war of attrition situation (Ziegelmeyer et al., 2005), with both players deciding in the last possible round. However, if one player understands that this situation makes him a first mover anyways, it is preferable to decide in the first period in order to minimize the waiting cost. Subsequently, the co-player's best response would be a decision in the second period. Given a decision in the second period, the first player's best response is a decision in the third period. As this evidently leads back to the "war of attrition" situation, there is no NE. Note that this does not change the described optimal predictions conditional on the position in the order. If players predict simultaneously, the rationale for the first mover applies to both. While this setting obviously precludes an investigation of rational timing, it offers an insight into behavior under uncertainty with best responses being conditional on the co-player's (unpredictable) timing.

⁷ The time bonus is displayed to participants in an easily accessible form. For *low cost* and *high cost*, "10...2 * project value" is displayed, for *signal dependent*, subjects are shown the exact time bonus.

We argue that our setting resembles actual decision situations in the context of social learning in the sense that all players are confronted with a gamble, betting on becoming a second mover by the decision to wait.⁸ Second movers gain the opportunity to observe first movers' decisions, enabling more accurate predictions, i.e. higher payoffs. This leads to an optimization problem characterized by comparing constant waiting cost and potential yet uncertain gains due to higher prediction accuracy. The probability of becoming second mover is evidently not calculable when the co-player has not acted yet. However, the higher waiting costs are relative to potential gains in accuracy, the less attractive the gamble becomes. Thus, as higher signals are related to higher waiting cost, the gamble's attractiveness decreases with higher signal strength. In this respect despite the incentive to always outwait the other, timing might reveal private information. In this case first movers can derive valuable information from the timing of their co-player, a process we denote as *anticipation*. Delayed decisions might be related to weaker signals and $E(m_i)$ is now smaller than $E(M)$. Overall efficiency could increase compared to the exogenous case, since first movers are able to give more accurate predictions. Risk preferences might have a significant impact in this context, as risk-averse players might tend to decide early, rejecting the gamble in favor of a fixed time bonus. Since we have fixed matching and seven repetitions of the game, players might try to build up reputation by signaling the willingness to become second mover no matter the amount of waiting cost. However, it is never optimal to wait another period if the co-player has already decided, given that such a behavior only increases waiting cost and cannot reveal further information. We denominate this behavior as *excessive delay*. Furthermore, it is always optimal to predict in the last period of a project rather than not predicting at all. In sum, there is no well-defined optimal solution to the gamble since the probability of becoming second mover is not calculable. Consequently, we are not testing for optimal Bayesian updating, but compare the results from endogenous and exogenous ordering in a more realistic social learning environment.

⁸ Note that the matching of players across the seven repetitions of both experiments is held constant to enable reputation effects related to individual preferences. For instance, consider a risk-averse player who develops a reputation of always deciding first. Understood by the second player, this should have a profound impact on the overall results. We chose this setting since the opportunity to build up reputation effects is a central feature of actual social learning environments where the set of participants remains mostly constant.

3. EXPERIMENTAL DESIGN

We run two separate experiments in a between-subjects design with fixed matching. Our benchmark experiment (*Exp1*) has an exogenously determined decision order, while the second experiment (*Exp2*) implements endogenous ordering. The latter comprises three treatments (*high cost*, *low cost*, *signal dependent*) in which waiting cost are varied.

The experiments took place at the Laboratory for Behavioral Economics at the University of Göttingen in December 2012 with 228 Undergraduate students in different fields of study⁹ participating in 13 sessions (58 in *Exp1*, 56/58/56 in the treatments of *Exp2*). *Exp1* took 40 minutes on average, and *Exp2* 80 minutes. 1000 ECU converted to 1.10 Euro in the exogenous game, to 0.80 Euro in the *high cost* endogenous treatment, and to 1.20 Euro in the *low-cost* and the *signal dependent*. The varying of conversion factors across treatments is intended to keep the average payoff per hour constant. The average payoff in *Exp1* (*high cost/ low cost/ signal dependent*) amounts to 10.6 Euro (12.3/12.0/12.3) including a 2.50 Euro show-up fee. Participants were recruited using the online recruiting system ORSEE (Greiner, 2004) and were allowed to take part in only one session. The games were programmed and conducted with the software z-Tree (Fischbacher, 2007). See the Appendix for our instructions.

⁹ Participants had the following fields of study: humanities 11%, Law studies 12%, social sciences 14%, science 26%, economics and business administration 35%. The mean age was 23.7 years, and 55% of participants were female.

4. RESULTS

Our dataset consists of 228 participants making predictions in seven projects which provides us with 1596 observations in total. We analyze our results in three steps: first, we measure informational efficiency against the rational benchmark across our experiments and treatments; second, we present data on the extent of strategic delay in the endogenous game; and third, combining the changes in informational efficiency and costs of strategic delay, we assess the net impact of introducing endogenous ordering on social welfare.

4.1 INFORMATIONAL EFFICIENCY

We define informational efficiency as the overall accuracy of predictions. Given that predictions are observable, rational second movers should infer first movers' signals and predict correctly, which would imply an efficient internalization of the information externality. The more accurate predictions become overall, the more efficient is the transmission of information. Table 1 shows that subjects overall fail to predict accurately. Recall that the rational prediction would have first movers adding 50 points= $E(M)$ to their private signal and the second movers subtracting these 50 points from the observed prediction to perfectly infer the first movers' signals. Our results suggest that the average prediction error for first movers is significantly greater than 25, the value expected for rational players ($t=4.5887$, $p=.0000$ for *Exp1*; $t=2.3345$, $p=.0215$ for *high cost*; $t=2.2878$, $p=.0237$ for *low cost*; $t=2.1060$, $p=.0372$ for *signal dependent*; two-sided t-test). For second movers, we find the average prediction error to be significantly smaller for all treatments ($z=4.246$, $p=.0000$ for *Exp1*; $z=4.606$, $p=.0000$ for *high cost*; $z=6.146$, $p=.0000$ for *low cost*; $z=5.468$, $p=.0000$ for *signal dependent*; Wilcoxon signed-rank test for matched data), which indicates observational learning. However, subjects fail to reduce the absolute deviation to zero ($t=14.5937$, $p=.0000$ for *Exp1*; $t= 8.6657$, $p=.0000$ for *high cost*; $t=13.4155$, $p=.0000$ for *low cost*; $t=11.4616$, $p=.0000$ for *signal dependent*; two-sided t-test).

This specific failure could reflect the poor performance of first movers, which we analyze in more detail in subsection 4.1.2. Overall absolute deviation is by far greater than 12.5, the theoretical minimum ($t=13.1748$, $p=.0000$ for *Exp1*; $t=12.9355$, $p=.0000$ for *high cost*; $t=12.9987$, $p=.0000$ for *low cost*; $t=12.7777$, $p=.0000$ for *signal dependent*; two-sided t-test).¹⁰

¹⁰ The minimum average prediction error could be lower if timing reveals information, i.e. subjects conduct anticipation. However, decisions might be given simultaneously in *Exp2* in which case the minimum error amounts to 25. Given the share of decisions in the same period the theoretical minimum error without

Comparing the experiments, *Exp1* and *Exp2* differ, as second movers in *Exp2* show significantly higher prediction accuracy ($z=3.685$, $p=0.0002$; Mann-Whitney U-Test), and thus the overall efficiency is higher. There is no significant difference for first movers ($z=1.255$, $p=0.2094$; Mann-Whitney U-Test).

	<i>Exp1</i>		<i>Exp2</i> (<i>high cost</i>)		<i>Exp2</i> (<i>low cost</i>)		<i>Exp2</i> (<i>signal dependent</i>)	
	absolute deviation	decision period	absolute deviation	decision period	absolute deviation	decision period	absolute deviation	decision period
First mover	n=203 33.33 (25.86)	1	n=105 30.08 (22.28)	1.34 (.70)	n=137 28.57 (18.26)	1.59 (.97)	n=129 28.71 (19.98)	1.47 (.89)
Second mover	n=203 24.93 (24.34)	2	n=102 18.79 (21.90)	2.37 (.69)	n=135 15.85 (13.73)	2.61 (.92)	n=126 16.82 (16.47)	2.51 (.86)
Same period			n=182 31.35 (23.22)	1.3 (.96)	n=131 28.12 (18.75)	1.77 (1.43)	n=134 28.84 (18.39)	1.88 (1.49)
Overall	n=406 29.13 (25.43)	1.5	n=389 27.71 (23.20)	1.59 (.95)	n=403 24.16 (18.01)	1.99 (1.21)	n=389 24.90 (19.14)	1.95 (1.20)

Table 1: Informational Efficiency and Strategic Delay

Note: Values are averaged for the respective group; standard deviations presented in parentheses.

When comparing the endogenous treatments to *Exp1*, the data might indicate a better performance of first movers in *low cost* and *signal dependent*. The average absolute deviation in *Exp1* is about 17% (19%) higher compared to *low cost* (*signal dependent*). Note, however, that there are no differences measured by the absolute deviations at the conventional levels of significance ($z=0.637$, $p=.5239$ for *high cost*; $z=1.361$, $p=.1734$ for *low cost*; $z=1.210$, $p=.2261$ for *signal dependent*; Mann-Whitney U-Test). For second movers, the average prediction error is significantly lower in contrast to *Exp1* ($z=2.190$, $p=.0285$ for *high cost*; $z=3.185$, $p=.0014$ for *low cost*; $z=2.960$, $p=.0031$ for *signal dependent*; Mann-Whitney U-Test). Even though there are no significant differences in first movers' performance across experiments, this result might partly be driven by the on average weaker performance of first movers in *Exp1*.

anticipation is 18.48 in *high cost*, 16.67 in *low cost* and 16.9 in *signal dependent*. Obviously, actual absolute average deviations far outreach these hypothetical values.

Despite the better performance of second movers, we only find a positive and significant net effect on informational efficiency in *low cost* ($z=0.339$, $p=.7346$ for *high cost*; $z=1.760$, $p=.0784$ for *low cost*; $z=1.468$, $p=.142$ for *signal dependent*; Mann-Whitney U-Test) when considering overall performance. This is clearly due to the large number of simultaneous decisions, which preclude observational learning. 37.56% of subjects in *Exp2* decided simultaneously (46.4% in *high cost*, 32.5% in *low cost*, 34.2% in *signal dependent*). Thus we conclude, that, on the one hand, endogenous ordering increases overall informational efficiency by improved observational learning. On the other, allowing for simultaneous decisions might fully compensate this effect or even reverse the results in an extreme scenario.

It is interesting to take a closer look on the distribution of rationality. Therefore, we define decisions as being rational if prediction value does not deviate from the optimal value derived by solving the equations in the theoretical framework by more than 5 points (equal to the width of each payoff interval) in absolute terms. If second mover i follows a non-optimal prediction, her decision is assumed to be rational if the prediction value falls into the interval $[m_i, m_i+100]$, which can be considered a rather lax criterion. The cumulative distribution of subjects giving rational decisions shows that around 25.9% (11.8%) decide rationally in one or less projects of *Exp1* (*Exp2*). Moreover, around 79% (56%) do not reach more than three rational predictions, while 3.5% give more than 4 rational predictions in *Exp1*, and 10% in *Exp2*. We also apply a Fisher exact test, which shows a statistically significant relationship between the number of rational predictions and the ordering regime (Fisher's exact=0.003). The differences between *Exp1* and *Exp2* regarding prediction accuracy are mirrored by differences in the number of rational predictions. Considering the relationship between the treatments of *Exp2* and the number of rational predictions, we obtain significant results (Fisher's exact=0.55).

Furthermore, we test for learning effects by running a Skillings-Mack (SM) test for the differences of the absolute prediction errors over projects. For all the treatments, we find no significant differences.¹¹ However, it is to note that we find some predictions showing a weak or even a misunderstanding of the game, i.e. subjects expecting the co-player's signal to be 0, negative or greater than 100. According to this definition there are 38 outliers (9.36% of all

¹¹ We apply a Skillings-Mack (SM) test since we have repeated measures given by the predictions of the participants and also missing values when participants did not give a prediction in a project. For *Exp1* we find (SM=9.933, $p(\text{no-ties})=0.1275$ and empirical $p(\text{ties})\sim 0.1280$); for *high cost* (SM=4.004, $p(\text{no-ties})=0.6761$ and empirical $p(\text{ties})\sim 0.658$); for *low cost* (SM=6.048, $p(\text{no-ties})=0.4178$ and empirical $p(\text{ties})\sim 0.421$) and for *signal dependent* (SM=4.504, $p(\text{no-ties})=0.06088$ and empirical $p(\text{ties})\sim 0.585$).

predictions) in *Exp1*; 15.52% in the first project, 10.34% in the second and 7.93% in later projects, a decrease that might point to some learning effects. For the treatments of *Exp2*, we find fewer predictions characterized by a weak understanding of the game (4.34% of all predictions for *high cost*, 3.2% in *low cost* and 2.55% in *signal dependent*).

To summarize, we find increases in informational efficiency for endogenous timing when decisions are not taken simultaneously. Interestingly, our results point to a better performance of first movers for *low cost* and *signal dependent*, though differences are not significant. This result might be understood as evidence for anticipation effects that require both players in a decision pair to follow a similar strategy where timing depends on signal strength. However, this intuition is somewhat misleading: we show in the next subsection that, besides some evidence for anticipation in *Exp2*, there is a systematic deviation from rational behavior of first movers driving the results in *Exp1*.

After considering first movers' behavior, we turn to the second movers to investigate whether weak observational learning and differences between the experiments merely reflect the poorer performance of first movers.

4.1.1 FIRST MOVER PERFORMANCE

In order to understand the performance of first movers in detail, we test whether expectations of the co-player's signals are derived rationally as proposed by Eq.1. We find that the average expectation (= prediction value – signal) of a co-player's signal for *Exp1* and *high cost* is 41.3 and 46.6 respectively and thus significantly smaller than the rational value of $50=E(M)$. For *low cost* (49.0) and *signal dependent* (48.9), there is no such significant deviation from 50 on average ($t=-4.5698$, $p=.0000$ for *Exp1*; $t=-2.4291$, $p=.0158$ for *high cost*; $t=-0.7831$, $p=.4342$ for *low cost*; $t=-0.9725$, $p=.3317$ for *signal dependent*; two-sided t-test).¹²

We check for anticipation effects as a way of improving prediction accuracy. We define anticipation as the systematic adjustment of predictions in response to the decision period of the co-player. If anticipation is present, the co-player's decision period should have a significant effect on prediction values. We find anticipation effects for *signal dependent* which can be best shown by considering Figure 1.

¹² These results are partially driven by the outliers mentioned; therefore, we tested for the differences after taking out these values. However, we find the same significant effects (average expectation for *Exp1* is 45.0 with $t=-3.7452$ and $p=.0002$; average expectation for *high cost* is 47.4 with $t=-2.3252$ and $p=.0208$; two-sided t-test).

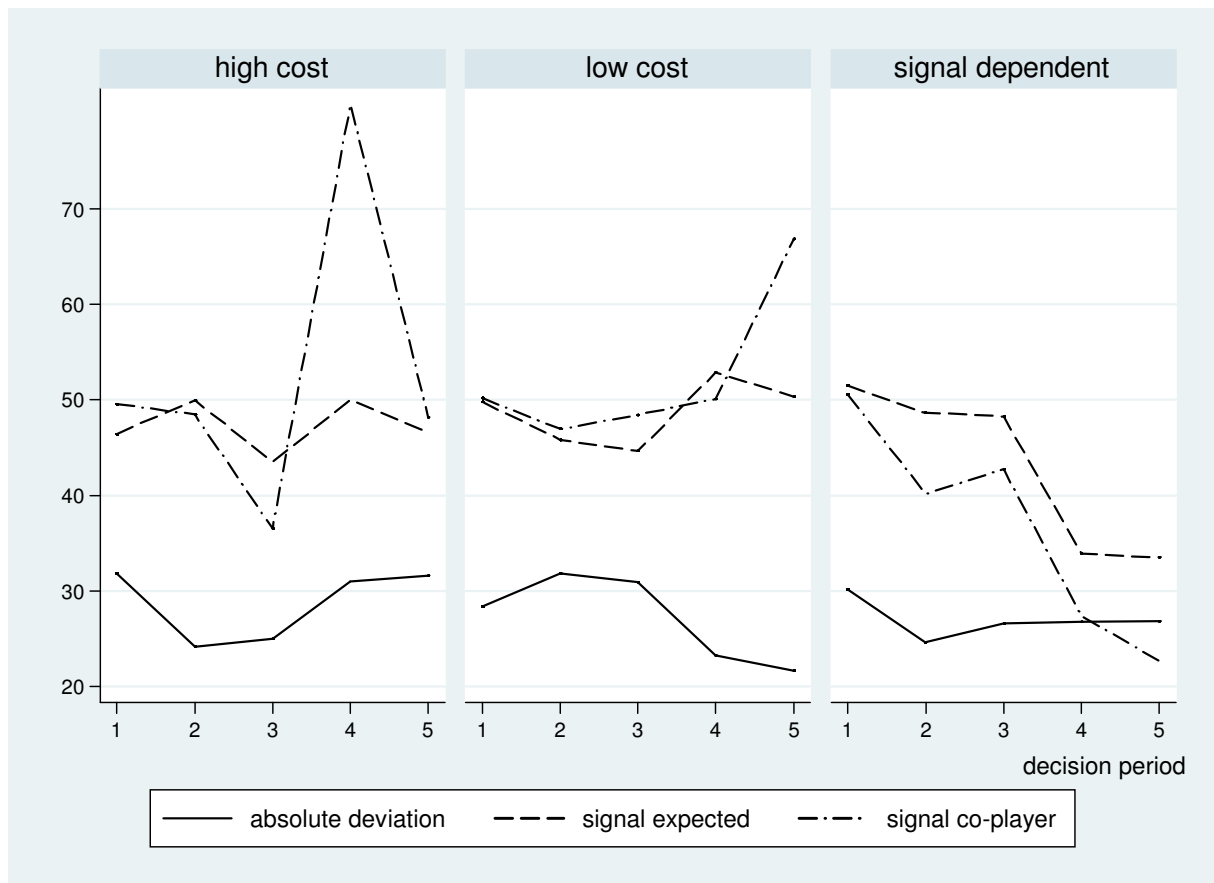


Figure 1: Performance of first movers

Figure 1 presents the development of the mean absolute deviation, the mean expectation of the co-player's signal and the actual mean of the co-player's signal over periods.

As the graph for *signal dependent* shows, first movers correctly expect decreasing signal strengths over decision periods. Simply assuming the co-player's signals to be 50 would yield higher prediction errors in later periods. To establish this result, we estimate a model with the randomly determined signals as the dependent variable and decision period as the explanatory variable for each treatment.¹³ Signal strength is estimated to decrease significantly by around 7.9 points per period in *signal dependent* ($t=-8.48$; $p=.0000$), with the respective constant amounting to 66.88 points.¹⁴ The basic requirement for anticipation is fulfilled, i.e. decision

¹³ We run a pooled OLS regression applying robust Driscoll and Kraay standard errors. Hence, we control for unobservable heterogeneity, heteroskedasticity, serial correlation in the idiosyncratic errors of order (2) and cross-sectional dependence.

¹⁴ For *high cost* the marginal effect of an additional period is estimated to reduce signal strength by -2.9 with a constant of 56.6. The effects is not significant ($t=-1.65$; $p=.151$). For *low cost* the coefficient is 2.55 and significant at the 5% level ($t=2.85$; $p=.021$), while the constant is 50. Applying fixed effects procedure also yields a significant decrease of signal strength estimated to be -6.56 points per period in *high cost*. For the other treatments results change only slightly.

periods reveal information about signal strength. To check whether this additional information is used, we turn to the player's expectations of their co-player's signal conditional on the decision period. Expectations of the co-player's signal are estimated to decrease significantly by 4.3 points per period ($t=-5.52$; $p=.001$) in *signal dependent*; the regression shows a constant of 56 points.¹⁵ Mirrored by decreasing prediction values over periods, anticipation thus improves, though not perfectly, prediction accuracy compared to the routine of expecting 50 for the co-player's signal in *signal dependent*. It remains unclear whether anticipation only occurs due to the less complex design of waiting cost in *signal dependent* or the weighing of time bonus in between *low* and *high cost*. However, we do not find anticipation for *low cost* and *high cost*. Signal strength slightly increases over periods in *low cost*, whereas expectations on co-player's signals do not change. In *high cost* signal strength significantly decreases over decision periods; however, again the additional information is not reflected in player's expectations.

We conclude that the better performance of first movers in *low cost* and *signal dependent* can be best explained by the occurrence of a systematic downward bias of expectations by first movers in *Exp1* and *high cost*. We additionally find anticipation for *signal dependent*, which improves prediction accuracy compared to the rational routine of expecting 50 as the co-player's signal.

4.1.2 SECOND MOVER PERFORMANCE

We now turn to the question of whether the poor performance of first movers is responsible for inefficient observational learning. Therefore, analogously to the previous analysis for the first movers, we consider the optimal routine described in Eq.2 in our theoretical framework. Recall that the optimal response for second movers when assuming rational behavior and no anticipation of first movers is given by $z^*_{t=2} = [z^*_{t=1} - E(M)] + m_i = m_1 + m_2 = w$; whereby t indicates the position in the decision sequence. We have already shown that predictions on average are not perfect. However, perfect predictions by second movers can only be obtained when first movers follow the rational routine described in Eq.1. Accordingly, we only consider observations for second movers that follow potentially rational predictions by the respective first movers. This applies to decisions following predictions that are equal or greater than 51 points ($=\min\{M\} + E(M)$) and smaller or equal than 151 points ($=\max\{M\} +$

¹⁵ The regression procedure is implemented as before. For *high cost*, we get a marginal effect of .025 points ($t=-0.02$; $p=.981$) and for *low cost* the marginal effect is -.133 points ($t=-0.25$; $p=.807$), thus both are not significant. Using fixed effect procedure does not change coefficients substantially and significances remain the same.

$E(M)$).¹⁶ Expectations of second movers should be characterized as perfectly revealing the first mover's signals. We thus test whether $[z^*_{t=1} - E(M)]$ equals the expectation of the second movers on average. For *Exp1 (high cost/low cost/signal dependent)*, we find the absolute deviation of the second movers' expectations from the optimal expectation amounting to 22.94 (12.19/11.46/13.86) on average and thus to be significantly greater than zero ($p=.0000$ for all treatments; two-sided t-test). The deviation from the optimal expectation is significantly higher for *Exp1* in contrast to the treatments of *Exp2* ($z=4.453$, $p=.0000$ for *high cost*; $z=4.853$, $p=.0000$ for *low cost*; $z=3.220$, $p=.0013$ for *signal dependent*; Mann-Whitney U-Test).

These results might be too pessimistic for *signal dependent*, since we have shown that for the first movers expectations of the co-player's signal are significantly correlated with decision periods. Second movers could adjust their calculation of first movers' signals respectively, thus outperforming the rule of Eq.2. As shown above, first movers reduce their expectation toward the co-player's signal by a rate of 4.3 points per period in the *signal dependent*. Therefore, we check whether second movers account for that systematic adjustment. We run a regression implementing the same routine as before of the second mover's expectation on decision period. We find that second movers are somewhat able to adjust for first movers' anticipation by reducing their own expectation by a rate of 6.2 points per period ($t=-6.36$; $p=.001$) with a constant of 69.6 in *signal dependent*.¹⁷

However, calculating the average of the absolute deviations of second mover expectations from realized first mover signals, we find a deviation of 25.9 for *Exp1*, 20.3 for *high cost*, 16.6 for *low cost* and 17.7 for *signal dependent*. Again, this calculation only considers decisions taken after potentially rational predictions of first movers. If second movers had strictly followed the rational routine $[z^*_{t=1} - E(M)]$ to guess the first mover's signal, the

¹⁶ Note that, for the analysis of second movers, relying solely on decisions following potentially rational predictions might be problematic. Players in a decision pair might have observed earlier irrational decisions thus causing project-interdependent assumptions regarding the co-player's behavior. However, as this keeps the analysis simple and our results in this section are very robust, we refrain from integrating project-interdependent effects.

¹⁷ Although we showed that expectations of first movers in *the high cost* are not related to the decision period, second movers adjust their expectation significantly by a rate of -8.8 points ($t=-3.72$; $p=0.01$). Results for *low cost* show an insignificant marginal effect of 1.5 points ($t=1.01$; $p=0.352$) for decision period on the second mover's expectation. The results do not change substantially when we only use decisions of second movers following potentially rational decisions of first movers or when we use a fixed effects procedure.

difference to realized signals for *Exp1* would be reduced to 14.4, 13.1 for *high cost*, 13.9 for *low cost* and 13.7 for *signal dependent*.

We conclude that the non-optimal performance of second movers does not result from the poor performance of first movers; rather, it is a source of inefficiency in itself. This effect is strongest for *Exp1*, thus second movers in *Exp2* treatments perform better on average by a more efficient observational learning. Consequently, c.p. informational efficiency tends to be higher when timing is endogenous.

We essentially see two driving forces that improve informational efficiency when ordering is endogenous. Firstly, we would argue that there is a self-selection conditional on the understanding of the mechanism of observational learning. The rational routine for second movers is somewhat more complicated to understand, as one has to comprehend the expectations and potential anticipation of the preceding player. Thus, players with a deeper level of reasoning tend to decide later and more frequently achieve observational learning.¹⁸

While this might explain the superior performance of second movers in *Exp2*, it does not explain equal or in tendency even higher levels of rationality for first movers. On the contrary, if players with a deeper understanding tend to decide as second movers, first movers should perform even worse due to self-selection in *Exp2* as compared to *Exp1*. Therefore, we would argue that the level of understanding hinges on the structure of the decision situation. Having subjects decide when to act induces considerations about the advantages and disadvantages of being the first and second mover. By inducing these reflections about the game itself, subjects are more likely to realize the relevance of the co-player's signal. These considerations might add to the level of understanding for both the first and second movers in case of endogenous ordering, thereby eliciting more rational behavior overall.

¹⁸ Therefore, one might consider a level-k approach to define an appropriate model of behavior in our experiment (Nagel, 1995; Stahl and Wilson, 1994, 1995; Crawford and Iriberry, 2007), thus rationalizing predictions conditional on the first mover's assumed depth of reasoning. However, even interdependent expectations regarding the depth of reasoning of the co-player might not explain the differences in informational efficiency between *Exp1* and *Exp2*. However, an extensive analysis regarding the expectation on the depth of reasoning is beyond the scope of this paper.

4.2 STRATEGIC DELAY

Strategic delay is the central feature of *Exp2* (Table 1 also includes the average decision periods). While an average decision period of 1.5 is predetermined for *Exp1*, it is 1.84 for pooled data of *Exp2*. The subjects react sensitive towards waiting cost: in *high cost*, decisions are taken earlier (1.58) compared to *low cost* (1.99) and *signal dependent* (1.94). Only 17 predictions are not given in the subsequent period after a co-player has decided, and thus excessive delay only occurs in 2% of all projects.

On the individual level, distinct strategies regarding timing can be revealed. There are around 16% of participants in *Exp2* always predicting in the first round (26.9% in *high cost*, 12.1% in *low cost* and 9% in *signal dependent*). These participants minimize waiting cost without trying to gain additional information by outwaiting the co-player. In 81 of the 91 projects with simultaneous decisions in *high cost*, predictions are given in the very first period; 46 of 66 in *low cost* and 45 of 67 in *signal dependent*.

In turn, only 4.1% of participants always predict as second mover or in the last period (3.57% in *high cost*, 5.17% in *low cost* and 3.57% in *signal dependent*). Thus, only few try to maximize accuracy regardless of signal strength or waiting cost.

Combining the analyses on informational efficiency and strategic delay, we now investigate the overall effects on social welfare.

4.3 SOCIAL WELFARE

We measure social welfare by aggregate payoffs. To assess welfare effects, we calculate average waiting cost and accuracy bonuses. Given that waiting costs are manipulated over treatments of *Exp2*, we calculate hypothetical waiting cost for *Exp1* according to the payoff structure of the respective *Exp2* treatment in order to enable a comparison.

In *Exp1* subjects earn on average 708ECU per projects as accuracy reward compared to slightly higher 719ECU in the *high cost*. Subjects take decisions relatively early in the *high cost* yielding an average time bonus of 886ECU, which is equivalent to average waiting cost of 117ECU. Applying the same waiting cost structure to *Exp1* gives an average time bonus of 924ECU or average waiting cost of 103ECU. Thus, the total expected payoff in a project of *Exp1* is 1632ECU and 1605ECU for the *high cost*. On aggregate for the whole game (7 projects), this gives a difference of 189ECU or 21 Cent, which is around 2.1% of the average payoff (excluding show-up fee). Following the same procedure leads to a 126ECU (14Cent) lower payoff for the *low cost* game aggregated for seven projects. The *signal dependent* game has a lower payoff of 21ECU (2 Cent). Obviously, these differences are of low relevance.

Another way of looking at this result is to exclude same period decisions, as there might be several real-world situations where simultaneous decisions are highly unlikely, e.g. high frequency trading in financial markets. When excluding simultaneous decisions from calculating the averages, effects on social welfare do not change substantially. The largest treatment effect is given for the *signal dependent*, in which the average increase in total payoff amounts to 530ECU, which convert to 58 Cent or 5.9% of average payoff (for *high cost* difference amounts to -13ECU and for *low cost* to 25ECU).

Overall, gains in informational efficiency are realized at the expense of increased waiting cost, such that no relevant effects on social welfare are elicited by introducing endogenous ordering. In the given gambling structure, waiting cost and informational efficiency turn out to be strongly interdependent, causing the absence of net effects on social welfare. Placing different relative weights on waiting cost shows no influence on social welfare, given that participants adjust their timing of decisions accordingly.

5. CONCLUDING REMARKS

The present study investigates informational efficiency in a game of social learning, comparing exogenous and endogenous ordering of choices. By quantifying the effect of observational learning and waiting cost, we show the welfare effects of these different regimes of ordering. Based on the model by Gul and Lundholm (1995), we run a two-player prediction game with a benchmark treatment of exogenous ordering and three treatments of endogenous ordering. Rather than the classic binary action sets following the seminal Anderson and Holt (1997) paper, we introduce a continuous action space to more precisely determine the success of observational learning. We refrain from implementing an optimal timing conditional on signal strength to expose subjects to a situation where gambling on the co-player's uncertain action is required. We argue that both the continuous action space and gambling situation that our subjects faced depict actual decisions in social learning environments more closely than the informational cascade games characterized by binary decisions and exogenous ordering.

In our treatments, endogenous timing enhances the rationality of predictions and thus their accuracy, yet also leads to higher waiting cost. Subjects react sensitively to changes in waiting cost and adjust their timing accordingly. This leads to earlier and often simultaneous decisions that inhibit observational learning. For lower waiting cost, subjects tend to wait longer, which fosters observational learning, yet increases waiting cost to the same degree. Thus, there are no overall positive welfare effects in our endogenous treatments. However, despite the specific incentive to always outwait the co-player, we rarely find war of attrition situations that would massively reduce welfare. We suggest that making subjects take a timing decision in the endogenous game fosters a deeper level of reasoning in general, which leads to a more efficient observational learning. Additionally, observational learning might be improved by a self-selection according to the understanding of second mover advantages. Our results show that introducing an endogenous rather than exogenous ordering regime leads to higher informational efficiency but does not increase overall social welfare.

We add to the literature on social learning by introducing an experiment that enables comparison between exogenous and endogenous ordering of choices. This allows us to combine the discussions following the seminal urn experiment by Anderson and Holt (1997) with the studies on endogenous ordering following SgROI (2003). Both strands of literature investigate the success of social learning and informational efficiency, yet fail to compare the two settings. We qualify the extent of informational efficiency in a unitary setting across regimes of ordering. While informational efficiency is effectively increased with the

introduction of endogenous ordering, as suggested by previous studies, we cannot conclude that this leads to a positive effect on social welfare. However, it also does not deteriorate welfare altogether, as situations with extreme waiting cost are rare. Our results suggest that social learning is fairly effective when implementing a continuous action space and endogenous timing. Therefore, the informational inefficiency in situations of rational herding emphasized by numerous studies is limited to specific decision situations and should not be generalized.

ACKNOWLEDGEMENTS

The authors are grateful to Kilian Bizer and Markus Spiwoks for their helpful comments and to Simon Schütz who provided excellent student assistance in programming the experiment.

REFERENCES

- Alevy, J.E., Haigh, M.S., List, J.A., 2007. Information Cascades: Evidence from a Field Experiment with Financial Market Professionals. *Journal of Finance* 62 (1), 151-80.
- Anderson, L.R. Holt, C.A., 1997. Information Cascades in the Laboratory. *The American Economic Review* 87 (5), 847-862.
- Anderson, L.R., 2001. Payoff Effects in Information Cascade Experiments. *Economic Inquiry* 39 (4), 609-615.
- Banerjee, A.V., 1992. A simple model of herd behavior. *Quarterly Journal of Economics* 107 (3), 797-817.
- Bikhchandani, S., Hirshleifer, D., Welch, I., 1992. A theory of fads, fashion, custom, and cultural change in informational cascades. *Journal of Political Economy* 100 (5), 992-1026.
- Bikhchandani, S., Hirshleifer, D., Welch, I., 1998. Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades. *The Journal of Economic Perspectives* 12 (3), 151-170.
- Çelen, B., Hyndman, K., 2012. An experiment of social learning with endogenous Timing. *Review of Economic Design* 16 (2-3), 251-268.
- Çelen, B., Kariv, S., 2004. Distinguishing Informational Cascades from Herd Behavior in the Laboratory. *The American Economic Review* 94 (3), 484-498.
- Chamley, C., 2004. *Rational Herds: Economic Models of Social Learning*. Cambridge: Cambridge University Press.
- Chamley, C., Gale, D., 1994. Information revelation and strategic delay in a model of investment. *Econometrica* 62 (5), 1065-1085.
- Cipriani, M., Guarino, A., 2005. Herd Behavior in a Laboratory Financial Market. *The American Economic Review* 95 (5), 1427-1443.
- Crawford, V.P., Iriberry, N., 2007. Level-k auctions: can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions?. *Econometrica* 75 (6), 1721-1770.

- Dominitz, J., Hung, A.A., 2009. Empirical Models of Discrete Choice and Belief Updating in Observational Learning Experiments. *Journal of Economic Behavior & Organization* 69 (2), 94-109.
- Drehmann, M., Oechssler, J., Roider, A., 2005. Herding and Contrarian Behavior in Financial Markets: An Internet Experiment. *The American Economic Review* 95 (5), 1403-1426.
- Fahr, R., Irlenbusch, B., 2011. Who follows the crowd – Groups or individuals?. *Journal of Economic Behavior & Organization* 80 (1), 200-209.
- Fischbacher, U., 2007. Z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2), 171-178.
- Frisell, L., 2003. On the Interplay of Informational Spillovers and Payoff Externalities. *The RAND Journal of Economics* 34 (3), 582-592.
- Goeree, J.K., Palfrey, T.R., Rogers, B.W., McKelvey, R.D., 2007. Self-correcting information cascades. *Review of Economic Studies* 74 (3), 733-762.
- Greiner, B., 2004. An online recruitment system for economic experiments. *GWDG Berichte* 63, 79-93.
- Gul, F., Lundholm, R., 1995. Endogenous Timing and the Clustering of Agents' Decisions. *Journal of Political Economy* 103(5), 1039-1066.
- Hung, A.A., Plott, C.R., 2001. Information Cascades: Replication and an Extension to Majority Rule and Conformity-Rewarding Institutions. *The American Economic Review* 91 (5), 1508-1520.
- Ivanov, A., Levin, D., Peck, J., 2013. Behavioral biases in endogenous-timing herding games: An experimental study. *Journal of Economic Behavior & Organization* 87, 25-34.
- Kübler, D., Weizsäcker, G., 2004. Limited Depth of Reasoning and Failure of Cascade Formation in the Laboratory. *Review of Economic Studies* 71 (2), 425-441.
- Lee, I. H., 1993. On the Convergence of Informational Cascades. *Journal of Economic Theory* 61 (2), 395-411.
- Levin, D., Peck, J., 2008. Investment dynamics with common and private values. *Journal of Economic Theory* 143 (1), 114-139.
- Nagel, R., 1995. Unraveling in Guessing Games: An Experimental Study. *The American Economic Review* 85 (5), 1313-1326.

- Nöth, M., Weber, M., 2003. Information Aggregation with Random Ordering: Cascades and Overconfidence. *The Economic Journal* 113 (484), 166-189.
- Oberhammer, C., Stiehler, A., 2003. Does Cascade Behavior in Information Cascades Reflect Bayesian Updating?. Max Planck Institute of Economics Strategic Interaction Group Discussion Paper No. 2003-01.
- SgROI, D., 2003. The right choice at the right time: A herding experiment in endogenous time. *Experimental Economics* 6 (2), 159-180.
- Stahl, D.O., Wilson, P.W., 1994. Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization* 25 (3), 309-327.
- Stahl, D.O., Wilson, P.W., 1995. On Players' Models of Other Players: Theory and Experimental Evidence. *Games and Economic Behavior* 10 (1), 218-254.
- Weizsäcker, G., 2010. Do we follow others when we should? A simple test of rational expectations. *The American Economic Review* 100 (5), 2340-2360.
- Willinger, M., Ziegelmeyer, A., 1998. Are More Informed Agents Able to Shatter Information Cascades in the Lab?, in Cohendet, P., Llerena, P., Stahn, H., Umbhauer, G. (Eds.), *The Economics of Networks: Interaction and Behaviours*. Berlin, Heidelberg: Springer, 291-305.
- Zhang, J., 1997. Strategic Delay and the Onset of Investment Cascades. *The RAND Journal of Economics* 28 (1), 188-205.
- Ziegelmeyer, A., Bracht, J., Koessler, F., Winter, E., 2008. Fragility of Information Cascades: An Experimental Study Using Elicited Beliefs. Max Planck Institute of Economics Strategic Interaction Group Discussion Paper No. 2008-94.
- Ziegelmeyer, A., My, K.B., Vergnaud, J.C., Willinger, M., 2005. Strategic Delay and Rational Imitation in the Laboratory. Max Planck Institute of Economics Strategic Interaction Group Discussion Paper No. 2005-35.

APPENDIX

Instructions for Experiment 1

The Game

In this game you and a co-player will estimate the value of a project. The value of the project consists of two parts: your own information and your co-player's information.

Your information and the information of your co-player are randomly determined numbers between 1 and 100. Therefore, project value that you have to estimate is always between 2 and 200. All of the possible information is equally likely.

Your Information (1 to 100)	+	Co-player's Information (1 to 100)	=	Project value (2 to 200)
Example: 25	+	50	=	75

There are 7 projects in which you will estimate the project value. In every project, it will be randomly determined if you or your co-player will give the estimation first. The first estimation is always displayed to the other player. Once both players have made their estimation, the next project begins.

You will have the same co-player in all projects. You have a maximum of one minute for each estimation. If you do not type in an estimation in time, you will not receive a payoff for this project!

The payoff

You will receive a precision bonus in every project, which depends on how precise your estimation was. The precision bonus depends on the deviation of your estimation from the correct project value. 1000 ECU equals a payoff of 1.70€. Additionally, you will receive an independent payoff of 2.50€. The following table clarifies the precision bonus:

Distance from correct project value	Precision bonus (in ECU)	Example: The project value is 100.	
		Your estimation was...	The precision bonus is...
0 – 5 points	2000	...96	...2000 ECU
6 – 10 points	1600	...109	...1600 ECU
11 – 15 points	1200	...87	...1200 ECU
16 – 20 points	800	...118	...800 ECU
21 – 25 points	400	...75	...400 ECU
from 26 points	0	...12	...0 ECU

Example

At the beginning of a project your information is 45. Therefore, you know that the project value is at least 45 plus the information of your co-player. Your co-player decides before you and estimates a project value of 120. You decide after him and estimate a project value of 105. **The correct project value is 95.** Thus, you receive a precision bonus of 1600 ECU, as your estimation deviated from the correct project value by 10 points.

Instructions for Experiment 2.

Note that the instructions refer to the *high cost* treatment. The differences from the other treatments are indicated as follows: information in square brackets corresponds to the *signal dependent* treatment, braces corresponds to the *low cost* treatment.

The Game

In this game you and a co-player will estimate the value of a project. The value of the project consists of two parts: your own information and your co-player's information.

Your information and the information of your co-player are randomly determined numbers between 1 and 100. Therefore, the project value that you have to estimate, is always between 2 and 200. All of the possible information is equally likely.

Your Information (1 to 100)	+	Co-player's Information (1 to 100)	=	Project value (2 to 200)
Example: 25	+	50	=	75

There are 7 projects in which you will give an estimation of the project value. All projects have 5 rounds of 2 minutes each. You must decide in which round you want to give your estimation.

All projects end once both players have given their estimation. Subsequently, the next project starts. You will have the same co-player in all projects. The following table provides an example of the course of the game:

project 1					project 2	
round 1	round 2	round 3	round 4	round 5	round 1	...
2 min.	2 Min.	2 min.	2 min.	2 min.	2 min.	...

At the beginning of each project, both players receive their information. Your co-player's information is unknown to you. You will have to decide in every round if you want to give an estimation (YES/NO). If you allow 2 minutes per round to elapse, you will not get a payoff for this project! If you choose NO, please wait for the next round of the project. If you choose YES, you will be told if your co-player will give an estimation in the same round. Subsequently, you will enter your estimation. Meanwhile, you will see an overview of the last rounds and, if applicable, the estimation of your co-player. If you decide before your co-

player, your estimation will also be shown to him. The following table exemplifies the course of the game and your possible actions:

	Round 1	Round 2	Round 3	Round 4	Round 5
Action by Player 1	NO	NO	NO	YES! Enters the estimation	Project completed!
Action by Player 2	NO	YES! Enters the estimation	Wait for the co-player...		

The payoff

The total payoff consists of two parts: the accuracy bonus (I.) and the time bonus (II.). For every round you wait with your estimation, your time bonus will be reduced. The precision bonus is higher, the closer your estimation gets to the correct project value. 1000 coins equal a payoff of 0.80€ {1.20€}, [1.20€]. Additionally, you will receive an independent payoff of 2.50€.

I. Precision bonus

You receive a bonus in every project which depends on the precision of your estimation, based upon its distance to the correct project value. The following table clarifies the precision bonus:

Distance from the correct project value	Precision bonus (in ECU)	Example: The project value is 100.	
		Your estimation was...	The precision bonus is...
0 – 5 points	2000	...96	...2000 coins
6 – 10 points	1600	...109	...1600 coins
11 – 15 points	1200	...87	...1200 coins
16 – 20 points	800	...118	...800 coins
21 – 25 points	400	...75	...400 coins
ab 26 points	0	...12	...0 coins

II. Time bonus

You receive a time bonus in every project, depending on the size of the project value {on the size of your information}. For every round you wait with your

estimation, your time bonus will be reduced. The following table clarifies the time bonus:

Estimation round	in Time bonus	Example: The project value is 100.	
		Estimation in round...	Time bonus...
1	10{5} x project value [10 x Information]	...1	...1000{500} ECU [1000 ECU]
2	8{4} x project value [8 x Information]	...2	...800{400} ECU [800 ECU]
3	6{3} x project value [6 x Information]	...3	...600{300} ECU [600 ECU]
4	4{2} x project value [4 x Information]	...4	...400{200} ECU [400 ECU]
5	2{1} x project value [2 x Information]	...5	...200{100} ECU [200 ECU]

Example:

At the beginning of a project, your information is 45. Therefore, you know that the project value is at least 45 plus the information of your co-player. Your co-player decides before you and estimates in round 3 that the project value is 120. You decide in round 4 and estimate that the project value is 105. **The correct project value is 95.** Therefore, you receive a time bonus of 380 (time bonus in round 4 = 4 x project value) { 190 (time bonus in round 4 = 2 x project value) } [180 (time bonus in round 4 = 4 x information)]. Additionally, you receive a precision bonus of 1600 coins, as your estimation deviates from the correct project value by 10 points.

Chapter VIII

THE IMPACT OF COMMUNICATION REGIMES ON GROUP RATIONALITY: EXPERIMENTAL EVIDENCE

with Till Proeger

Published:

Cege Discussion Papers, no. 185, Göttingen University, February
2014. Download: <http://www.uni-goettingen.de/de/60920.html>

ABSTRACT

The performance of groups has been thoroughly investigated in experimental economics, showing that groups are overall more rational deciders than individuals. However, superior group performance in economic experiments has primarily been shown for face-to-face decision making, which has ceased to be the prevalent form of communication in many IT-based organizations. To test the robustness of higher group rationality under conditions of virtual communication, we conduct a social learning experiment. We find that virtual communication leads to a substantial deterioration of group rationality for a judgmental task, while there is no effect for a purely intellectual task. Further, we show that higher cognitive abilities of group members have no impact for the judgmental task, yet increase rationality for the intellectual task. Our results have potential implications for the design of communication structures within decentralized organizations relying on virtual communication.

Keywords: cognitive abilities; communication; group composition; group performance; laboratory experiment; methodology; social learning

JEL Classification: C9; D8

HIGHLIGHTS

- We investigate the influence of communication regimes on group rationality.
- We compare face-to-face and chat-based decision making in a social learning game.
- For an intellectual task, rationality is independent of the communication regime.
- For a judgmental task, rationality deteriorates for virtual communication.
- Group member's cognitive abilities raise rationality in the intellectual task only.

1. INTRODUCTION

Experimental economics' understanding of group performance has advanced significantly in the past decade. Comparing individual and group results for the major experiments in behavioral economics, the core finding is that groups are more rational deciders overall (Kugler et al., 2012). By overcoming cognitive limitations, groups are shown to be "*less behavioral than individuals*" (Charness and Sutter, 2012, p. 159) and thus regularly closer to theoretically rational solutions. This is interpreted as a partial vindication of the assumptions on rational behavior for the prediction of real-world economic decision making.

By incorporating group cooperation in economic contexts, behavioral research more accurately reflects the widespread team decision making in actual organizations. However, economic experiments to date have given little consideration to the effects of different external restraints on group decision making. Specifically, as pointed out in Kugler et al.'s (2012) literature review, the impact of varying modes of communication is yet to be investigated, particularly for the case of virtual communication and personal group discussions. This specific issue has gained relevance with the fundamental shift in the nature of organizational decision making towards digital communication and decentralized decision making. An increasing share of team decisions in international organizations is taken in geographically dispersed locations, using means of virtual communication rather than personal meetings (Balliet, 2010). To date, economic experiments have not considered this aspect's potential consequences for group rationality.¹ Of the thirty-seven economic group experiments reviewed in Kugler et al. (2012), thirty implement direct face-to-face discussions or a no-communication voting procedure. While seven studies feature computer-mediated discussions, none focusses on potential differences from face-to-face interaction.

In contrast, empirical studies in organizational psychology and field studies on IT-management have covered the effects of virtual decision making, overall pointing to rather negative consequences for team performance, e.g. in the meta-study by Baltes et al. (2002). De Guinea et al.'s (2012) recent contribution points to overall negative effects, particularly in short-term teams. Mesmer-Magnus et al. (2011), using the extent of information sharing as a measure for group success, show that low levels of virtuality are supportive, while increased levels rather constrain effective information sharing. However, the derivation of clear

¹ In contrast, group cooperation in social dilemma situations with computer-mediated communication has been investigated, e.g. by Bicchieri and Lev-On (2007). Accordingly, the meta-study by Balliet (2010) finds face-to-face communication to be more effective in fostering group cooperation than virtual communication.

behavioral patterns remains difficult, due to the ubiquitous lack of clear theoretical benchmarks and financial incentives (Charness and Sutter, 2012). The same holds true for the broad results of small group research in social psychology (for a review see Laughlin, 2011), which suggests that groups are more successful if the demonstrability of the respective correct solutions is high (Laughlin et al., 2002). The performance of groups can thus be predicted by the respective task's position on a continuum from intellective to purely judgmental (Laughlin, 1980).² With demonstrability as the core influence on group performance, virtuality might reduce a group's ability to effectively convey information, leading to the deterioration of performance and thus reduce group rationality premiums.

We build on this notion to further economic small group research by investigating the robustness of higher group rationality when solutions have to be agreed upon by means of virtual communication. Therefore, we replicate a recent study on group performance in a social learning game that features personal communication presented by Fahr and Irlenbusch (2011, henceforth: F&I). All other factors constant, we have participants communicating and deciding via chat in two treatments, one featuring a purely intellective task and another that introduces a judgmental element. By contrasting both studies, we assess the impact of virtual communication on group rationality. Our secondary aim in this paper is to investigate the impact of group composition according to cognitive abilities on group rationality. Such group composition effects have not been considered to date in economic group research (Kugler et al., 2012), yet seem useful to give insight into reasons for group success. This might help to determine whether groups in economic contexts tend to perform like their most capable members or if the process of group discussions itself is pivotal for superior group performance by generating additional insight.

In comparison to the benchmark data from F&I, we find virtual communication to have a profound impact on decision quality. Virtual communication considerably reduces the rationality of group decisions for a judgmental task, while there is no effect for the purely intellective task. Secondly, higher cognitive abilities of group members have a positive effect for the intellective task and no effect for the judgmental task. Accordingly, even high cognitive abilities in groups fail to eliminate the detrimental effect of virtual communication in the judgmental task. Overall, in the case of judgmental tasks, we can conclude that groups rely on face-to-face group discussions to achieve superior rationality when compared to

² Intellective tasks possess a clear normative criterion to evaluate the quality of subjects' decisions, while judgmental tasks involve uncertainty about the actual quality of decisions (Cox and Hayne, 2006).

individuals. For a purely intellectual task, superior group performance holds regardless of the communication regime.

The remainder of this paper is organized as follows. In section 2, the experimental design is introduced, before section 3 presents our behavioral hypotheses. Section 4 describes our results and section 5 draws a conclusion.

2. EXPERIMENTAL DESIGN

Our experimental design replicates F&I, who built their group study on the urn game by Anderson and Holt (1997) and the respective seminal herding models by Banerjee (1992) and Bikhchandani et al. (1992). Using this experimental setting as a working horse yields the advantage of building on a large and well-established strand of literature on social learning, with Weizsäcker (2010) offering a comprehensive meta-study. As in the seminal game, two urns are considered, one containing two red and one blue marbles and the other two blue and one red. One urn is randomly selected prior to each period. Six subjects sequentially and in a randomized order guess which urn was selected. Along the decision sequence, all guesses become public information. Before taking a choice, each player is shown one marble, which is randomly drawn from the selected urn. At the end of every period, the correct urn is revealed. Rational decisions require correct Bayesian updating. We base our analysis on the basic benchmark model of rational behavior employed by the studies following Anderson and Holt (1997).³

There are two treatments: in the private information treatment (*PIT*), players observe all prior decisions taken in the sequence, as described above; in the full information treatment (*FIT*), players are instead shown the marbles drawn by previous players. As deduced by F&I, this setting allows for a distinction between judgmental and intellectual tasks. In *FIT*, subjects have to correctly apply a counting rule regarding prior drawn marbles, a considerably straightforward intellectual task. By contrast, in *PIT*, subjects additionally need to interpret the observed decisions with respect to the preceding players' decision quality. Thus, there is an additional element of uncertainty, which serves as a judgmental aspect to the otherwise intellectual task.

In addition to the six individual players, a group of three players is matched to one of the six individual players ("focal individual") during the entire game. Both the group and the focal

³ See e.g. Goeree et al. (2007) for a comprehensive theoretical analysis.

individual face identical situations in all periods. Unlike the focal individual's the group decision does not become public information. In each of the fifteen periods, groups are required to come up with a unitary solution. In both treatments, guessing the correct urn pays 1€ for each group member.

To closely replicate F&I, we implement the decision sequences generated in their treatments. We match groups of three players and have them decide in the same situations as their groups. It was common knowledge that the decision sequences shown to groups were generated by individual players facing the identical game. This procedure provides us with a large number of observations for group decisions in similar situations. While all other factors remained constant, we had groups communicate via chat rather than deciding face-to-face, as in the benchmark experiment. Henceforth, the benchmark treatments are denoted as *FIT/PIT f2f*, in contrast to *FIT/PIT chat*. To rule out the possibility that distrust in the preprogrammed decision sequences drive our results, we ran an additional treatment reported in subsection 4.3.

Furthermore, as we aimed to determine the impact of varying cognitive abilities within groups, participants took the cognitive reflection test (Frederick, 2005)⁴ beforehand, whereby subjects were paid 1€ for answering all three questions correctly. Groups were subsequently matched according to their performance in the test. We classify individuals in two categories, as in Oechssler et al. (2009): participants with two or more correct answers in the CR-test and those with one or less correct answer. Groups exclusively composed of participants of the first category are referred to as *high*, groups composed of the second category as *low* and groups composed of both categories as *mixed*.

For both treatments there were four sessions, giving us a total of 93 (24/21/24/24) participants in *PIT* and 90 (24/24/21/21) for *FIT*. For the additional treatment serving as a robustness check, we conducted 8 sessions, with 9 subjects each. We used zTree (Fischbacher, 2007) and ORSEE (Greiner, 2004). Experiments were carried out at the Göttingen Laboratory of Behavioral Economics using a standard subject pool across all disciplines. 46% of participants

⁴ The cognitive reflection test by Frederick (2005) is a simple task involving three questions, which tests the individual tendency of choosing intuitive responses over more reflected solutions. The CRT is correlated with cognitive abilities and with performance on heuristics-and-biases tasks, as reported by Toplak et al. (2011).

in our sample were female. Including show-up fees, subjects earned 10.40€ on average. The sessions lasted around sixty minutes.⁵

3. HYPOTHESES

Given that group success crucially depends on the demonstrability of correct solutions (Laughlin et al., 2002), our basic hypothesis is that the mode of communication has a detrimental influence on overall rationality when compared to the baseline face-to-face setting by F&I. This is expected to hold for both treatments, i.e. the purely intellectual and the judgmental task. We thus assume that the positive effects of group decision making will not hold regardless of the mode of communication for a social learning game. Therefore, we formulate:

Hypothesis 1 (“Rationality and virtual communication”).

- a) Virtual communication reduces rationality in the full information treatment (*FIT*).
- b) Virtual communication reduces rationality in the private information treatment (*PIT*).

Our second hypothesis is concerned with the effect of group composition for the particular case of cognitive abilities. Since higher cognitive abilities are associated with more rational behavior on an individual level (see e.g. for the beauty contest game, Burnham et al., 2009; Brañas-Garza et al., 2012), we expect groups composed of subjects scoring well on the CR-test to achieve more rational decisions. We thus imply that it is not the process of group discussion itself, but rather the individual cognitive abilities that increase group performance.

Hypothesis 2: (“Rationality and cognitive abilities”).

- a) Groups composed of individuals showing higher cognitive abilities act more rationally in the full information treatment (*FIT*).
- b) Groups composed of individuals showing higher cognitive abilities act more rationally in the private information treatment (*PIT*).

⁵ The original instructions were in German and are available from the authors upon request. A translation is documented in the appendix.

4. RESULTS

For the basic analysis of rationality in group decisions, we build on the same assumptions as F&I. Rational behavior requires players in *PIT* to identify and correctly interpret informative and uninformative prior decisions. A decision is uninformative, i.e. reveals no additional information, if Bayesian updating prompts a preceding player to pick a specific urn regardless of their private information. Rational players thus apply a counting rule restricted to informative decisions.

To qualify decisions as being in line with Bayesian rationality, it is assumed that all players generally believed that their predecessors decided in accordance with Bayes' Rule. The distinct violation of this behavior in a *PIT* setting, i.e. a decision not following a cascade of prior decisions, is assumed to prompt subsequent deciders to believe that respective individuals followed their private signal.⁶ Also, if the restricted counting rule leads to an indifference situation regarding which urn to choose, players are assumed to expect previous deciders in these situations followed their private signals.⁷

By contrast, in *FIT*, rational decisions imply the use of a simple counting rule, as all prior draws are informative. Again, if applying the counting rule gives the same probability for both urns, players are expected to decide in accordance with their private signals.

Both treatments include situations whereby the private signal and the majority of preceding drawn marbles or decisions are congruent, as well as situations in which the private signal opposes the decision history. We denominate a situation whereby players need to discard their private signals in order to act rationally as a *non-private situation*. Past studies have pointed out the relevance of these decision situations, with the reluctance to discard private signals being the main cause for irrational decisions in the urn game (Weizsäcker, 2010). We thus put an emphasis on the analysis of the varying performance in *non-private situations*.

With regard to the CR-test taken by participants beforehand, we find that 25.68% of all participants answered none of the three questions correctly, 19.67% give one and 27.87% two

⁶ A cascade is defined as a situation in which it becomes rational to follow the preceding pattern of decisions regardless of private signals. For instance, the player on the third position in a decision sequence rationally follows a uniform decision of the two preceding players regardless of her own private signal. Thus, a cascade situation is established, which continues for the rest of the period, rationally ending in a unanimous decision of all players.

⁷ These assumptions are in line with behavior regularly observed in social learning games and are even more plausible when taking into account a small, but positive error rate (see e.g. Weizsäcker, 2010).

correct answers, while 26.78% achieve three correct answers. Accordingly, our matching procedure gives us 37.7% *high* groups, 32.79% *low* groups and 29.51% *mixed* groups. Recall that all group types were presented similar preprogrammed decision situations, thus differences in rationality can be ascribed to differences in cognitive abilities.⁸

Since we are interested in comparing our data to the results obtained by observing randomly matched groups under face-to-face communication, we estimate a probability weighted average of decision rationality for our matched groups. This calculation follows a simple procedure accounting for the probability that a specific group type is formed under a random matching routine and the fractions of correct answers in the CR-test.⁹

Table 1 presents the aggregate data for decision rationality in our treatments and Table 2 summarizes the results of F&I. If not mentioned otherwise, all testing procedures are performed “two-sided” and every individual or group is treated as one observation only.

treatment	share rational (obs.)				share rational (obs.) in <i>non-private situations</i>			
	<i>high</i>	<i>low</i>	<i>mixed</i>	<i>w-average</i>	<i>high</i>	<i>low</i>	<i>mixed</i>	<i>w-average</i>
<i>FIT chat</i>	.9758 (165)	.9267 (150)	.9556 (135)	.9650 (450)	1.0000 (25)	.8462 (26)	.9048 (21)	.9021 (72)
<i>PIT chat</i>	.9056 (180)	.8733 (150)	.9111 (135)	.8984	.5357 (28)	.3043 (23)	.5455 (22)	.5304 (73)

Table 1: Rationality by group type for both treatments with the total number of observations in parentheses.

Percentages are given relative to the respective total number of observations. *w-average* denotes the weighted average.

⁸ For *FIT chat* and *PIT chat* there are potentially 180 preprogrammed decision situations derived from the 12 groups of F&I. Each group is presented 15 situations. Thus, not all decisions are covered by each group type, e.g. for the *mixed* groups in *PIT* only 129 situations were actually played (149 for *high* groups, 141 for *low* groups). Differences between the actual number of situations played and the total number of observations occur due to a small number of situations that were played more than once, which was caused by programming restrictions of z-Tree. However, since there are no systematic differences in the situations presented to the different group types, we can rule out that varying color histories triggered differences in rationality on the group type level.

⁹ The probability of a subject to answer one or less questions in the CR-test correctly amounts to 0.4535, consequently the probability of answering two or more correctly is equal to 0.5465. Thus, a *high* group should appear 16.32% ($=0.5465^3$) of the times given a random matching procedure; a *low* group in 9.33% ($=0.4535^3$) and a *mixed* group in 74.35% ($=1-0.5465^3-0.4535^3$) of the times. The weighting factors for *high*, *low*, *mixed* are then .4329, .2844, 2.52 respectively.

treatment	share rational (obs.)		share rational (obs.) in <i>non-private situations</i>	
	groups	individuals	Groups	individuals
<i>FIT f2f</i>	.9278 (180)	.9093 (1080)	.7419 (31)	.707 (215)
<i>PIT f2f</i>	.9722 (180)	.9028 (1080)	.8966 (29)	.6436 (202)

Table 2: Rationality for groups and individuals as reported in Fahr and Irlenbusch (2011) with the total number of observations in parentheses. Percentages are given relative to the respective total number of observations.

4.1 PERFORMANCE IN *FIT*

Analyzing decision rationality with respect to the mode of communication, we find that the overall weighted average in *FIT chat* is not significantly lower in comparison to *FIT f2f* (Satterthwaite's unequal-variance t-test, $t=-1.0708$, $p=.2927$). In contrast, for *non-private situations*, group rationality is around 16 percentage points higher in *FIT chat* (Satterthwaite's unequal-variance t-test, $t=-1.7932$, $p=.0952$), pointing to a positive effect of virtual communication in *non-private situations*. Thus, the implementation of virtual communication might even foster rational decision making for a purely intellectual task. However, we conclude that virtual communication does not lead to a deterioration of rationality for the full information condition (*FIT*) and thus find no evidence in support of H1a.

Considering differences with respect to cognitive abilities, we cannot reject the null of equal overall rationality across group types (Kruskal-Wallis test with ties, $\chi^2(2)=3.421$, $p=.1807$; for *non-private situations* $\chi^2(2)=3.234$, $p=.1985$). However, there is a significant difference in rationality between *high* and *low* groups (Mann-Whitney test, $z=1.811$, $p=.0701$; for *non-private situations* $z=1.824$, $p=.0682$). *Mixed* groups show a level of rationality, that is neither significantly different from *low* groups (Mann-Whitney test, $z=-.0370$, $p=.7116$) nor from *high* groups (Mann-Whitney test, $z=.673$, $p=.5011$). We conclude that there is, albeit weak, evidence in support of H2a since at least *high* groups outperform *lows*. High cognitive ability of all group members has a small positive impact on group rationality compared to groups in which all members show low cognitive ability.

4.2 PERFORMANCE IN *PIT*

In *PIT chat*, as indicated by the weighted average, group rationality is significantly lower compared to *PIT f2f* (Satterthwaite's unequal-variance t-test, $t=3.3085$, $p=.0027$). Additionally, groups in *PIT chat* show no significantly different overall rationality compared to individual players in *PIT f2f* (Satterthwaite's unequal-variance t-test, $t=.3613$, $p=.7189$). In *non-private situations* groups not only perform worse under virtual communication

(Satterthwaite's unequal-variance t-test, $t=4.0059$, $p=.0003$), they show even lower levels of rationality when compared to individual players (Satterthwaite's unequal-variance t-test, $t=1.7771$, $p=.0831$). The rationality premium for groups communicating face-to-face amounts to 36.8 percentage points in *non-private situations*. In sum, we find strong evidence in support of H1b. Accordingly, groups using chat more often fail to discard their private signal in the crucial *non-private situations*, which leads to deteriorating overall rationality for the judgmental task (*PIT*).

Concerning cognitive abilities in *PIT chat*, we find that overall rationality is not significantly different across group types (Kruskal-Wallis test with ties, $\chi^2(2)=1.046$, $p=.5929$; for *non-private situations* $\chi^2(2)=1.826$, $p=.4014$). Differences between *mixed* and *high* groups are particularly small. In *non-private situations*, *low* groups discard their signals less often compared to *high* and *mixed* groups. This could be interpreted as supporting the notion that the most capable team members drive overall group rationality. However, the difference between pooled *mixed* and *high* groups on the one side and *low* groups on the other is not significant (Mann-Whitney test, $z=.983$, $p=.3258$; for *non-private situations* $z=1.285$, $p=.1989$).

Nevertheless, the rationality of *high* groups in *PIT chat* is still significantly lower compared to the randomly matched groups in *PIT f2f* (Mann-Whitney test, $z=-1.780$, $p=.0751$; for *non-private situations* $z=-2.102$, $p=.0355$).¹⁰ Thus, even groups showing higher cognitive abilities are unable to compensate the negative effect of virtual communication. We conclude that there is no evidence in support of H2b, thus groups showing higher cognitive abilities are not capable of deciding more rationally in the judgmental task.

4.3 ROBUSTNESS CHECK FOR *PIT*

Obviously, it could be argued that subjects might have primarily suspected the preprogrammed decision histories in *PIT chat* to be manipulated. This distrust could in turn have driven the reluctance to discard private signals, implying that virtual communication per se cannot be held responsible for the strong in rationality. To clarify this, we test for the

¹⁰ This result can be seen as giving additional support to the robustness of our main findings. Since subjects in the cognitive reflection test performed rather poorly in comparison to the results presented by Oechssler et al. (2009) and Frederick (2005), the estimated weighted averages may be somewhat too low. However, our main treatment effect of decreasing rationality due to the virtual communication regime proves robust, since even the high cognitive ability groups are not capable to achieve similar levels of rationality as randomly matched groups communicating face-to-face.

robustness of our results by running an additional treatment that concisely replicates the benchmark treatment by F&I, but has groups communicate via chat. Consequently, groups observed decisions taken by individual players who were now physically present in the lab and groups are assigned to an individual player (focal individual), facing identical decision situations. We ran 8 sessions with 9 subjects, giving us 8 randomly matched groups each taking 15 decisions. We find 89.17% of group decisions to be rational with only 60.9% in *non-private situations*. The respective focal individuals achieve 96.66% rational decisions overall with 95.65% in *non-private situations* (Wilcoxon matched-pairs signed-ranks test, $z=-1.616$, $p=.1060$; for *non-private situations* $z=-2.188$, $p=.0287$).¹¹ Thus, focal individuals performed even better than the respective groups and not vice versa. At all, individuals show an overall rationality of 91.11% with 69.66% in *non-private situations*, which is not significantly different from the respective group performance (Mann-Whitney test, $z=.788$, $p=.4307$; for *non-private situations* $z=.207$, $p=.836$).

At the same time, group rationality is not significantly different from the groups in *PIT chat* (Satterthwaite's unequal-variance t-test, $t=.4520$, $p=.6614$; for *non-private situations* $t=-1.5284$, $p=.1424$). Nonetheless, it is again significantly lower than in *PIT f2f* (Mann-Whitney test, $z=2.293$, $p=.0219$; for *non-private situations* $z=2.039$, $p=.0415$).

We can thus rule out distrust in the preprogrammed decision histories as being the driving force for our treatment effect. Groups in *PIT* rely on personal communication to achieve relatively higher levels of rationality when compared to individuals.

5. CONCLUSION

Economic group research has arrived at the conclusion that groups are more rational economic agents, which may rehabilitate assumptions of rational behavior in actual decision situations. We argue that the experimental studies showing higher rationality need to be extended to encompass virtual communication, which has emerged as a prevalent form of decision making within organizations. To further this point, we compare the rationality of groups in a social learning game featuring face-to-face decision making and chat

¹¹ As pointed out by F&I, tests within the *PIT* treatment regarding differences between individuals and groups have to be performed using a testing procedure for matched pairs (Wilcoxon signed rank test). This is due to the interdependence of groups' and the focal individuals' decisions. Recall that the focal individuals' decisions become public information and thus influence other individuals' decisions that are presented to groups as well as to focal individuals in subsequent periods. Accordingly, observations of groups and focal individuals within *PIT* are not independent.

communication. We show that group rationality remains high for an intellectual task, but strongly deteriorates for a judgmental task. Our group matching procedure shows that higher cognitive abilities lead to increased rationality for an intellectual task but not for a judgmental task. Apparently, groups are able to perform better given more cognitively capable members in intellectual tasks, regardless of virtual communication. For the judgmental task, in contrast, even high cognitive ability groups using virtual communication are unable to achieve the rationality of those using face-to-face communication. Thus, the process of group discussions appears to be critical for group rationality instead of the members' cognitive abilities. These results stress the importance of the demonstrability of solutions for superior group performance. Once judgmental reasoning is hindered by non-personal communication, group performance deteriorates even below the average level of individuals, which is in line with experimental results in social psychology (Laughlin et al., 2002). We would thus argue that superior group rationality in many economic studies involving judgmental aspects might not hold for conditions of virtual communication. Considering real-world decision making, our results hold relevance when choosing between different potential communication regimes for group decisions. Our results suggest that virtual communication is a suitable tool for rather simple decisions that have clear and demonstrable solutions. In this case, more capable members efficiently convey information, which enables groups to take better decisions. For judgmental tasks, personal conversations appear to foster more successful group decision making. While personal meetings may be more costly, we argue that they are bound to increase group performance for judgmental tasks. Inversely, if the organizational structure only allows for virtual communication, having individuals take decisions involving judgmental aspects might yield superior results. We thus argue that communication structures in decentralized organizations should be carefully crafted according to the specific tasks at hand in order to optimize group and individual performance.

ACKNOWLEDGEMENTS

We are grateful to René Fahr and Bernd Irlenbusch for their kind cooperation and access to their experimental data. We would also like to thank Kilian Bizer for his helpful comments. Simon Schütz provided excellent student assistance in programming the experiment.

REFERENCES

- Anderson, L.R., Holt, C.A., 1997. Information Cascades in the Laboratory. *The American Economic Review* 87 (5), 847-862.
- Balliet, D., 2010. Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution* 54 (1), 39-57.
- Baltes, B., Dickson, M., Sherman, M., Bauer, C., LaGanke, J., 2002. Computer-Mediated Communication and Group Decision Making: A Meta-Analysis. *Organizational Behavior and Human Decision Processes* 87 (1), 156-179.
- Banerjee, A.V., 1992. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107 (3), 797-817.
- Bicchieri, C., Lev-On, A., 2007. Computer-mediated communication and cooperation in social dilemmas: an experimental analysis. *Politics, Philosophy & Economics* 6 (2), 139-168.
- Bikhchandani, S., Hirshleifer, D., Welch, I. (1992). A theory of fads, fashion, custom, and cultural change in informational cascades. *Journal of Political Economy* 100 (5), 992-1026.
- Brañas-Garza, P., Garcia-Muñoz, T., González, R. H., 2012. Cognitive effort in the beauty contest game. *Journal of Economic Behavior & Organization* 83 (2), 254-260.
- Burnham, T. C., Cesarini, D., Johannesson, M., Lichtenstein, P., Wallace, B., 2009. Higher cognitive ability is associated with lower entries in a beauty contest. *Journal of Economic Behavior & Organization* 72 (1), 171-175.
- Charness, G., Sutter, M., 2012. Groups make better Self-Interested Decisions. *The Journal of Economic Perspectives* 26 (3), 157-176.
- Cox, J., Hayne, S., 2006. Barking up the right tree: are small groups rational agents?. *Experimental Economics* 9 (3), 209-222.
- De Guinea, A.O., Webster, J., Staples, D., 2012. A meta-analysis of the consequences of virtualness on team functioning. *Information & Management* 49 (6), 301-308.
- Fahr, R., Irlenbusch, B., 2011. Who follows the crowd - Groups or individuals?. *Journal of Economic Behavior & Organization* 80 (1), 200-209.

- Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2), 171-178.
- Frederick, S., 2005. Cognitive reflection and decision making. *The Journal of Economic Perspectives* 19 (4), 25-42.
- Goeree, J.K., Palfrey, T.R., Rogers, B.W., McKelvey, R.D., 2007. Self-correcting information cascades. *The Review of Economic Studies* 74 (3), 733-762.
- Greiner, B., 2004. An online recruitment system for economic experiments. *GWDG Berichte* 63, 79-93.
- 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.
- Laughlin, P.R., 1980. Social combination processes of cooperative problem-solving groups on verbal intellectual tasks, in: Fishbein, M. (Ed.), *Progress in Social Psychology*. Hillsdale NJ: Erlbaum, 127-155.
- Laughlin, P.R., 2011. *Group Problem Solving*. Princeton and Oxford: Princeton University Press.
- 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.
- Mesmer-Magnus, J., DeChurch, L., Jimenez-Rodriguez, M., Wildman, J., Shuffler, M., 2011. A meta-analytic investigation of virtuality and information sharing in teams. *Organizational Behavior and Human Decision Processes* 115 (2), 214-225.
- Oechssler, J., Roider, A., Schmitz, P.W., 2009. Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization* 72 (1), 147-152.
- Toplak, M.E., West, R.F., Stanovich, K.E., 2011. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition* 39 (7), 1275-1289.
- Weizsäcker, G., 2010. Do we follow others when we should? A simple test of rational expectations. *The American Economic Review* 100 (5), 2340-2360.

APPENDIX

The instructions documented below refer to the *PIT* treatment. The differences to the *FIT* treatment are indicated in brackets.

General Information

In this game, there are two urns, each with three marbles. Urn A contains two red and one blue marbles. Urn B contains two blue and one red marbles. The number and distribution of the marbles will remain constant during the entire game.

At the beginning of every round, the computer will select one of the two urns (A or B), with equal likelihood. You are not informed which urn has been chosen for this round. Six players now have to guess sequentially which urn has been chosen. Once it is your turn, one of the three marbles will be chosen by the computer, with equal likelihood, and shown to you. Every player can only see the marble that has been drawn for him/her.

Besides the individual marbles, each player can observe [*the previous marbles drawn for*] the decisions of the other players. For instance, this means that the second player sees [*the marbles drawn for the first player*] the decision of the first player, the third player sees the [*marbles*] decisions of the first and second players, and so forth.

The course of the game

In this game, you do not decide on your own, but rather in a group with two other players. You can communicate using the chat function to take a unanimous decision. The chat and all other data will be recorded anonymously. At first, you will answer three questions, and will have two minutes to answer each question. Your group membership will be determined according to the results of the questions. Accordingly, the game will start once all players have answered the questions. With your group, you will be randomly assigned to one of the six positions in the decision sequence. The decisions [*marbles drawn*] you see are from a prior experiment with individual players, who played the game under the exact same conditions as you. After your group has been shown the randomly determined marble, you have to take a common guess within 3 minutes. Once all groups have taken their guesses, the round ends and the correct urn will be shown to all players of your group. This is repeated for 15 rounds.

Your payoff

For a correct guess, each group member receives a payoff of 1€. The gains from all rounds will be added and paid to you at the end of the game. You will additionally be paid 1€ for the correct answer to all three questions at the beginning. Please note: If your group fails to give a

common answer within 3 minutes (and an additional 30 seconds), you will not receive a payoff for the respective round.