Designing Incentive Systems for Truthful Forecast Information Sharing Within a Firm

Case Study, Theory and Experiments

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2017

vorgelegt von
Diplom-Wirtschaftsinformatikerin Lisa Scheele
aus
Nienburg (Weser)

Referent:	Prof. Ulrich W. Thonemann, Ph.D.
Koreferent:	Prof. Dr. Fabian J. Sting
Tag der Promotion:	01.02.2018

Contents

List of Figures

List of Tables

List of Abbreviations

act.	Actual
avg.	Average
beh.	Behavioral
BIC	Bayesian information criterion
CLER	Cologne Laboratory for Economic Research
comp.	Computerized
corr.	Correlation
ECU	Experimental currency unit
excl.	Excluding
GLS	Generalized least squares
IC	Initial conditions
ML	Maximum likelihood
n. a.	Not available
OLS	Ordinary least squares

OM	Operations management
ORSEE	Online recruitment system for economic experiments
PBE	Perfect Bayesian equilibrium
pp	Percentage points
RSS	Sum of squared residuals
SKU	Stock keeping unit
std.	Standard
std. dev.	Standard deviation

List of Symbols

α	Critical ratio of Operations
AS_{it}	Actual unit sales of product i in month t
b	Unit sales bonus
β	Lying aversion factor
$\tilde{\beta}$	Log-transformed lying aversion factor
C	Co-domain of candidate equilibrium distortion strategy
c	Forecast correction value in computerized order decisions model
C_O	Fixed compensation of Operations
c_o	Unit overage cost of Operations
c_o^C	Unit overage cost of the company
C_S	Fixed compensation of Sales
c_u	Unit underage cost of Operations
c_u^C	Unit underage cost of the company
D	Random variable of demand
d	Realization of demand
δ	Distortion value
δ^{pref}	Preferred distortion value of Sales
δ^{sep}	Distortion value in separating equilibrium

$\boldsymbol{\delta_i^{sep}}$	Distortion vector of subject i in separating equilibrium
δ_R^{sep}	Distortion value in separating equilibrium of reference point model
δ_{std}^{sep}	Distortion value in separating equilibrium under standard model assumptions
$\boldsymbol{\delta_i}$	Distortion vector of subject i
δ_θ	Equilibrium distortion value in naïveté model
$\Delta\Pi_C$	Difference in expected profit of the company
$\Delta\Pi_O$	Difference in expected payoff of Operations
$\Delta\Pi_S$	Difference in expected payoff of Sales
E	Random variable of market uncertainty
ε	Realization of market uncertainty
η_x	Regression coefficient of variable x
$\mathbb{E}_X$	Expected value with respect to random variable X
F	Cumulative distribution function of market condition
f	Probability density function of market condition
f_ξ	Probability density function of market condition under trust-based belief
F_Z	Cumulative distribution function of random variable Z in trust model
FC_{it}	Unit sales forecast of product i in month t
FCA_{it}	Forecast accuracy of product i in month t
G	Cumulative distribution function of market uncertainty
g	Probability density function of market uncertainty
γ	Forecast error penalty factor
γ_O	Behavioral cost factor of Operations

$\tilde{\gamma}$	Log-transformed forecast error penalty factor
H	Cumulative distribution function of demand
I	Interval
i	Variable for subjects or products
k	Number of parameters
κ	Alignment factor
$\mathcal{L}$	Likelihood function
λ	Loss aversion factor
log	Natural logarithm
m	Unit profit margin of the company
μ	Belief system of Operations
μ^*	Belief system of Operations in equilibrium
μ^{sep}	Belief system of Operations in separating equilibrium
μ_R^{sep}	Belief system of Operations in separating equilibrium of reference point model
μ_β	Fixed effect of lying aversion factor
$\mu_{\tilde{\beta}}$	Fixed effect of log-transformed lying aversion factor
$\boldsymbol{\mu}$	Fixed effects vector
μ_γ	Fixed effect of forecast error penalty factor
$\mu_{\tilde{\gamma}}$	Fixed effect of log-transformed forecast error penalty factor
μ_θ	Belief system of Operations in naïveté model
μ_ξ	Belief system of Operations in trust model

$\mathcal{N}$	Normal distribution
n	Number of observations
ω	Psychological cost parameter for over- and underforecasting
$P(X = x)$	Probability that a random variable X takes on value x
$\bar{p}$	Threshold cost
p_o	Unit overforecasting penalty
p_u	Unit underforecasting penalty
Φ	Random variable of market condition
ϕ	Realization of market condition
$\hat{\phi}$	Demand forecast
$\hat{\phi}^R$	Reference forecast
$\hat{\phi}_c$	Equilibrium forecast of Sales in computerized order decisions model
$\hat{\phi}_\xi$	Equilibrium forecast of Sales in trust model
Π_C	Expected profit of the company
Π_S	Expected payoff of Sales
$PLAN$	Dummy variable for the planning phase
$POST$	Dummy variable for the post-implementation phase
$\mathbf{\Psi}$	Variance-covariance matrix of random effects
q	Order quantity
q^*	Response function of Operations in equilibrium
q^{FB}	First-best order quantity
q^{sep}	Response function of Operations in separating equilibrium

q_R^{sep}	Response function of Operations in separating equilibrium of reference point model
q_c	Order function of Operations in computerized order decisions model
q_θ	Equilibrium response function of Operations in naïveté model
q_ξ	Equilibrium response function of Operations in trust model
$\mathbb{R}$	Set of real numbers
$r_{\beta,i}$	Random effect of lying aversion factor of subject i
$r_{\gamma,i}$	Random effect of forecast error penalty factor of subject i
$\boldsymbol{r_i}$	Random effects vector of subject i
r_O	Reservation utility of Operations
r_S	Reservation utility of Sales
ρ	Correlation of forecast error penalty and lying aversion factors
$\tilde{\rho}$	Correlation of log-transformed forecast error penalty and lying aversion factors
s	Signaling function of Sales
s^*	Signaling function of Sales in equilibrium
s^{pref}	Preferred signaling function of Sales
s^{sep}	Signaling function of Sales in separating equilibrium
s_R^{sep}	Signaling function of Sales in separating equilibrium of reference point model
s_c	Candidate equilibrium signaling function of Sales
s_θ	Equilibrium signaling function of Sales in naïveté model

$SAFE$	Number of safe choices
SG	Dummy variable for the order of play
σ_β	Standard deviation of lying aversion factor
$\sigma_{\tilde{\beta}}$	Standard deviation of log-transformed lying aversion factor
σ_γ	Standard deviation of forecast error penalty factor
$\sigma_{\tilde{\gamma}}$	Standard deviation of log-transformed forecast error penalty factor
σ_O	Response strategy of Operations
σ_S	Signaling strategy of Sales
σ_u^2	Standard deviation of independent error term
σ_v^2	Standard deviation of subject- or product-specific error term
t	Variable for time periods
$T2$–$T8$	Dummy variables for Treatments 2–8
$T2R$	Dummy variable for Treatment 2R
$T5R$	Dummy variable for Treatment 5R
t_T	Period within a treatment
T_i	Number of observations of subject i
τ	Weight on gain-loss utility
θ	Naïveté factor
u_{it}	Independent error term of subject or product i at time t
u	Distortion function of Sales
u^{sep}	Distortion function of Sales in separating equilibrium
u_c	Candidate equilibrium distortion function of Sales
U_O	Expected utility of Operations
U_S	Expected utility of Sales

U_S^{FC}	Forecast-dependent part of the utility of Sales
U_S^N	Consumption utility of underforecasting
U_S^P	Consumption utility of overforecasting
U_S^R	Utility of Sales in reference point model
U_S^{SB}	Quantity-dependent part of the utility of Sales
v_i	Error term specific to subject or product i
v	Value function of prospect theory
$\bar{v}$	Value function of reference point model
w	First derivative of the distortion function of Sales
ξ	Trust factor
Z	Random variable in trust model

1. Introduction[1]

"On the one hand, we want to benefit from cheating, while on the other, we want to be able to view ourselves as honest, honorable people. This is where our amazing cognitive flexibility comes into play: As long as we cheat by only a little bit, we can secure the benefits and still view ourselves as marvelous human beings."

Dan Ariely (Behavioral Economist)

1.1. Motivation

For many companies, the challenge to manage the internal supply chain from sourcing to serving the end customer is growing (PwC, 2015). On the one hand, customer needs are now more sophisticated and volatile than ever. On the other hand, operations networks are becoming more global and complex in structure. A recent analysis by McKinsey & Company shows that organizations are becoming more mature in managing their supply chains, which is reflected in decreasing and converging high-level supply chain benchmarks, such as overall fill rates and supply chain costs (Karlsson et al., 2017). However, when taking a more granular look, the performance along individual dimensions, such as demand forecasting, is still heterogeneous. For example, among consumer goods companies with similar levels of demand volatility, the

[1] This thesis is an extended version of Scheele et al. (2017), which is joint work with Ulrich Thonemann and Marco Slikker. It benefited from the comments of four anonymous referees and the editors of *Management Science*. Preliminary research results have been presented at the following conferences: *Behavioral Operations Management Conference* (Washington, D.C., 2012), *INFORMS Annual Meeting* (Phoenix, AZ, 2012), *European Conference on Operational Research* (Rome, 2013), *MSOM SCM SIG Conference* (Seattle, WA, 2014). An earlier working paper version was awarded the *2014 INFORMS Behavioral Operations Management Best Working Paper Award* and was presented in the finalists session at the *INFORMS Annual Meeting* (San Francisco, CA, 2014).

gap in forecast accuracy between median and top-quartile firms is still at 17 pp (Karlsson et al., 2017).

The demand forecast is a major input factor in the internal planning process, which ranges from inventory management and production planning to sourcing decisions. The operational performance of a company depends to a large extent on how well it matches supply and demand. Achieving a good match requires accurate information to be available to the decision maker who is in charge of the supply decision. At many companies, different departments are responsible for collecting demand information and making production or order decisions (e.g., Shapiro, 1977; Eliashberg and Steinberg, 1993; Balasubramanian and Bhardwaj, 2004; Oliva and Watson, 2009; Özer and Uncu, 2013). Typically, a sales or marketing division (Sales) collects demand information and creates a demand forecast, which is then used by an operations division (Operations) to make a production or order decision. The rationale behind this task allocation is that Sales is close to the customer and has good information about future demand, while Operations has good information about the effect that production and order decisions have on cost. If the incentive system of Sales is not properly designed, the demand forecast could be biased, which means that Operations would receive inaccurate demand information and consequently make suboptimal decisions (Chen, 2005; Lawrence and O'Connor, 2005; Özer et al., 2011).

Figure 1.1 shows an example of biased forecasts at a global pharmaceutical company. The graphs show the demand forecasts that were produced three months in advance and the actual monthly demand for four products over a period of 48 months. Most demand forecasts are greater than the actual demand, with an average demand forecast inflation of 16.2 %. Further analyses revealed that Sales personnel expect production quantities to increase in their demand forecasts. Higher production quantities increase the availability of products, which results in increased sales and sales bonuses. It is therefore not surprising that Sales has a tendency to inflate demand forecasts. Note, however, that the demand forecasts and demands are correlated, and the inflation is moderate. This implies that actual forecasts contain information about future demands, as opposed to the predictions of an economic model with expected-payoff-maximizing decision makers, which predicts that forecasts are entirely uninformative (see Theorem 1 in Özer et al. (2011) and Theorem 1 in Section 3.2). As Özer et al. (2011) show in an inter-firm setting, this effect can be explained by a human desire to trust and be trustworthy, which is further promoted in long-term relationships. One could expect similar factors to influence forecast information sharing within a firm. However, as the data of Özer et al. (2011) and Figure 1.1 suggest, these factors are not sufficient to remove biases from sales forecasts.

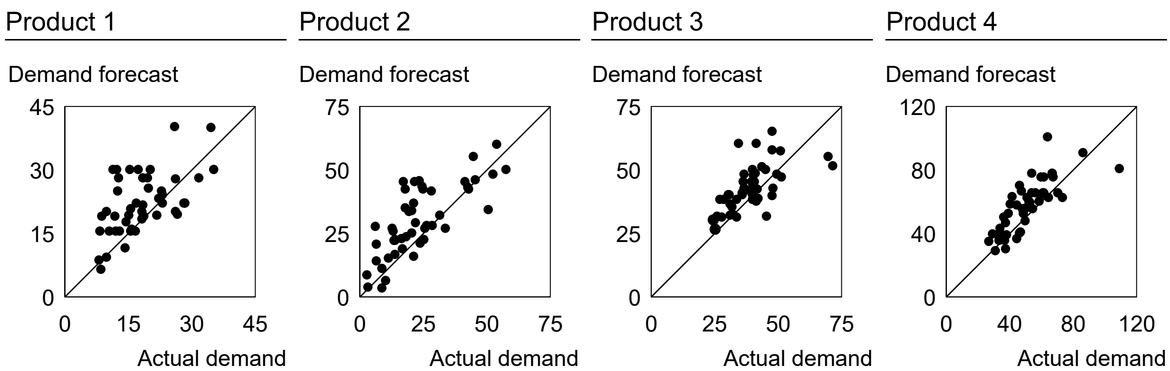


Figure 1.1.: Monthly forecasts and actual demand of example products (in thousand units)

Inspired by the example above, the objective of our research is to design incentive systems for Sales that lead to truthful demand forecasts. The economics literature has analyzed incentive systems and has shown that monetary incentives play an important role in motivating individuals (e.g., Prendergast, 1999; Lazear, 2000; Gibbons, 2005). Properly designed incentive systems can further foster cross-functional collaboration (Ellinger, 2000) and improve supply chain performance (Cohen et al., 2007). However, incentive system designers are faced with certain challenges, such as choosing the right set of performance indicators (Kerr, 1975) and predicting how human decision makers react to incentives (Fehr and Falk, 2002; Camerer and Loewenstein, 2004). We address these issues in our analyses and the design of incentive systems for truthful forecast information sharing within a firm.

1.2. Context and Contribution

There exists a rich body of operations management (OM) literature on supply chain coordination. One branch of research in this field investigates how the actions of a supply chain dyad can be aligned if *information is distributed asymmetrically*, i.e., if one actor is in possession of information that the other actor needs to make an informed decision. Empirical investigations in the semiconductor industry have shown, for example, that a downstream buyer who has superior access to market information uses this advantage to manipulate the production decisions of the upstream supplier (Cohen et al., 2003; Terwiesch et al., 2005). In contrast, it has been shown that sharing sales and market information truthfully across organizational boundaries can significantly improve the operational performance of the supply chain (e.g., Gaur et al., 2005; Cui et al., 2015).

The focus of earlier publications was to *design contracts* that provide incentives for the better informed party to reveal its private information truthfully to the uninformed party (Cachon, 2003). In many cases, these contracts require the better informed actor to “put his money where his mouth is,” i.e., to send a costly signal that credibly conveys the private information (Chen, 2003, p. 342). Alternatively, the uninformed party can try to motivate the informed actor to release its information, such as by offering a menu of contracts. Both approaches are used in the analyses of coordination mechanisms for truthful forecast information sharing. For example, Cachon and Larivière (2001) study how a manufacturer can credibly signal high demand to a supplier through an upfront payment to reserve capacity. Özer and Wei (2006) compare capacity reservation and advance purchase contracts in a supplier-manufacturer relationship and derive conditions under which both types of contracts reveal the private demand information of the manufacturer. Various extensions of the problem and alternative contract forms have been analyzed (e.g., Mishra et al., 2007; Li and Zhang, 2008; Shin and Tunca, 2010; Gümüş, 2014; Chen et al., 2016). Besides the use of formal contracts, repeated interactions between supply chain partners can support truthful forecast information sharing because the uninformed party can use review strategies to detect and punish untruthful behavior (Ren et al., 2010).

While the general structure of the problem is similar, few publications have analyzed the issue of truthful forecast information sharing *within a firm*. In contrast to inter-firm information exchange, a different set of mechanisms is available to coordinate the actions of decision makers who belong to the same organization. In particular, performance-based incentive systems, such as the ones that we consider, allow for enforceable ways of "contracting" between different departments. Research in Marketing and in OM has addressed the question of how to incentivize Sales to provide accurate forecasts. Among the first to approach this problem, Gonik (1978) describes a sales incentive system, which ties the bonuses of sales personnel to their forecasts. In essence, bonus payments under the Gonik scheme are piecewise linear functions of revenues, where the self-selected forecast determines the kink in the function. By adjusting the parameters of this incentive plan, a rational sales agent can be incentivized to provide an accurate estimate of demand, which simultaneously serves as a sales target (Mantrala and Raman, 1990). Alternatively, Sales can be offered a menu of linear contracts to elicit private knowledge about the market and be motivated to exert sales effort (Chen, 2005; Chen et al., 2016). Celikbas et al. (1999) present a model where Sales is penalized for overforecasting and Operations is penalized for understocking. Assuming that the demand forecast of Sales is used as an upper bound in the production decision of Operations, they find that the parameters of such an incentive system can be set such that demand forecasts and production quantities are coordinated.

The traditional OM approach to problems of inter- and intra-firm forecast information sharing builds upon the assumption of rational decision makers who maximize their own payoffs. This assumption has been relaxed in a growing stream of *behavioral research* on these topics. In their study of forecast information sharing under wholesale price contracts, Özer et al. (2011) show that trust and trustworthiness between a supplier and a manufacturer can explain why forecasts are considerably more truthful than expected under rational decision-making assumptions. Their experimental data indicates that the degree of trust depends on environmental factors, such as the cost of capacity, the level of demand uncertainty, and the cultural background (Özer et al., 2011, 2014). Furthermore, trust and trustworthiness seem to be affected by the type of information that an informed supply chain partner communicates: Offering information that enables the uninformed party to make a decision generates higher trust than offering advice or even making the decision on behalf of the uninformed party (Özer et al., 2017). For situations where the upstream party is better informed, Beer et al. (2017) investigate both theoretically and experimentally if a supplier can signal its trustworthiness to a buyer through pre-contractual investments. They find that the prices paid by the buyer, the quality delivered by the supplier, and ultimately the overall supply chain profit are higher if a supplier invests in the upfront signal. For situations where two firms in a supply chain dyad receive different demand information, the experimental data of Hyndman et al. (2013) suggests that pre-play communication can help them to align their capacity decisions even if the information exchange is nonbinding. Ebrahim-Khanjari et al. (2012) develop a model that captures the social characteristics of a salesperson (e.g., selfishness and loyalty) and derive theoretically how these characteristics influence the degree of trust of a retailer in the salesperson's forecast. In situations where a supplier intends to design a menu of contracts but does not know the holding cost of a buyer, the model of Voigt and Inderfurth (2012) and the experimental data of Inderfurth et al. (2013) suggest that a holding cost signal can convey credible information and hence improve supply chain performance only if the trustworthiness of the buyer and the trust of the supplier are well aligned. They also find that there are situations of insufficient trust and underreporting of the cost position, where communication harms the supply chain. Özer and Zheng (2017) synthesize the existing knowledge of the role of trust and trustworthiness in supply chain information sharing.

Our research builds on the work of Özer et al. (2011, 2014). We extend their model by transferring it to an intra-firm context and adding forecast error penalties to the payoff function of the informed party (Sales). Our theoretical and empirical findings contribute to the existing literature on forecast information sharing in several ways: First, we provide robustness to the finding that Sales is more trustworthy than predicted by expected-profit-maximizing behavior even in the absence of formal incentives and that Operations anticipates this trustworthiness when receiving the demand signal. We support this finding empirically with experimental data and insights from a case study. Second, we show how forecast-based incentives influence the forecast decision of Sales and subsequently the order decision of Operations. Third, by combining these insights, we derive conditions for Sales incentive schemes that enable truthful forecast information sharing.

Research on behavioral OM has evolved significantly over the last two decades (reviews are provided in Gino and Pisano, 2008; Bendoly et al., 2010). One stream of research in this field focuses on individual decision biases, typically in newsvendor-type situations (e.g., Schweitzer and Cachon, 2000; Bostian et al., 2008; Ho and Zhang, 2008; Katok and Wu, 2009; Ho et al., 2010; Becker-Peth et al., 2013; Kremer et al., 2014; Zhang et al., 2016). Another stream of research analyzes the behavioral drivers of strategic, interactive decision making in supply chains. Besides trust and trustworthiness, supporting elements of cooperation also include concerns for fairness (Cui et al., 2007; Katok and Pavlov, 2013; Katok et al., 2014; Ho et al., 2014) and the existence of personal relationships (Loch and Wu, 2008). An inherent characteristic of research in behavioral OM is the combination of different research methods (Croson et al., 2013). A well-established method for discovering human decision biases and social preferences in supply chain interactions is controlled laboratory experiments (Katok, 2011). In this thesis, we combine the classical approach of model building with results of laboratory experiments. More specifically, we build a model based on previous research and behavioral theory and use experimental data to estimate the behavioral parameters of our model. We show that our results are robust by conducting an out-of-sample validation experiment and provide evidence that the main insights remain valid under repeated interaction. In the field of OM, there has furthermore long been a call for more case and field research to complement the traditional methods of modeling, optimization and simulation (e.g., Meredith, 1998; Fisher, 2007; Barratt et al., 2011). We follow this call and complement the theoretical and experimental approach by a case study that investigates the use of forecast-based incentives in practice. We hence contribute methodologically to the growing body of behavioral OM research by using a multi-method approach of formal modeling, human-subject experiments and field research to develop and test sales incentive schemes for truthful forecast information sharing.

Our results have important implications. They show that firms can incentivize truthful forecast information sharing by including a forecast-based component in the incentive system of Sales. Our analyses indicate that behavioral factors play an important role in decision making. Taking these factors into account is crucial for predicting and influencing human forecasting behavior.

1.3. Outline

In the following, we briefly outline the structure of the thesis. Chapters 2, 3 and 4 contain the main results of our research – each with a different methodical focus. Chapter 5 serves to provide further robustness to the previous results.

Chapter 2 contains a case study of a global pharmaceutical company, which motivated the theoretical and experimental research of the following chapters. We characterize the initial situation at the company, describe the process of introducing forecast accuracy incentives into the target agreements of product managers at a local sales unit, and report the quantitative and qualitative results of this pilot project.

In *Chapter 3*, we develop a game-theoretic model that describes the forecast exchange between Sales and Operations. We specify the sequence of events and develop utility functions based on behavioral theory for both parties. We define three different types of incentive systems for Sales, all of which include a bonus for sales but differ in the way they penalize forecast errors. We present a solution concept for the signaling game that we consider and derive Pareto-dominant separating equilibria for the forecasting behavior of Sales and the ordering behavior of Operations. Based on the theoretical results we develop a set of hypotheses.

Chapter 4 reports the results of laboratory experiments that test how well our model explains the behavior of actual decision makers. We first describe the design of our main experiment, address the laboratory protocol, and discuss the results of the experiment. Based on the experimental data, we then estimate the behavioral parameters of our model and test the hypotheses. We use the insights of the main experiment to define a new set of treatments that we test in a validation experiment with a new group of subjects. We show that close-to-truthful information sharing can be achieved by our approach.

Chapter 5 summarizes various analyses in support of the results of Chapters 3 and 4. We first take a look at alternative models that relax the assumption of Bayesian belief updating and test different reference points in the valuation of the forecast error penalties. We then report the results of additional experiments, where we replace the human Operations player with a computer and let subjects play the forecasting game repeatedly with the same player instead of different players. We conclude the chapter with a set of supporting analyses based on the data of the main experiment. These include regressions to identify time and order effects, treatment-specific parameter estimations, measures to ensure the overall understanding of the experiment, and an analysis of subjects' risk attitudes.

We conclude our work in *Chapter 6*. Besides summarizing our main results, we lay out how these results can be used to guide the design of sales incentive systems in practice. We also address the limitations of our research and provide directions for future research.

2. Case Study: Forecast Accuracy Incentives at a Pharmaceutical Company

In order to address our research question of whether and how properly designed incentive systems can remedy systematic biases in sales forecasts, we conducted field research at a Western European pharmaceutical company (hereinafter called PharmaCo), which ranks among the top 20 global players in the industry with net sales of over EUR 10 billion and more than 40,000 employees worldwide. We use this case study to motivate our research question and to contextualize the theoretical and experimental research that we conduct. Our field research follows a mixed methods design, combining both quantitative and qualitative methods of analysis (e.g., Yin, 2009; Johnson and Onwuegbuzie, 2004). Information was collected from 2009 to 2010. We conducted 14 semi-structured interviews with representatives from different functional areas in six different countries. We also participated in three workshops where representatives from local sales organizations and global supply chain management functions discussed the matter of forecast accuracy and the potential effects of incentive-based interventions. These field observations were complemented by information from a range of documents, such as standard operating procedures, organization charts and internal survey data that the company made available for the purpose of our research. In addition, we had access to a global database of forecast and sales data, which we used for quantitative analyses.

We will next describe the organization, forecasting process and forecasting performance at the company at the beginning of our field research (Section 2.1). We will then present the setup (Section 2.2) and results (Section 2.3) of the pilot project.

2.1. Initial Situation

Over the last decade, operational topics such as forecasting and planning processes have become increasingly important in the pharmaceutical industry (e.g., Ebel et al., 2013). Since sales margins used to be high and reputation risks immanent, high service levels had been the major objective for many years. Meanwhile, however, increasing price pressure due to health care reforms and declining research efficiency were forcing the industry to become more cost competitive. This transition from a mainly sales-focused to a cost-conscious organization enabled us to accompany the introduction of forecast accuracy incentives in a pilot project at PharmaCo.

The product portfolio of PharmaCo was split into two segments. One segment consisted of pharmaceutical products that a physician must prescribe to a patient (*prescription medicines*). The other segment involved products that were sold through pharmacies without prescription (*consumer products*). The market for prescription medicines was relatively stable. Except for effects caused by product lifecycle-related variations (e.g., during the launch phase and when a product loses its exclusivity) or exceptional disruptions, such as regulatory interventions, demand was generally driven by well-predictable patient populations and the share of prescribing doctors. The demand for consumer products was more volatile as it was subject to the listing decisions of pharmacies and consumer preferences.

PharmaCo was organized into two types of business units: sales units and production units (see Figure 2.1). *Sales units* had the primary purpose of marketing and selling products to a local or regional market. Within a sales unit, products were assigned to individual product managers, who were responsible for creating monthly demand forecasts for their products. *Production units* produced active pharmaceutical ingredients, bulk ware or finished goods and supplied the latter to the global network of sales units. PharmaCo planned and produced entirely to stock. The organizational interface between both units were demand managers on the part of the sales units and supply managers on the part of the production units. Demand managers collected forecasts from local product managers, translated them into net requirements and communicated these to the supply managers in the relevant production units. Supply managers aggregated the net requirements from all sales units and generated production plans.

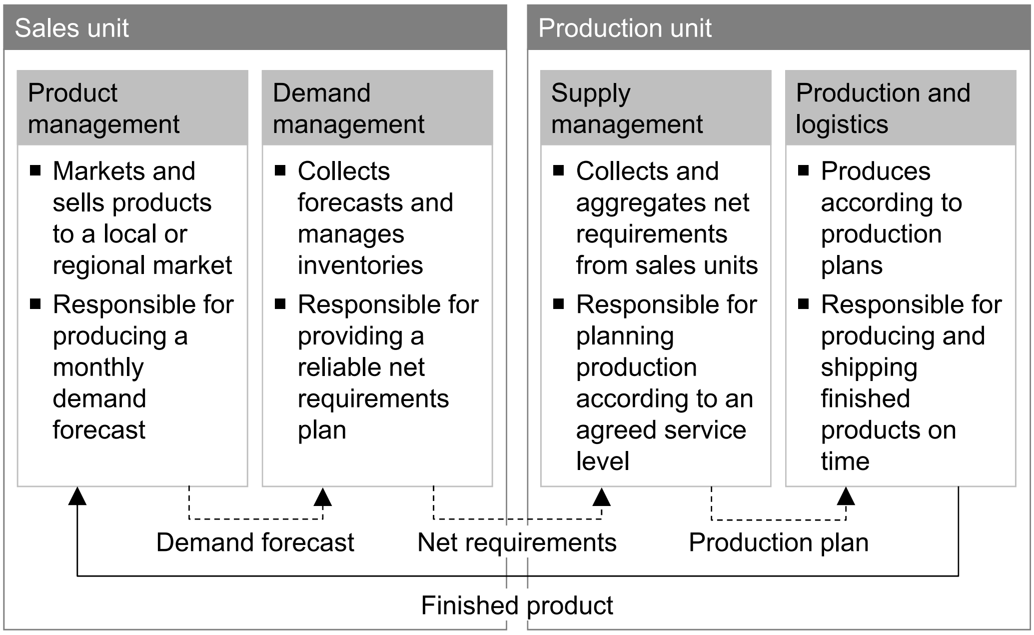


Figure 2.1.: Simplified organization and planning process at PharmaCo

Sales forecasts were made three months in advance and were updated once a month. The forecasts were discussed in monthly consensus meetings between product managers and demand managers. While product managers prepared and presented the forecasts, demand managers challenged the underlying assumptions. In case of disagreement, product managers had the decision right.

The forecast performance at PharmaCo was measured by means of a forecast accuracy metric. The forecast accuracy of a particular stock keeping unit (SKU) was measured as the absolute relative deviation between the forecast and actual sales

$$FCA_{it} = \max\left(1 - \left|\frac{AS_{it} - FC_{it}}{AS_{it}}\right|, 0\right), \tag{2.1}$$

where AS_{it} were actual unit sales of SKU i in month t and FC_{it} was the unit sales forecast for SKU i in month t that had been made three months in advance. For an aggregate view across multiple SKUs, PharmaCo used a non-weighted average of the individual values.

As shown by the example in Figure 1.1, demand forecasts at PharmaCo were systematically biased at the time of our investigation. The forecast accuracy differed considerably between the sales units and individual product managers. Figure 2.2 shows that the average forecast accuracy across all sales units was 67 % in prescription medicines and 58 % in consumer products in 2009. Comparing the average forecast accuracy to that of the top five sales units in each segment indicated a considerable room for improvement.

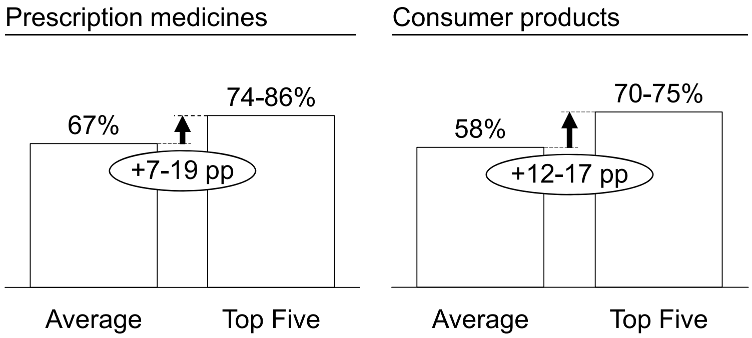


Figure 2.2.: Forecast accuracy at PharmaCo

The forecast bias was not tracked at PharmaCo at the time of our investigation. We measured the bias as the share of inflated forecasts, i.e., forecasts that were greater than actual demand. If forecasts are unbiased, the expected value of this measure is 50 %. Values above 50 % indicate a tendency for overforecasting. Values below indicate a tendency for underforecasting. At the beginning of our research, the average forecast bias across all sales units was 58 % in both segments, prescription medicines and consumer products. A more fine-granular analysis revealed that the forecast bias was significantly greater than 50 % in nearly half of all sales units (Wilcoxon signed-rank test, $p < 0.05$).

An internal survey at PharmaCo revealed that half of the demand managers in the sales units where not satisfied with the forecast accuracy. 70 % of them believed that the low forecast accuracy was due to a lack of motivation and the wrong incentives. Only 30 % believed that the poor performance was due to a lack of skills. Our interviews supported these results. Product managers prioritized their sales targets over a good forecast performance. The forecast for a particular month was made three months in advance, often without taking into account marketing campaigns or other unusual demand patterns. Product managers took any opportunity to increase sales, for example, by means of a marketing campaign, and relied on sufficient safety stock to cover the peak.

> *"At the moment, we have a huge campaign on [product name]. I know the forecast process is key, but at the end of the day I want to push sales and add to the bottom line of the company – even if this hurts my forecast accuracy."*
>
> (Product manager, consumer products, South American sales unit)

Demand managers reported that product managers often submitted overly optimistic forecasts to ensure sufficient supply and safety stocks. To some extent, demand managers could react to inflated forecasts by temporarily building up more inventory than planned. However, if the forecast inflation was too high, they had to cancel orders, which in turn disrupted production plans or led to excess stock at the plants.

> *"The only thing they [product managers] experience is running out of stock if the forecast was too low. To overcome this problem, they report high forecasts. They always plan with buffers in their forecasts. If the forecast accuracy is 70–80 %, we can handle it, but sometimes it is much worse. In the end, we [demand managers] have to cancel orders."*
>
> (Demand manager, Eastern European sales unit)

Next to the low forecast accuracy, demand managers were especially frustrated by the lack of effort and cooperation on the part of product management.

> *"For us, accurate forecasts are the basis of our job and we think it is the responsibility of product managers to improve the forecast. But in the end, they seem to have other objectives. Forecasting is not one of their priorities, even though it should be. Sometimes they don't come to the consensus meetings. There is always an excuse."*
>
> (Demand manager, Western European sales unit)

Some demand managers even felt that they could develop a better forecast than product managers.

> *"I tell you one thing: We can do it much better than them [product managers]. We have tracked this over the last months. Our forecasts were better than those made by product managers. But the problem is, if we did the forecast, they would blame every shortage of supply on us."*
>
> (Demand manager, Western European sales unit)

While demand and supply managers were not allowed to modify the forecast, they would, however, adapt the net requirements to avoid excessive stock and potential destruction due to expired shelf life times.

> *"The forecast is a holy thing. We cannot change it. But if we mistrust the forecast, we adjust the net requirements and the stocks together with the demand managers."*
>
> (Head of supply management, Southern European production unit)

Overall, there seemed to be a lot of tension between the different functional areas. To understand this conflict of interest in more detail, we analyzed the incentive systems of product managers and demand managers. Figure 2.3 depicts the share of different objectives in the variable compensation (end-of-year bonus payment) of product managers and demand managers in the European country where we conducted the pilot project. The incentive systems of other countries were similar. Next to individual qualitative objectives (such as the completion of a certain project) and participation in the overall company profit, both divisions were incentivized by a set of performance indicators that were related to their respective core functions. The incentive system of product managers contained a bonus for revenues and market shares, but did not contain any operational objectives, such as forecast quality. Demand managers were incentivized to optimize stock levels, keep requirements plans stable and achieve the agreed service levels. This clear division of objectives seemed to be one of the reasons behind the silo mentality that we encountered in the interviews and the resulting unsatisfactory forecast quality.

One idea to improve the collaboration between product and demand managers at PharmaCo was to make product managers accountable for forecast accuracy by including corresponding

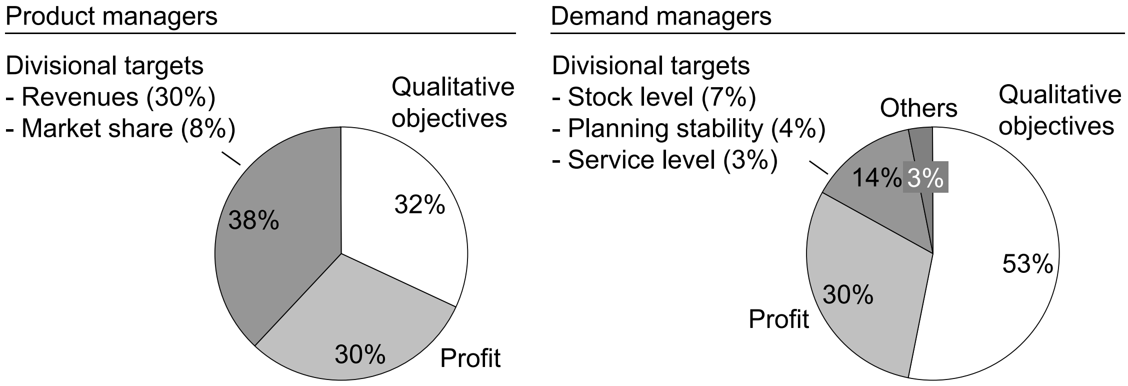


Figure 2.3.: Incentive system at PharmaCo

objectives in their target agreements and linking them to their variable compensation. We will next describe how this idea was tested in a local pilot project.

2.2. Introduction of Forecast Accuracy Incentives

The pilot project was implemented in a major European country that comprised both a sales unit and a production unit. The head of product management and the heads of demand and supply management were involved from the beginning of the project. Figure 2.4 shows the timeline of the project. The initiative was announced in July 2009. In the following six months, a project team developed a formal "inventory policy," which explained the relationship between inventories and service levels and highlighted the importance of accurate forecasts. The policy defined how forecast accuracy was measured and that forecast accuracy targets should be included in the incentive systems of product managers. To increase the acceptance of the proposed changes in product management, the policy did not specify at which weight or percentage to include forecast accuracy in the target agreements. This decision was left to the heads of product management and their team leaders. In December 2009, all stakeholders signed the inventory policy and it became effective in January 2010.

In response to the new policy, the incentive systems in product management were changed. Figure 2.5 shows an example of the product management team of a major product group in prescription medicines. The changes in the other product groups and in the segment of

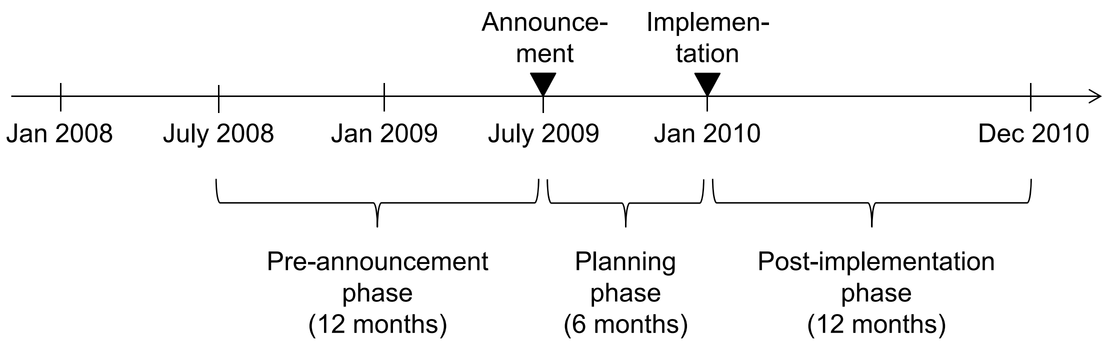


Figure 2.4.: Timeline of the pilot project

consumer products were similar. Compared to the previous situation (see Figure 2.3), the share of qualitative objectives of product managers and their team leaders was reduced and a forecast accuracy target was added with 6 % and 4 % weights respectively. The incentive system of the head of product management did not change. Note, however, that the overall financial impact of the forecast accuracy incentive was comparably low. For example, the variable compensation of a product manager was 12 % of the total annual salary, which corresponds approximately to one and a half months of salary. Of that part, only 6 % was linked to the forecast accuracy target.

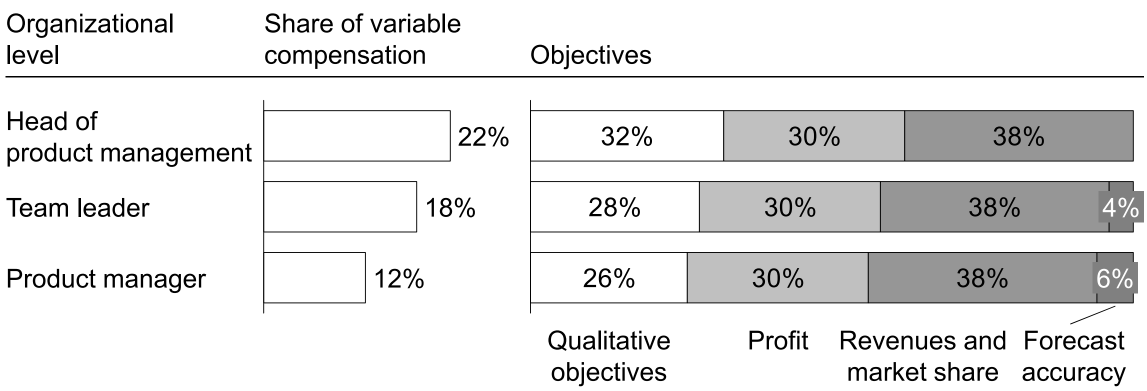


Figure 2.5.: New incentive system of product management

2.3. Results

Figures 2.6 and 2.7 show how the forecast accuracy and forecast bias developed in the pilot sales unit. We analyze forecasts that were made in the 12 months before the project was announced (*Pre*), forecasts that were made in the planning phase (*Plan*), and forecasts that were made in the 12 months after implementation (*Post*). Note that the length of the pre-announcement phase and post-implementation phase are comparable because they both cover a whole year. The aggregate data of the planning phase must be interpreted with caution because it might be distorted by seasonal effects.

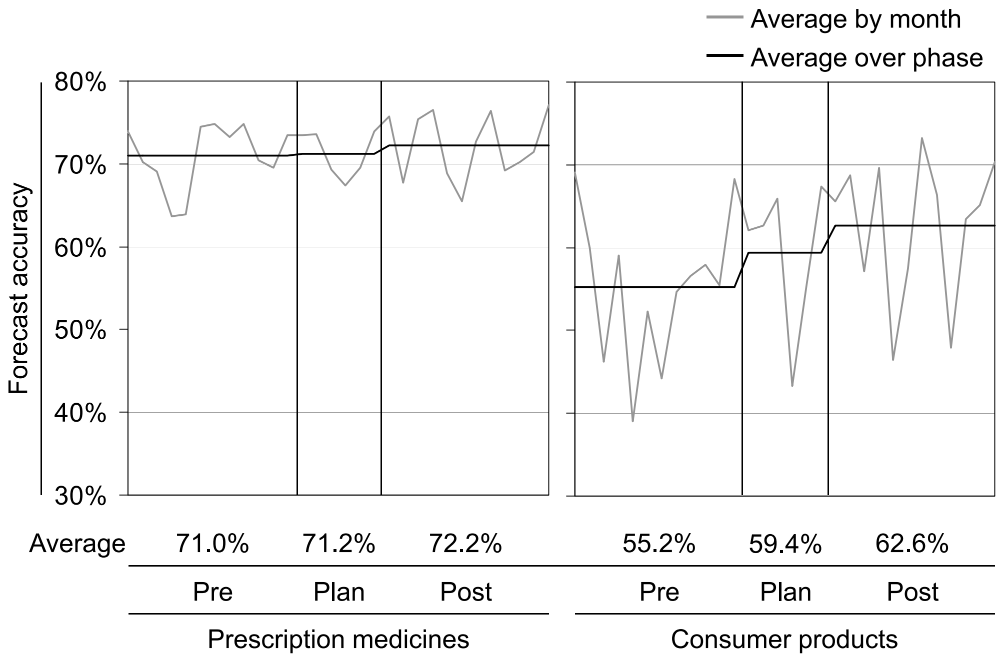


Figure 2.6.: Development of forecast accuracy in pilot sales unit

The average forecast accuracy was 71.0 % in prescription medicines and 55.2 % in consumer products before the forecast incentive project was announced. In prescription medicines, the forecast accuracy improved only by a small degree to a value of 72.2 % with the change in incentives. The difference to the baseline value is weakly significant (Wilcoxon signed-rank test, $p = 0.059$). A random effects generalized least squares (GLS) regression of the forecast accuracy (FCA) on dummy variables of two out of three project phases ($PLAN$ and $POST$)

indicates a slightly greater improvement. We formulate the regression function as follows:

$$FCA_{it} = \text{Intercept} + \eta_{PLAN} \cdot PLAN_{it} + \eta_{POST} \cdot POST_{it} + v_i + u_{it}, \qquad (2.2)$$

where SKUs are indexed by i and months by t. The error term is split into a product-specific part ($v_i \sim \mathcal{N}(0, \sigma_v^2)$) and a part that is independent across all observations ($u_{it} \sim \mathcal{N}(0, \sigma_u^2)$) to account for the grouped structure of the data. Table 2.1 reports the results. Compared to the pre-announcement phase, the forecast accuracy improved by 2.9 pp in the post-implementation phase (z-test, $p = 0.015$). A possible reason for the limited improvement in forecast accuracy could be an unexpected governmental decrease of reference prices in the post-implementation phase. With the announcement of the price adjustment, wholesalers stopped buying until after the price cut was effective. As a result, demand was shifted between the months of May and June, which had a negative effect on the forecast accuracy in these two months. We repeat the above regression, but exclude the data of May and June in both the pre-announcement and the post-implementation phase. Table 2.1 shows that the forecast accuracy improved by 3.9 pp after the introduction of the new incentive system and that this effect is highly significant (z-test, $p = 0.003$).

	Prescription medicines		Consumer products
Model	Full	Excl. May/June	Full
Observations	1,911	1,657	1,450
Groups	76	76	55
Intercept	0.677 *** (0.020)	0.674 *** (0.021)	0.537 *** (0.022)
PLAN	0.004 (0.014)	0.012 (0.015)	0.052 ** (0.022)
POST	0.029 ** (0.012)	0.039 *** (0.013)	0.078 *** (0.018)

Note: Significance of estimates (z-test): *** p < 0.01, ** p < 0.05, * p < 0.1.
Standard errors are reported in parentheses.

Table 2.1.: Estimation results of forecast accuracy by project phase

In consumer products, the forecast accuracy increased significantly from 55.2 % to 62.6 % after the implementation of the new target agreements (Wilcoxon signed-rank test, $p = 0.002$). This segment was not affected by the governmental intervention. Table 2.1 reports the results of a random effect GLS regression as specified in Equation (2.2). The estimate for the increase in forecast accuracy between the pre-announcement and post-implementation phase is 7.8 pp and is highly significantly different from zero (z-test, $p < 0.001$).

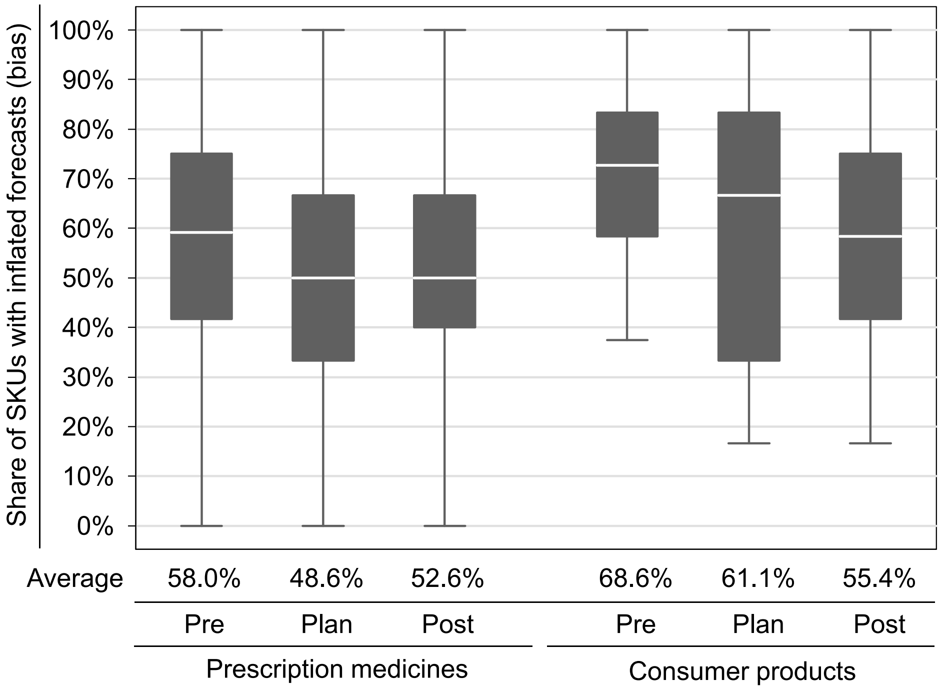


Figure 2.7.: Development of forecast bias in pilot sales unit

With respect to the forecast bias, the results are visible in both product segments (see Figure 2.7). The average forecast bias in the pre-announcement phase was 58.0 % in prescription medicines and 68.6 % in consumer products. Both values were significantly greater than 50 % (Wilcoxon signed-rank test, $p < 0.004$). After the new incentives were introduced, the bias reduced to 52.6 % in prescription medicines and 55.4 % in consumer products. Both values are significantly smaller than in the pre-announcement phase (Wilcoxon signed-rank test, $p = 0.044$ in prescription medicines and $p < 0.001$ in consumer products). This analysis is supported by a random effects logistic regression of the forecast bias ($BIAS$) on dummy variables of two out of three project phases ($PLAN$ and $POST$):

$$\log\left(\frac{P(BIAS_{it}=1)}{1-P(BIAS_{it}=1)}\right) = \text{Intercept} + \eta_{PLAN} \cdot PLAN_{it} + \eta_{POST} \cdot POST_{it} + v_i, \qquad (2.3)$$

where SKUs are indexed by i and months by t. $P(BIAS_{it} = 1)$ denotes the probability that the forecast of product i is positively biased in month t. The error term v_i captures product-specific deviations ($v_i \sim \mathcal{N}(0, \sigma_v^2)$). Table 2.2 reports the estimated logarithmic odds from which we can calculate the probability for a positive bias by project phase. In prescription medicines, the probability of a positive bias was 58.5 % in the pre-announcement phase, which reduced to 48.2 % in the planning phase (z-test, $p = 0.002$) and to 52.1 % in the post-implementation phase (z-test, $p = 0.017$). The estimates of the reduced dataset excluding the months of May and June are similar. In consumer products, the probability of a positive bias was 69.6 % in the pre-announcement phase, which reduced to 61.3 % in the planning phase (z-test, $p = 0.016$) and to 55.6 % in the post-implementation phase (z-test, $p < 0.001$). Hence, even though the forecast bias was not explicitly incentivized, it was significantly reduced as a result of the forecast accuracy incentives.

	Prescription medicines		Consumer products
Model	Full	Excl. May/June	Full
Observations	1,911	1,657	1,450
Groups	76	76	55
Intercept	0.342 *** (0.104)	0.328 *** (0.106)	0.827 *** (0.106)
PLAN	-0.413 *** (0.131)	-0.400 *** (0.135)	-0.367 ** (0.152)
POST	-0.258 ** (0.108)	-0.237 ** (0.118)	-0.604 *** (0.127)

Note: Significance of estimates (z-test): *** p < 0.01, ** p < 0.05, * p < 0.1.
Standard errors are reported in parentheses.

Table 2.2.: Estimation results of forecast bias by project phase

As the above analyses show, the introduction of forecast accuracy incentives in the pilot country led to improvements in both forecast accuracy and forecast bias. Even though the monetary incentive was comparably small, not only the quality of the forecasts but also the overall awareness for the importance of good forecasts increased, as interviews of the post-implementation phase revealed.

> *"The weight [on the forecast accuracy target] is still too low. However, for the break-through moment it was good. For the first time, product managers are con-*

> *cerned about the quality of their forecasts. We can see that the forecast accuracy is improving."*
>
> (Head of supply management, pilot country)

> *"The most important point was to include forecast accuracy in the objectives of product managers. This showed that it matters."*
>
> (Team leader product management, prescription medicines, pilot country)

In addition, it was made clear that the possibilities for incentivizing forecast accuracy would always be limited due to the importance of other targets in the incentive systems of product managers.

> *"You can never put a lot of weight on forecast accuracy in the variable compensation of product managers. Their task is to manage a brand, so there will never be a large bonus payment linked to the forecast accuracy target. But the crucial factor is that they now have to discuss their achievements with their boss and nobody likes to present an objective that he did not meet."*
>
> (Head of product management, prescription medicines, pilot country)

The case example illustrates that incentives for forecast accuracy can have a positive effect on the quality of demand forecasts, even if the monetary impact is comparably small. We could witness an improvement in forecast accuracy and a reduction in forecast bias. However, the forecast accuracy is still not at target level and there is still some systematic bias in the forecasts. Furthermore, there are uncontrollable factors in a single-case field study, which could have affected the above results. We take the example of PharmaCo therefore as an inspiration to approach the problem by formal model building and controlled laboratory experiments in order to increase the reliability of the conclusions. We will next translate the information exchange between product management and demand management into a game-theoretic model (Chapter 3) and test it in a series of experiments (Chapter 4).

3. Game-Theoretic Model

3.1. Model Development

We consider a company where Sales is responsible for demand forecasting and Operations is responsible for ordering. Sales has better information about the market demand than Operations and provides a demand forecast to Operations. Based on the demand forecast, Operations decides on the order quantity. We will next describe the demand model and sequence of events in detail (Section 3.1.1), derive the utility functions of Sales (Section 3.1.2) and Operations (Section 3.1.3), and consider the perspective of the company (Section 3.1.4).

3.1.1. Demand Model and Sequence of Events

The demand of the product $D = \Phi + E$ is stochastic and consists of two components, a market condition Φ and a symmetric market uncertainty E with mean zero. Sales and Operations both know the distributions of Φ and E, but only Sales knows the realization of the market condition ϕ. The distribution function of the market condition Φ is $F(\cdot)$ and the density function is $f(\cdot)$. The distribution function of the market uncertainty E is $G(\cdot)$ and the density function is $g(\cdot)$. The distribution function of demand is the convolution of the market condition and market uncertainty, i.e., $H(\cdot) = F(\cdot) \circ G(\cdot)$. All distribution functions are continuous, continuously differentiable, and supported on $\mathbb{R}$.

The sequence of events is as follows (see Figure 3.1): First, nature draws a realization of the market condition ϕ. Sales observes the market condition ϕ and sends a nonbinding demand forecast $\hat{\phi}$ to Operations. Operations receives the forecast $\hat{\phi}$, updates the belief about the distribution of the market condition Φ, and orders quantity q. Then, nature draws a realization of the market error ε. If the demand $d = \phi + \varepsilon$ is less than or equal to the order quantity q, all demand is filled and $q - d$ units are left over (overage quantity). If the demand d is greater than the order quantity q, q units of the demand are filled and $d - q$ units cannot be filled (underage quantity). Finally, Sales and Operations receive their compensations.

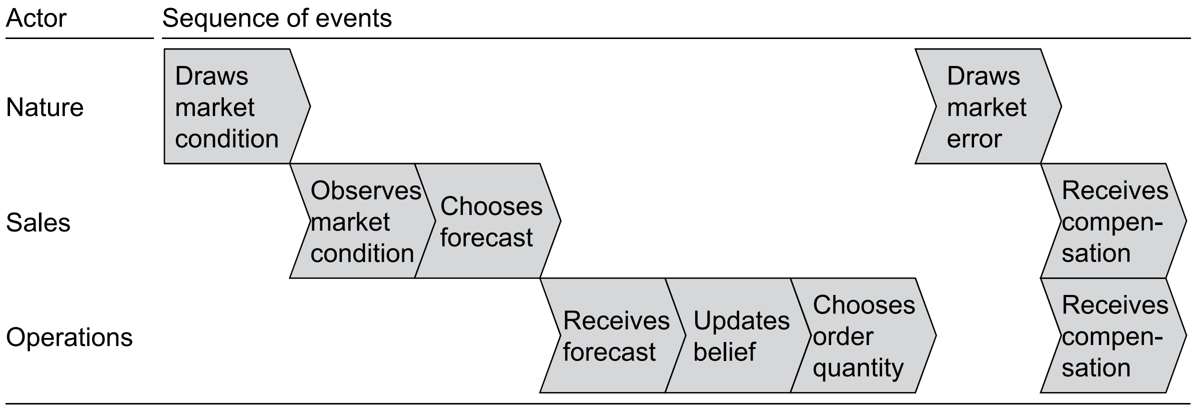


Figure 3.1.: Sequence of events

3.1.2. Utility of Sales

Expected payoff function.

The incentive system of Sales consists of a fixed compensation, a sales bonus, and a forecast error penalty. The *fixed compensation* is denoted by $C_S > 0$ and is independent of sales and demand forecasts. The *sales bonus* is proportional to the number of units sold and Sales receives a unit sales bonus of $b > 0$. The *forecast error penalty* is proportional to the deviation of the demand forecast $\hat{\phi}$ from the demand realization d. For each unit that the demand forecast $\hat{\phi}$ exceeds the demand realization d, a unit overforecasting penalty of $p_o \geq 0$ is subtracted from the compensation. For each unit that it falls below the realized demand, a unit underforecasting penalty of $p_u \geq 0$ is subtracted. Given market condition ϕ, the expected payoff of Sales is

$$\Pi_S(q, \hat{\phi} \mid \phi) = C_S + b\mathbb{E}_E \min(\phi + E, q) - \mathbb{E}_E \left(p_o[\hat{\phi} - (\phi + E)]^+ + p_u[(\phi + E) - \hat{\phi}]^+ \right). \quad (3.1)$$

The behavioral OM literature has identified various behavioral factors that affect decision makers' utilities. We next present a model that uses two behavioral factors to address important individual and social preferences and, as we will see, that can explain and predict actual forecasting and ordering behavior quite well. There are other behavioral aspects that could be relevant in our setting, some of which can be captured by the factors that we use and others that require modifications or extensions of the model. We postpone this discussion to Section 5.1.

Forecast error penalty factor.

An expected-payoff-maximizing decision maker is indifferent to the values of the individual components of the compensation and bases decisions on the total expected payoff (Equation 3.1). However, actual decision makers do not necessarily base decisions on total expected payoffs, but might form *mental accounts* and evaluate elementary outcomes or sets of elementary outcomes (Thaler, 1985, 1999; Kőszegi and Rabin, 2006). If elementary outcomes are segregated, a value is assigned to each elementary outcome and the evaluation of the total outcome is based on the sum of the values of the elementary outcomes. If elementary outcomes are integrated, the sum of the elementary outcomes (i.e., the total payoff) is valued. The concept of mental accounting has proven to be a powerful tool to explain human behavior in OM decisions (e.g., Ho and Zhang, 2008; Katok and Wu, 2009; Becker-Peth et al., 2013).

Thaler (1985, 1999) uses the value function of *prospect theory* (Kahneman and Tversky, 1979) to model the evaluation of outcomes. The function assigns value $v(x)$ to an (elementary or aggregate) outcome x relative to a reference point, where positive and negative deviations from this reference point have different values. In particular, prospect theory establishes that "losses loom larger than gains" (Kahneman and Tversky, 1979, p. 279). The disutility that a subject experiences from losing is greater than the utility experienced from gaining the same amount. The nontrivial issue in reference-dependent utility models is defining the reference point. While some researchers use the status quo, i.e., the endowment at the time of making a decision (e.g., Kahneman et al., 1990), others suggest that reference points are based on expectations (Kőszegi and Rabin, 2006), goals (Heath et al., 1999) or other salient pieces of information (Ockenfels et al., 2015).

For our incentive system of Sales, the elementary outcomes are the fixed compensation, sales bonus, and forecast error penalties. Whether elementary outcomes are segregated or integrated in the evaluation process depends on the situation and the frame of the decision problem (Thaler, 1985). Because theory does not provide clear guidance on how people evaluate outcomes in a setting like ours, we conducted a pre-experiment with 48 students at the University of Cologne using Thaler's (1985) "who is happier?" approach.

PRE-EXPERIMENT *Mr. A and Mr. B work for the sales division of a company. At the beginning of each month, Mr. A and Mr. B must provide a demand forecast for the following month. At the end of the month, they both receive a fixed compensation and a performance-based compensation. Last month, Mr. A provided a demand forecast of 1,500 units, demand was 1,000 units, and 1,000 units were sold. He receives a sales bonus of EUR 100 for the quantity sold in addition to his regular salary. Mr. B also provided a demand forecast of 1,500 units, demand was 1,000 units, and 1,000 units were sold. He receives a sales bonus of EUR 150 for the quantity sold, minus a penalty of EUR 50 for the deviation between the demand forecast and actual demand, i.e., Mr. B also receives EUR 100 in addition to his regular salary.*

Who is happier? Mr. A (40), Mr. B (1), no difference (7).

Although the total payoffs of Mr. A and Mr. B are the same, 83 % of subjects believe that Mr. A, who receives only a sales bonus, is happier than Mr. B, who receives both a sales bonus and a penalty for the forecast error. The result of the pre-experiment is in line with mental accounting practices where elementary outcomes are segregated, losses are weighted more heavily than gains and the reference point is zero.

In the more complex incentive system of Sales that we use in our model, Sales might form different kinds of (expectations-based) reference points. For example, Sales might compare the forecast decision with alternative possible forecast decisions that result in reference forecast error penalties. We discuss such alternative reference points in Section 5.1.2 and proceed with a reference point of zero. We operationalize this reference point using the linear version of the value function, $v(x) = [x]^+ - \gamma[-x]^+$, $\gamma \geq 1$, which we apply to the elementary outcomes of the payoff function (Equation 3.1).

Note that the forecast error-related part of the decision problem of Sales is similar in structure to the order decision of a newsvendor. One of the biases that has been identified in newsvendor-type situations is *ex-post inventory error minimization*, which is proposed as an explanation for pull-to-center ordering behavior (Schweitzer and Cachon, 2000; Ho et al., 2010; Kremer et al., 2014). The underlying rationale is that subjects experience a negative utility for each unit that the order quantity deviates from realized demand. An analogous pattern might exist for over- and underforecasting. In addition to the monetary forecast error penalties, human decision makers could feel a negative utility for each unit that the forecast deviates from realized demand. Following the approach of Schweitzer and Cachon (2000), we could add a term $-\omega \left| \hat{\phi} - (\phi + E) \right|$ to the payoff function of Sales, where ω is a psychological cost parameter for over- and underforecasting, which would increase the unit over- and underforecasting penalty factors p_o and p_u to $p_o + \omega$ and $p_u + \omega$. In our model, we multiply the unit over- and underforecasting penalty factors p_o and p_u by γ, which has a similar effect as adding ω to them.

An important difference between the sales bonus payoff stream and the payoffs resulting from forecast error penalties is not captured by our pre-experiment. This difference goes back to the findings of Ellsberg (1961), who distinguishes between *risk* (i.e., uncertainty that can be represented by measurable probabilities) and *ambiguity* (i.e., uncertainty with unknown probabilities). For any choice of forecast, Sales can quantify the risk of over- and underforecasting because the probability distribution of the market error E and hence the probability distribution of the forecast error are known. With respect to the sales bonus, however, the outcome depends not only on the distribution of E but is also determined by the decision of Operations. Since the order decision of Operations is subject to individual beliefs, preferences and biases, the sales bonus is subject to ambiguity. There exists a large body of theoretical models and experimental evidence, which shows that most people are ambiguity averse (for a review, see Camerer and Weber, 1992). The literature suggests that ambiguity aversion is prevalent in strategic games, where the ambiguity aversion relates to preferences and payoffs, and hence the actions of the opponent player (Pulford and Colman, 2007; Eichberger et al., 2008; Kelsey and le Roux, 2015). Ambiguity aversion can influence decision making in our setting. A higher valuation of risky payoffs from forecast error penalties compared to ambiguous payoffs from sales bonus payments would be reflected in a higher behavioral parameter γ.

The discussion indicates that there are multiple theories that suggest a stronger relative effect of forecast error penalties on the utility of Sales compared to sales bonuses. We do not differentiate between the individual psychological drivers but rather quantify their combined effect in the *forecast error penalty factor* γ.

Lying aversion factor.

Human decisions are affected by concerns about the well-being of others (e.g., Rabin, 1998; Fehr and Falk, 2002; Cooper and Kagel, 2016). Particularly relevant for our setting is the insight that people are more trustworthy than standard theory suggests because they experience a disutility when lying to others (Gneezy, 2005; Charness and Dufwenberg, 2006; Vanberg, 2008; Lundquist et al., 2009; Hurkens and Kartik, 2009; Erat and Gneezy, 2012; Gneezy et al., 2013; López-Pérez and Spiegelman, 2013). There has been extensive discussion in the economics literature about whether and how *lying aversion* can be distinguished from other social preferences, such as fairness and inequity aversion (e.g., Gneezy, 2005) as well as analyses of the underlying motives for this behavior, such as altruism, guilt aversion or belief-dependent lying aversion (e.g., López-Pérez and Spiegelman, 2013). Engaging in these discussions is beyond the scope of this thesis. Instead, we build on the robust finding that people tend to be more honest in social interactions than standard theory suggests, even in one-shot interactions like the one that we consider, where being trustworthy does not "pay" monetarily (Ashraf et al., 2006, p. 194). This phenomenon has also been observed in supply chain settings, where one party shares private information that another party needs to make a decision (e.g., Özer et al., 2011, 2014; Inderfurth et al., 2013; Beer et al., 2017). To capture the disutility of lying in our setup, we follow the modeling approach of Özer et al. (2011) and include the term $-\beta\left|\hat{\phi}-\phi\right|$ in the utility function of Sales. We refer to the factor $\beta \geq 0$ as the *lying aversion factor*.

Utility function.

With the above three components, we obtain the utility function of Sales:

$$U_S(q, \hat{\phi} \mid \phi) = C_S + b\mathbb{E}_E \min(\phi + E, q) - \gamma \mathbb{E}_E \left(p_o[\hat{\phi} - (\phi + E)]^+ + p_u[(\phi + E) - \hat{\phi}]^+ \right) - \beta \left| \hat{\phi} - \phi \right| . \tag{3.2}$$

We refer to the model with $\gamma = 1$ and $\beta = 0$ as the *standard model* and to models with $\gamma > 1$ or $\beta > 0$ as *behavioral models*.

In contrast to other models in the literature on sales force incentives (e.g., Albers, 1996), we have not included an *effort parameter* in the utility function of Sales for two reasons. First, in many industry contexts (including the example of PharmaCo in Chapter 2), demand is comparably stable and can be influenced by means of sales and marketing efforts only in the long term, whereas production and supply planning is done on a shorter time horizon. Our model therefore reflects a scenario where Sales has already invested time and effort to shape demand before the operational forecast is created. Second, our focus is on how behavioral factors influence the reaction of Sales to different forecast incentive schemes. To separate these effects from others, we focus on the trade-off between sales volume and forecast accuracy.

Incentive systems.

We consider three types of incentive systems for Sales. All incentive systems use the fixed compensation C_S and the unit sales bonus b, but differ in how they penalize forecast errors. In a *sales-bonus-only* incentive system, forecast errors are not penalized ($p_o = p_u = 0$). This incentive system corresponds to the initial incentive system at PharmaCo and the one analyzed by Özer et al. (2011). In an *absolute forecast error* incentive system, the absolute deviation between demand and forecast is penalized, i.e., over- and underforecasting are penalized at equal rates ($p_o = p_u > 0$). Such incentive systems are frequently used in practice (e.g., Reese, 2001; Dershem, 2007; McKenzie, 2011). Even though PharmaCo used a forecast accuracy measure instead of a forecast error measure, the new incentive system at the pilot sales unit was similar to an absolute forecast error incentive system because over- and underforecasting were also penalized at equal rates. In a *differentiated forecast error* incentive system, overforecasting is penalized more heavily than underforecasting ($p_o > p_u$). Even though this approach has been suggested in the literature (Gonik, 1978; Mantrala and Raman, 1990; Celikbas et al., 1999; Chen, 2005; Chen et al., 2016), such incentive systems do not seem to be used in practice except in rare cases (e.g., Gonik, 1978; Turner et al., 2007). However, as we will see, they can be more effective and efficient than the other incentive systems to incentivize truthful information sharing.

3.1.3. Utility of Operations

The incentive system of Operations consists of a fixed compensation and penalties for unfilled demand and leftovers. The *fixed compensation* of the incentive system is $C_O > 0$. The *penalty for unfilled demand* is proportional to the demand that cannot be filled. For each unit that the demand d exceeds the order quantity q, a unit underage cost of $c_u > 0$ is subtracted from the payoff. The *penalty for leftovers* is proportional to the number of units that are left over after demand has been filled. For each unit that the order quantity q exceeds the demand d, a unit overage cost of $c_o > 0$ is subtracted from the payoff. Following similar arguments as for the development of the expected utility function of Sales (see also Schweitzer and Cachon, 2000), we multiply the expected overage and underage penalties by the behavioral factor $\gamma_O \geq 1$ and obtain the following expected utility function of Operations:

$$U_O(q) = C_O - \gamma_O \mathbb{E}_D \left(c_o[q - D]^+ + c_u[D - q]^+ \right). \tag{3.3}$$

If Operations knew the actual market condition ϕ, the only uncertainty would be the market uncertainty E and the optimal order quantity would be

$$q^{FB}(\phi) = \phi + G^{-1}(\alpha), \tag{3.4}$$

where $\alpha = c_u/(c_u + c_o)$ is the critical ratio of Operations. We refer to $q^{FB}(\phi)$ as the first-best order quantity. Note that this quantity (as well as all other equilibrium order quantities that we will analyze later) is independent of γ_O.

3.1.4. Company Perspective

While our primary objective is to design incentive systems for truthful forecast information sharing, we next discuss the conditions under which such incentive systems also maximize the expected profit of the company. The *unit profit margin* is m and the unit overage and underage cost of the company are c_o^C and c_u^C, respectively. Taking into account that the company has to pay Sales and Operations their fixed and variable compensations, the expected profit of the company is

$$\begin{aligned}\Pi_C(q,\hat{\phi}) =& m\mathbb{E}_D(D) - c_o^C\mathbb{E}_D[q-D]^+ - c_u^C\mathbb{E}_D[D-q]^+ \\ &- \Big(C_S + b\mathbb{E}_D\min(D,q) - \mathbb{E}_D\left(p_o[\hat{\phi}-D]^+ + p_u[D-\hat{\phi}]^+\right)\Big) \\ &- \left(C_O - \mathbb{E}_D\left(c_o[q-D]^+ + c_u[D-q]^+\right)\right).\end{aligned} \tag{3.5}$$

We consider situations where the unit overage and underage costs of Operations are aligned with the objectives of the company ($c_o = \kappa c_o^C$ and $c_u = \kappa c_u^C$ with factor $\kappa > 0$) such that the expected-utility-maximizing solution of Operations q^{FB} maximizes the expected profit of the company.

The company faces a principal-agent problem, where the agents (Sales and Operations) might have participation constraints or reservation utilities, such as those resulting from outside options (e.g., Prendergast, 1999). Let r_S and r_O denote these reservation utilities. The principal's (company's) objective is to design an incentive system that maximizes the principal's expected profit Π_C while meeting the agents' participation constraints ($U_S \geq r_S$ and $U_O \geq r_O$).

If we neglect the behavioral factors in the utility functions of Sales and Operations (standard model), payoffs could be allocated arbitrarily between the company, Sales, and Operations. The company could maximize its expected profit by calibrating the incentive system such that the participation constraints are binding ($U_S = r_S$ and $U_O = r_O$). It could change the incentive parameters b, p_o, p_u, c_o and c_u to incentivize first-best order quantities and adapt the fixed compensations C_S and C_O to meet the participation constraints. The optimal profit of the company would be $\Pi_C(q^{FB}) = m\mathbb{E}_D(D) - c_o^C\mathbb{E}_D[q^{FB}-D]^+ - c_u^C\mathbb{E}_D[D-q^{FB}]^+ - r_S - r_O$.

When taking the behavioral factors into account, the valuation of the variable compensation elements is different for Sales and Operations than for the company. In that case, incentivizing first-best order quantities is not necessarily optimal for the company. To illustrate this, suppose that the participation constraints of Sales and Operations are binding for a given incentive system that leads to non-first-best order quantities. Suppose further that increasing the unit forecast error penalties of Sales leads to a change in forecasting behavior that allows Operations to make a first-best order decision. This change in incentives changes the expected variable compensation of Sales by $\Delta\Pi_S$ and increases the expected variable compensation of Operations by $\Delta\Pi_O$ and the expected operational profit of the company by $\Delta\Pi_C$. In order to keep the participation constraints satisfied and binding, the company must adapt the fixed compensations of Sales and Operations, i.e., it must change the fixed compensation of Sales C_S by $-\gamma\Delta\Pi_S$ and decrease the fixed compensation of Operations C_O by $-\gamma_O\Delta\Pi_O$. Hence, changing from a non-first-best to a first-best incentive system increases the company's expected profit only if the expected increase in the company's profit $\Delta\Pi_C$ exceeds the additional cost of the incentive system, i.e., if $\Delta\Pi_C > (\Delta\Pi_S - \gamma\Delta\Pi_S) + (\Delta\Pi_O - \gamma_O\Delta\Pi_O)$. To determine whether implementing such an incentive system is beneficial for the company, one would have to compute the expected profit of the current and the first-best solution and analyze whether the inequality is satisfied.

The inequality holds if the changes in the compensations of Sales and Operations are much smaller than the change in the profit of the company. We consider such situations, and require that the overage and underage costs of the company (c_o^C and c_u^C) be considerably higher than the variable incentive parameters (b, p_o, p_u, c_o, c_u). This assumption seems reasonable for real applications. However, if it does not hold, incentivizing first-best order quantities is not necessarily optimal from a company perspective.

In Section 3.2.2, we will show that different parameterizations of the incentive system of Sales result in first-best order quantities. Under some parameterizations, Sales distorts the demand forecast, but Operations is aware of this distortion and corrects the forecast to determine the actual market condition. Under other parameterizations, Sales reports the demand forecast truthfully and Operations can rely on the forecast without corrections. We are interested in identifying incentive systems where the latter case applies, i.e., where forecast information is shared truthfully. Such incentive systems are appealing because the sales forecast is often communicated and used beyond the organizational boundaries of Sales and Operations, where people might not be aware of the distortion and correction activities. If forecasts are truthful, they can be used by other departments for planning purposes without causing further biases. This aspect is not included in the utility and cost functions of our model, but it implies that an incentive system with first-best order quantities and truthful demand forecasts should be preferred over one with distorted forecasts.

3.2. Model Analysis

The setting that we consider is a two-stage dynamic game with incomplete information, where all the parameter values are common knowledge except for the realization of the market condition ϕ. Because the informed player (Sales) moves first (by sending a demand forecast), our setting corresponds to a signaling game. We next describe the equilibrium concept that we use (Section 3.2.1) and then analyze the separating equilibria of the game (Section 3.2.2).

3.2.1. Equilibrium Concept

To solve the signaling game that we consider, we use the concept of *perfect Bayesian equilibrium* (PBE) (for a formal definition, see Fudenberg and Tirole, 1991, pp. 324-326). This equilibrium concept has previously been employed in a similar context by Özer et al. (2011). In such an equilibrium, Sales chooses a demand forecast that maximizes the expected utility of Sales and takes into account the response strategy of Operations. Operations updates the belief about the market condition based on Bayes' rule whenever possible and chooses the order quantity that maximizes the expected utility of Operations, taking into account the (updated) belief of the market condition.

Let σ_S be a signaling strategy of Sales, where $\sigma_S(\hat{\phi} \mid \phi)$ denotes the probability that Sales chooses forecast $\hat{\phi}$ given a market condition ϕ $\left(\int_{\hat{\phi}} \sigma_S(\hat{\phi} \mid \phi) = 1\right)$. Let μ be a belief system of Operations, where $\mu(\phi \mid \hat{\phi})$ represents the belief that the market condition is ϕ given a demand forecast $\hat{\phi}$. Finally, let σ_O be an ordering strategy of Operations, where $\sigma_O(q \mid \hat{\phi})$ denotes the probability that Operations chooses quantity q in response to a forecast $\hat{\phi}$ $\left(\int_q \sigma_O(q \mid \hat{\phi}) = 1\right)$.

A PBE is a set of strategies σ_S and σ_O and posterior belief system μ that satisfy the following conditions:

$$\begin{aligned}
&\text{(i)} \quad \text{For all } \phi, \hat{\phi}: \quad \sigma_S(\hat{\phi} \mid \phi) > 0 \text{ if } \int_q \sigma_O(q \mid \hat{\phi}) U_S(q, \hat{\phi} \mid \phi) dq \\
&\qquad\qquad = \max_{\hat{\phi}' \in \mathbb{R}} \int_q \sigma_O(q \mid \hat{\phi}') U_S(q, \hat{\phi}' \mid \phi) dq, \\
&\text{(ii)} \quad \text{for all } q, \hat{\phi}: \quad \sigma_O(q \mid \hat{\phi}) > 0 \text{ if } \int_\phi \mu(\phi \mid \hat{\phi}) U_O(q \mid \phi) d\phi = \max_{q' \in \mathbb{R}} \int_\phi \mu(\phi \mid \hat{\phi}) U_O(q' \mid \phi) d\phi, \\
&\text{(iii)} \quad \text{for all } \phi, \hat{\phi}: \quad \mu(\phi \mid \hat{\phi}) = \frac{f(\phi)\sigma_S(\hat{\phi} \mid \phi)}{\int_{\phi' \in \mathbb{R}} f(\phi')\sigma_S(\hat{\phi} \mid \phi') d\phi'} \text{ if } \int_{\phi' \in \mathbb{R}} f(\phi')\sigma_S(\hat{\phi} \mid \phi') d\phi' > 0, \\
&\qquad\qquad \text{otherwise, } \mu \text{ is any probability distribution on } \mathbb{R}.
\end{aligned}$$

Condition (i) states that Sales chooses a demand forecast with positive probability if it maximizes the expected utility of Sales, taking into account the response strategy of Operations. Condition (ii) ensures that for any demand forecast sent by Sales, Operations chooses an order quantity that maximizes the expected utility of Operations, taking into account the (updated) belief of the market condition. Condition (iii) states that Operations updates the belief about the market condition based on Bayes' rule whenever possible.

In the remainder of this thesis, we restrict our analyses to differentiable signaling strategies (see the discussion in Mailath and von Thadden, 2013). To be more precise, we assume that signaling strategies are continuous everywhere and differentiable everywhere except for a countable number of points. For example, we expect that differentiability does not hold at points where behavior switches from overforecasting to underforecasting or vice versa since the utility function of Sales is not differentiable at such points.

Various types of equilibria can exist for the signaling game that we analyze (e.g., Sobel, 2009). In a *pooling equilibrium*, the demand forecast is uninformative about the market condition and Operations does not gain any insights beyond the prior belief. In a *partially separating equilibrium*, Sales sends with positive probability the same demand forecast for some, but not all market conditions, such that Operations can update the prior belief based on the demand forecast, but cannot distinguish market conditions with certainty. In a *separating equilibrium*, Sales sends different demand forecasts for each market condition, which allow Operations to infer the market condition with certainty. As laid out in Section 3.1.4, it is in the interest of the company to enable first-best order quantities. Therefore, our focus is on separating equilibria, where the signaling strategy of Sales is invertible, such that Operations can infer the true market condition from the demand forecast. In such an equilibrium, the utility function of Operations U_O is strictly concave in the order quantity q and the associated optimal order quantity is unique (see Proof of Theorem 1 in Appendix A.1). Since the utility function of Sales U_S is strictly concave in $\hat{\phi}$ for a given $q(\hat{\phi})$, we can limit our attention to pure strategies of Sales.

We define a *separating PBE* as the set of an invertible signaling function s^*, a response function q^*, and posterior belief system $\mu^*(\phi \mid \hat{\phi})$ that satisfy the following conditions:

$$\begin{aligned} &\text{(i)} && \text{For all } \phi: && s^*(\phi) \in \underset{\hat{\phi}}{\operatorname{argmax}} \left(U_S(q^*(\hat{\phi}), \hat{\phi} \mid \phi) \right), \\ &\text{(ii)} && \text{for all } \hat{\phi}: && q^*(\hat{\phi}) \in \underset{q}{\operatorname{argmax}} \left(\int_\phi \mu^*(\phi \mid \hat{\phi}) U_O(q \mid \phi) d\phi \right), \\ &\text{(iii)} && \text{for all } \phi, \hat{\phi}: && \mu^*(\phi \mid \hat{\phi}) = \begin{cases} 1 \text{ for } \hat{\phi} = s^*(\phi), \\ 0 \text{ otherwise.} \end{cases} \end{aligned}$$

A special case of separating equilibrium is a *truth-telling separating equilibrium*, where (i) Sales communicates a truthful demand forecast ($s^*(\phi) = \phi$), (ii) Operations determines the expected-utility-maximizing order quantity based on this demand forecast ($q^*(\hat{\phi}) = \hat{\phi} + G^{-1}(\alpha)$) and (iii) Operations believes that Sales reports the market condition truthfully ($\mu^*(\phi \mid \hat{\phi}) = 1$, if $\phi = \hat{\phi}$ and 0 otherwise). As discussed in Section 3.1.4, we will place particular emphasis on the analysis of incentive systems that result in truth-telling separating equilibria.

3.2.2. Analysis of Equilibria

The following theorem describes the conditions under which a Pareto-dominant separating equilibrium exists. In a Pareto-dominant equilibrium, all the players are at least as well off as in all the other equilibria and at least one (but not necessarily always the same) player is strictly better off relative to each of the other equilibria. This definition implies that if a Pareto-dominant equilibrium exists, it is unique. Since a Pareto-dominant equilibrium is a focal equilibrium, we cannot be sure but have reason to believe that players coordinate on this equilibrium (Stamland, 1999; Wang, 2006; Hyndman et al., 2013). All proofs are given in Appendix A.1.

Theorem 1. *If $p_o > (b(1-\alpha)-\beta)/\gamma$, there exists a Pareto-dominant separating equilibrium $(s^{sep}, q^{sep}, \mu^{sep})$:*

$$\begin{array}{lll} (i) & \text{For all } \phi: & s^{sep}(\phi) = \phi + \delta^{sep}, \\ (ii) & \text{for all } \hat{\phi}: & q^{sep}(\hat{\phi}) = \hat{\phi} - \delta^{sep} + G^{-1}(\alpha), \\ (iii) & \text{for all } \phi, \hat{\phi}: & \mu^{sep}(\phi \mid \hat{\phi}) = \begin{cases} 1 & \text{for } \phi = \hat{\phi} - \delta^{sep}, \\ 0 & \text{otherwise,} \end{cases} \end{array}$$

with distortion value

$$\delta^{sep} = \begin{cases} G^{-1}\left(\frac{b(1-\alpha)+\gamma p_u-\beta}{\gamma(p_u+p_o)}\right) & \text{for } p_o - p_u < 2\frac{b(1-\alpha)-\beta}{\gamma}, \\ 0 & \text{for } 2\frac{b(1-\alpha)-\beta}{\gamma} \le p_o - p_u \le 2\frac{b(1-\alpha)+\beta}{\gamma}, \\ G^{-1}\left(\frac{b(1-\alpha)+\gamma p_u+\beta}{\gamma(p_u+p_o)}\right) & \text{for } p_o - p_u > 2\frac{b(1-\alpha)+\beta}{\gamma}. \end{cases} \tag{3.6}$$

Otherwise, if $p_o \le (b(1-\alpha)-\beta)/\gamma$, there exists no separating equilibrium.

An important implication of Theorem 1 is that the demand forecast is distorted by a distortion value δ^{sep} that is independent of the market condition ϕ. Equation (3.6) distinguishes three cases that depend on the difference between the unit over- and underforecasting penalty. In the first case, the difference is small and Sales inflates the demand forecast ($\delta^{sep} > 0$). In the second case, the difference is medium and Sales provides a truthful demand forecast

($\delta^{sep} = 0$). In the third case, the difference is large and Sales deflates the demand forecast ($\delta^{sep} < 0$). In all cases, Operations anticipates the demand forecast distortion of Sales and corrects the forecast such that the order decision is based on the true market condition and hence $q^{sep} = q^{FB}$. Since our focus is on truthful forecast information sharing, we are particularly interested in finding incentive systems that fulfill the condition of the second case. Note that this condition can be achieved by many different incentive systems. For a given sales bonus b, for example, all differentiated forecast error incentive systems with $p_o - p_u = 2b(1-\alpha)/\gamma$ lead to truthful information sharing independently of the behavioral parameter β. For $\beta > 0$, truthful information sharing can even be achieved by a range of differences $p_o - p_u$.

In the proof of Theorem 1, we show that this equilibrium Pareto dominates all other separating equilibria. In particular, we show that if any other separating equilibrium exists, it must be based on a signaling strategy s that results in forecasts $s(\phi) \geq s^{sep}(\phi)$ for all $\phi \in \mathbb{R}$ with strict inequality for all ϕ in some interval $I \subseteq \mathbb{R}$.

We also show that Sales has a unique preferred forecasting strategy $s^{pref}(\phi) = \phi + \delta^{pref}$ with distortion value

$$\delta^{pref} = \begin{cases} G^{-1}\left(\frac{\gamma p_u - \beta}{\gamma(p_u + p_o)}\right) & \text{for } p_o - p_u < -\frac{2\beta}{\gamma}, \\ 0 & \text{for } -\frac{2\beta}{\gamma} \leq p_o - p_u \leq \frac{2\beta}{\gamma}, \\ G^{-1}\left(\frac{\gamma p_u + \beta}{\gamma(p_u + p_o)}\right) & \text{for } p_o - p_u > \frac{2\beta}{\gamma}, \end{cases} \tag{3.7}$$

that maximizes the forecast-dependent part of Sales' utility function

$$U_S^{FC}(\hat{\phi} \mid \phi) = -\gamma \mathbb{E}_E \left(p_o[\hat{\phi} - (\phi + E)]^+ + p_u[(\phi + E) - \hat{\phi}]^+\right) - \beta \left|\hat{\phi} - \phi\right|. \tag{3.8}$$

Comparing Equations (3.6) and (3.7), it is straightforward to see that for any market condition ϕ the equilibrium forecast $s^{sep}(\phi)$ is weakly higher than the preferred forecast $s^{pref}(\phi)$. Strict concavity and unimodality of U_S^{FC} further imply that sending signal $s^{sep}(\phi)$ is also weakly more costly for Sales than sending signal $s^{pref}(\phi)$. We only have $\delta^{sep} = \delta^{pref}$ (and hence minimal expected signaling costs in equilibrium) if $2(b(1-\alpha) - \beta)/\gamma \leq p_o - p_u \leq 2\beta/\gamma$, i.e., if $\delta^{sep} = \delta^{pref} = 0$. This requires a sufficiently large lying aversion of $\beta \geq b(1-\alpha)$. Hence, even though the equilibrium of Theorem 1 Pareto dominates all other separating equilibria, coordination between the two parties is always costly for Sales if the lying aversion is small.

The necessary condition for the existence of a separating equilibrium is for the unit overforecasting penalty p_o to be above the threshold cost $\bar{p} = (b(1-\alpha)-\beta)/\gamma$, which can be rewritten as $b(1-\alpha) < \gamma p_o + \beta$. The left-hand side of the inequality is $\partial U_S/\partial q$ evaluated at $q = q^{FB}(\phi)$. The right-hand side is the limit of $-\partial U_S/\partial\hat{\phi}$ as $\hat{\phi}$ goes to infinity. Hence, for any Sales type ϕ in a separating equilibrium, we require the incremental gain from an order quantity q that is greater than the equilibrium order quantity $q^{FB}(\phi)$ to be strictly smaller than the maximum incremental cost of sending a signal $\hat{\phi} > s^{sep}(\phi)$. If this condition holds for the standard model ($\gamma = 1$ and $\beta = 0$), it also holds for all other values $\gamma > 1$ and $\beta > 0$, which leads to the following corollary:

Corollary 1. *If a separating equilibrium exists for the standard model, there exists a separating equilibrium for the behavioral model.*

If the lying aversion is large compared to the unit sales bonus ($\beta \geq b(1-\alpha)$), the condition $p_o > \bar{p}$ holds for all overforecasting penalties $p_o \geq 0$. Then, the disutility of lying associated with each unit of forecast inflation ($-\beta$) exceeds the expected gain ($b(1-\alpha)$) independently of the forecast error incentives. For smaller values of the lying aversion factor ($\beta < b(1-\alpha)$), only forecast error incentive systems with $p_o > 0$ can fulfill the necessary condition for the existence of an equilibrium according to Theorem 1. The case distinctions for δ^{sep} further imply that absolute forecast error incentive systems lead to inflated forecasts in equilibrium, while differentiated forecast error incentive systems can be parameterized such that a truth-telling equilibrium exists. We summarize the different characteristics of the incentive systems in the following corollary:

Corollary 2. *For a sufficiently small lying aversion ($\beta < b(1-\alpha)$), the following properties hold:*

(a) *Sales-bonus-only incentive systems do not result in a separating equilibrium.*

(b) *Absolute forecast error incentive systems can be parameterized such that a separating equilibrium exists, but the demand forecasts are always inflated* ($\delta^{sep} > 0$)*. The equilibrium forecast inflation* δ^{sep} *approaches zero as* $p_o = p_u$ *go to infinity.*

(c) *Differentiated forecast error incentive systems can be parameterized such that a truthful separating equilibrium exists* ($\delta^{sep} = 0$)*.*

Theorem 1 also implies that differentiated forecast error incentive systems can achieve any level of forecast distortion more economically than absolute forecast error incentive systems, which results in the following corollary:

Corollary 3. *For every absolute forecast error incentive system with equilibrium forecast distortion* δ^{sep}*, there exists a differentiated forecast error incentive system with the same equilibrium forecast distortion* δ^{sep} *with lower unit forecast error penalties* p_o *and* p_u*, and hence with a lower expected forecast error cost for Sales.*

Our theoretical analyses provide insights into the expected effects of the incentive system on the forecast decisions of Sales and the order decisions of Operations. In the next section, we derive hypotheses about the expected behavior.

3.3. Hypotheses

In the development of our model, we included two behavioral factors: a forecast error penalty factor γ and a lying aversion factor β. We argued that loss aversion and other behavioral biases lead to a forecast error penalty factor γ that is greater than one. The disutility that people experience from reporting information untruthfully suggests that the lying aversion factor β is greater than zero, which is also what Özer et al. (2011) find in their experiments. If our modeling assumptions hold and a separating equilibrium exists ($p_o > \bar{p} = (b(1-\alpha) - \beta)/\gamma$), a model with $\gamma > 1$ and $\beta > 0$ explains the actual forecasts of Sales and order decisions of Operations better than a model with $\gamma = 1$ and $\beta = 0$, which is stated in the following hypothesis:

Hypothesis 1. *For incentive systems with* $p_o > \bar{p}$*,*

(a) *a forecast error penalty factor* γ *that is greater than one, and*

(b) *a lying aversion factor* β *that is greater than zero*

explain actual behavior better than the standard model with $\gamma = 1$ *and* $\beta = 0$*.*

Hypothesis 1 allows us to test whether differences between actual behavior and the standard model predictions can be explained by the behavioral parameters and their underlying theories. We cannot state the values of the expected forecast distortions and corrections because they depend on the values of γ and β, which we do not know. However, for the special case of the standard model, we know that the equilibrium forecast distortion is

$$\delta_{std}^{sep} = G^{-1}\left(\frac{b\,(1-\alpha)+p_u}{p_u+p_o}\right). \tag{3.9}$$

Comparing Equations (3.6) and (3.9), one can observe that forecasts in a behavioral model exhibit a directional shift from the standard model solution towards the market condition (effect of β) and a general shift downwards (effect of γ). If the standard model predicts overforecasting ($p_o - p_u < 2b(1-\alpha)$), the behavioral model predicts strictly smaller, and possibly even negative, forecast distortions. If the standard model predicts truthful forecasting ($p_o - p_u = 2b(1-\alpha)$), the behavioral model predicts either the same or negative forecast distortions. If the standard model predicts underforecasting, there exists no unambiguous directional prediction of the behavioral model. We summarize these properties in the following hypothesis:

Hypothesis 2. *For incentive systems with* $p_o > \bar{p}$ *and* $p_o - p_u \leq 2b(1-\alpha)$*,*

(a) *forecast distortions* $(\hat{\phi} - \phi)$ *are (weakly) smaller, and*

(b) *forecast corrections* $(q - \hat{\phi})$ *are (weakly) greater*

than predicted by the standard model.

In addition to the directional predictions above, we can derive the structure of the forecasting and ordering behavior from Theorem 1: Sales adds δ^{sep} to all market conditions and Operations subtracts this value from the forecast. Özer et al. (2011, 2014) observe a similar pattern in their experimental data. Hence, we expect to see the following correlations:

Hypothesis 3. *For incentive systems with* $p_o > \bar{p}$*, there exist positive correlations between*

(a) *the market condition ϕ and the demand forecast $\hat{\phi}$,*

(b) *the demand forecast $\hat{\phi}$ and the order quantity q, and*

(c) *the order quantity q and the market condition ϕ.*

For incentive systems with $p_o > \bar{p}$, there exists a separating equilibrium and Operations can derive the market condition from the demand forecast. For incentive systems with $p_o \leq \bar{p}$, there does not exist a separating equilibrium, but partially separating or pooling equilibria can exist. According to the definition of these equilibria, the market condition cannot always be derived from the demand forecast, which we formulate as follows:

Hypothesis 4. *For incentive systems with $p_o > \bar{p}$, the correlations between*

(a) *the market condition ϕ and the demand forecast $\hat{\phi}$,*

(b) *the demand forecast $\hat{\phi}$ and the order quantity q, and*

(c) *the order quantity q and the market condition ϕ*

are higher than for incentive systems with $p_o \leq \bar{p}$.

We chose the incentive system of Operations such that the order quantity that maximizes Operations' profit also maximizes the profit of the company. The better the information that Operations can deduct from the demand forecast, the better the order decisions it can make. This leads to the following hypothesis:

Hypothesis 5. *Under incentive systems with $p_o > \bar{p}$, the expected cost of Operations is lower than under incentive systems with $p_o \leq \bar{p}$.*

To test the above hypotheses and design incentive systems for truthful forecast information sharing, we must know the values of the behavioral parameters γ and β. To estimate these parameters and test whether human decision makers behave structurally as stated in the hypotheses, we use a laboratory experiment.

4. Laboratory Experiments

4.1. Main Experiment

4.1.1. Experimental Design

We conducted an experiment with two sessions and 32 subjects per session. In each session, 16 subjects were assigned to Sales and 16 subjects were assigned to Operations. The design of the experiment is shown in Table 4.1. All monetary parameters are in experimental currency units (ECUs) and an exchange rate of 1,000 ECUs per EUR 1 was used. In Session 1, we exposed the subjects to sales-bonus-only and absolute forecast error incentive systems. In Session 2, we exposed them to differentiated forecast error incentive systems. Within a session, half of the subjects played the treatments in the order shown in Table 4.1 and the other half in reverse order to reduce potential order effects. Within a treatment, the subjects of each subgroup were randomly matched with another subject of the same subgroup at the beginning of each period. Each treatment lasted eight periods. The decisions of each period were independent of previous periods, i.e., excess demand of previous periods was lost and leftovers of previous periods were discarded. More details, such as screenshots, instructions and questionnaires, can be found in Section A.3 of the Appendix.

If existent, Table 4.1 reports the equilibrium distortion values δ^{sep} of the standard model and directional predictions of the behavioral model. In the design of the experiment, we rely on the standard model as opposed to the behavioral model because we cannot specify parameter values for γ and β prior to the experiment. The treatments are chosen to cover a variety of incentive schemes. We include the sales-bonus-only incentive system (Treatment 1) to analyze a treatment where both the standard model and the behavioral model predict that no separat-

Session	Treatment ($b/p_o/p_u$)	Incentive system	Distortion value δ^{sep} Standard model	Behavioral model
1	1 (16/0/0)	Sales-bonus-only	n/a	n/a
	2 (14/3/3)	Absolute forecast error	n/a	n/a
	3 (12/7/7)	Absolute forecast error	44.0	< 44.0
	4 (10/10/10)	Absolute forecast error	20.2	< 20.2
2	5 (10/6/4)	Differentiated forecast error	38.4	< 38.4
	6 (10/8/2)	Differentiated forecast error	15.7	< 15.7
	7 (10/10/0)	Differentiated forecast error	0.0	≤ 0.0
	8 (10/12/2)	Differentiated forecast error	0.0	≤ 0.0

Note: Incentive parameters are in ECUs. Fixed compensations are C_S = 1,000 ECUs for Sales and C_O = 2,000 ECUs for Operations in all treatments.

Table 4.1.: Experimental design and model predictions of the main experiment

ing equilibrium exists, assuming that the lying aversion β is sufficiently small. Treatments 2–4 cover the absolute forecast error incentive system, including a setting (Treatment 2) where no separating equilibrium exists under the standard model. Since we are most interested in designing incentive schemes for truthful forecast information sharing, we include four treatments (Treatments 5–8) with different parameterizations of the differentiated forecast error scheme, including two treatments (Treatments 7 and 8) for which the standard model predicts truthful forecast information sharing.

In all treatments, the market condition Φ is normally distributed with a mean of 100 and a standard deviation of 30 and the market uncertainty E is normally distributed with a mean of zero and a standard deviation of 30. To avoid negative demand realizations, we drew realizations of Φ and E before the experiment and used them in both sessions. The ex-ante probability of a negative demand for a distribution of $D = (\Phi + E) \sim \mathcal{N}(100, 42.43^2)$ is only $P(d < 0) = 0.0092$. In line with Ho et al. (2010, p. 1896), we chose not to encounter subjects with a truncated normal distribution, which would have added unnecessary complexity to the experiment.

With respect to Operations, we are interested in analyzing how people process demand forecast information rather than how they make optimal order quantity decisions. Therefore, we chose a unit overage cost factor of $c_o = 10$ that is equal to the unit underage cost factor of

$c_u = 10$, which results in a critical ratio of $\alpha = 0.5$. With this critical ratio and a zero-mean normal distribution of the market uncertainty, the expected-utility-maximizing order quantity is equal to the market condition. With other critical ratios, computing the expected-utility-maximizing order quantity is more complex and deviations between the actual order quantities and expected-utility-maximizing quantities could be due to effects other than demand information processing, such as the pull-to-center effect (e.g., Schweitzer and Cachon, 2000). For a critical ratio of $\alpha = 0.5$, however, the data of Bostian et al. (2008) suggests that human decision makers choose order quantities close to the optimum, i.e., close to the mean of a symmetric distribution. We validated this presumption in our experiment by means of a post-experiment question on hypothetical order quantities. We report the results in Section 5.3.3.

4.1.2. Experimental Protocol

The experiment was conducted at the Cologne Laboratory for Economic Research (CLER) at the University of Cologne. The laboratory protocol was the same for all sessions and we followed the good practices proposed by Katok (2011). We used the experimental programming environment z-Tree (Fischbacher, 2007). Subjects were recruited from the subject pool of the CLER using the recruitment software ORSEE (Greiner, 2015). Subjects were students of economics, business administration, and information systems with an average age of 25.0 years. The instructions and screenshots of the software are contained in Sections A.3.1 and A.3.3 of the Appendix. Additional subject pool characteristics are reported in Section A.3.6 of the Appendix.

After entering the laboratory, until the experiment was officially concluded, subjects were not allowed to talk to each other and no communication was observed. After being randomly assigned to seats, subjects received a handout with instructions, including example calculations of the payoffs. The instructions were first read aloud by the instructor, then silently by each subject. Subjects had to click a button on the screen to indicate that they were ready to start the experiment. If subjects had questions, they could raise their hand and ask them quietly. The instructor did not answer the question but rather directed subjects to the relevant part of the instructions to ensure that one person was not better informed than another.

Subjects were told that they were working for a company with a Sales and an Operations department, where Sales forecasts customer demand and Operations decides on the production quantity. They were informed of the random nature of the demand and of the activities they were expected to perform in the experiment. The incentive systems and individual compensations were explained in detail and subjects were informed that they would be exposed to four different treatments with eight periods each. They were also informed that they would not play with the same person more than once within a treatment. Before the start of the actual experiment, subjects had to answer a pre-experiment quiz (see Section A.3.2 of the Appendix) to test if they had understood the instructions. To become acquainted with the software, subjects played four test periods before the 32 periods of the actual experiment started. All participants kept the role (Sales or Operations) that was assigned to them during the entire experiment.

The sequence of events within a period was as follows: At the beginning of the period, the computer randomly matched two players. The identity of the players was not revealed either during or after the experiment. Then, the realization of the market condition ϕ was privately shown to Sales. Sales could use decision support in the form of a table that showed how the expected payoff of Sales depends on the demand forecast $\hat{\phi}$ and the order quantity q (which Sales did not know, but could only estimate). Then, Sales entered a demand forecast $\hat{\phi}$ and Operations received the demand forecast. Operations could use decision support in the form of a table that showed how the expected payoff of Operations depends on the market condition ϕ (which Operations did not know, but could only estimate) and the order quantity q. Operations entered the order quantity q, a random error ε was generated, and the payoffs of Sales and Operations were computed. All results (except for the realization of the market condition) were shown to both players at the end of a period. The market condition and error term of a period within a treatment were the same for all subjects.

After the last period was played, subjects could increase their payout by making a set of Holt–Laury type lottery choices (Holt and Laury, 2002; Eckel and Grossman, 2008) (see Section A.3.4 of the Appendix). Then, subjects filled out a questionnaire with demographic data and some additional questions regarding the experiment (see Sections A.3.5 and A.3.6 of the Appendix). Finally, subjects were paid privately based on their performance. A session lasted approximately 120 minutes and subjects earned an average of EUR 21.23, including a fixed participation fee of EUR 2.50.

Compared to other experiments in economics or OM, the interactive decision task of our experiment is fairly complex. In Section 5.3.3, we take a closer look at the different measures that we took to support decision making and to ensure subjects' understanding of the game.

4.1.3. Overview of Results

The results of the experiment are shown in Figure 4.1. The figure shows three graphs for each of the eight treatments of the experiment. The left graph of a treatment shows the demand forecast $\hat{\phi}$ plotted against the market condition ϕ, the middle graph shows the order quantity q plotted against the demand forecast $\hat{\phi}$, and the right graph shows the order quantity q plotted against the market condition ϕ. Each dot corresponds to one observation in the experiment. The solid line in the left graph corresponds to the truth-telling solution. For incentive systems for which the standard model ($\gamma = 1, \beta = 0$) has a separating equilibrium, the dashed lines show the equilibrium forecasting and ordering predictions of the standard model.

Summary statistics are reported in Table 4.2. We use the metric *forecast distortion* (i.e., the difference between the market condition and the demand forecast) to analyze the forecasting behavior of Sales. We use the metric *forecast correction* (i.e., the difference between the order quantity and the demand forecast) to analyze the ordering behavior of Operations. In addition, we use the metric *order deviation* (i.e., the difference between the order quantity and the market condition) to characterize the quality of the order decision of Operations. The aggregate results of Table 4.2 already indicate that Sales subjects forecast considerably less than predicted by the standard model, which is in line with the predictions of the behavioral model (see Table 4.1). Operations subjects seem to anticipate this behavior and correct forecast distortions.

Figures 4.2 and 4.3 visualize the effect of an incentive system on the payoffs to Sales and Operations. We use the expected profit of Sales as well as the expected overage and underage costs of Operations to quantify these effects. Note that the profit of Sales cannot be directly compared across treatments because the unit sales bonus and forecast error penalties differ.

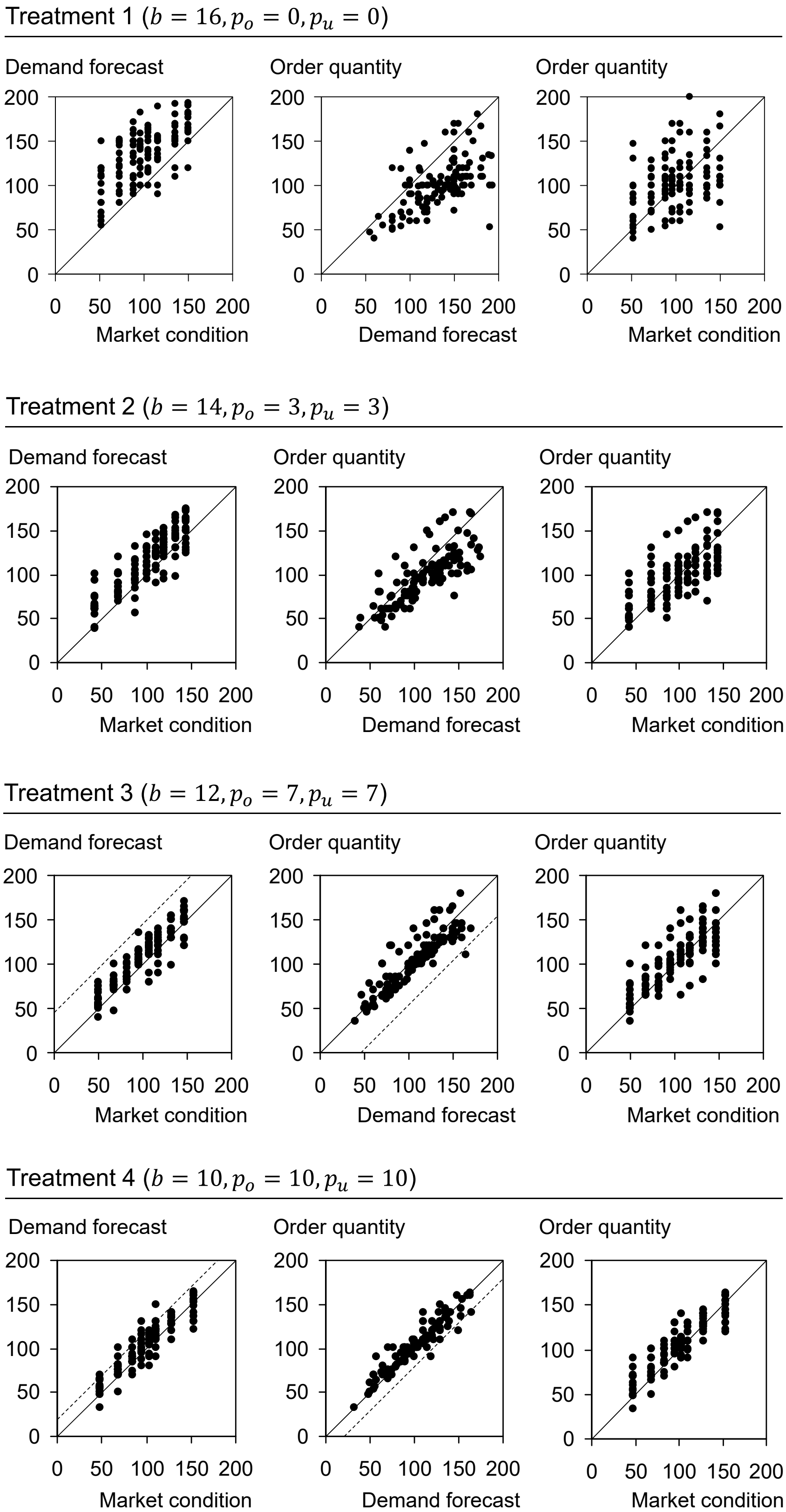


Figure 4.1.: Forecast and order decisions of the main experiment

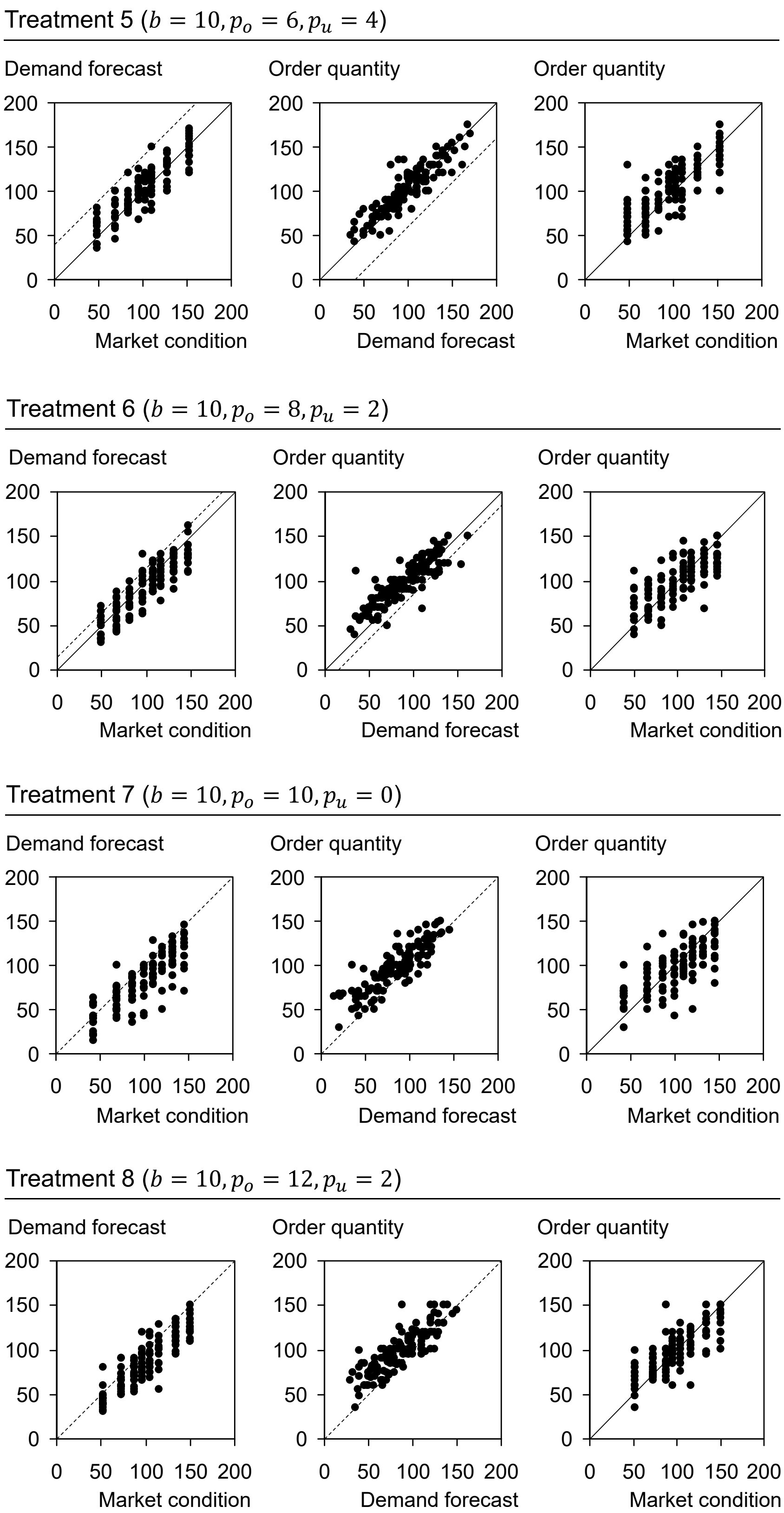


Figure 4.1.: Forecast and order decisions of the main experiment (continued)

Treatment ($b/p_o/p_u$)	Forecast distortion $\hat{\phi}-\phi$ Std. model	Actual average	Corr. $(\hat{\phi},\phi)$	Forecast correction $q-\hat{\phi}$ Std. model	Actual average	Corr. $(q,\hat{\phi})$	Order deviation $q-\phi$ Std. model	Actual average	Corr. (q,ϕ)
1 (16/0/0)	n/a	40.7 (41.6)	0.56	n/a	-38.9 (46.0)	0.44	n/a	1.9 (36.5)	0.30
2 (14/3/3)	n/a	13.3 (16.0)	0.87	n/a	-15.4 (20.3)	0.77	n/a	-2.0 (24.3)	0.68
3 (12/7/7)	44.0	6.9 (12.3)	0.92	-44.0	-3.9 (14.4)	0.89	0.0	3.0 (18.4)	0.82
4 (10/10/10)	20.2	3.9 (11.8)	0.93	-20.2	1.0 (9.7)	0.95	0.0	4.9 (14.1)	0.89
5 (10/6/4)	38.4	1.2 (14.4)	0.89	-38.4	3.2 (13.2)	0.90	0.0	4.4 (19.3)	0.79
6 (10/8/2)	15.7	-8.4 (14.2)	0.89	-15.7	7.1 (14.9)	0.87	0.0	-1.3 (20.2)	0.76
7 (10/10/0)	0.0	-16.7 (17.1)	0.85	0.0	12.3 (15.0)	0.86	0.0	-4.5 (22.1)	0.72
8 (10/12/2)	0.0	-14.8 (14.0)	0.89	0.0	11.6 (15.2)	0.85	0.0	-3.2 (19.3)	0.77

Note: Standard deviations are reported in parentheses. Correlation is the Pearson correlation coefficient.

Table 4.2.: Summary statistics of the main experiment

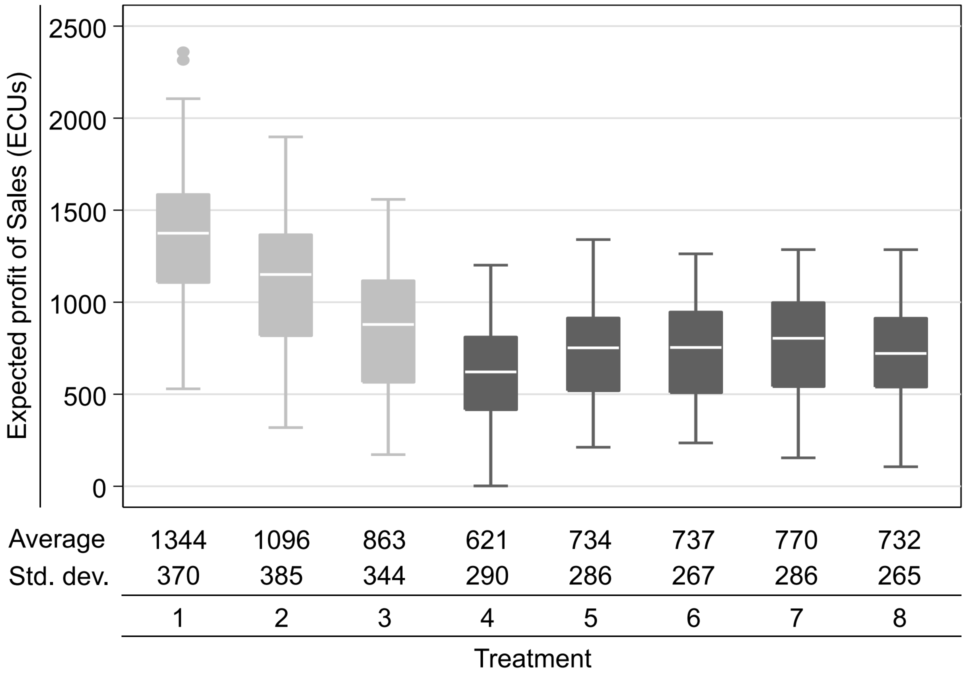
Especially in Treatments 1–3, the unit sales bonus is greater than in Treatments 4–8 and hence profits are comparably high. The unit overage and underage cost of Operations is the same in all treatments and the overall cost of Operations is comparable across treatments. Because the critical ratio of Operations is aligned with the critical ratio of the company, we can use the expected cost of Operations as a proxy for the profit of the company. Figure 4.3 suggests that the cost of Operations is particularly high in the sales-bonus-only Treatment 1 and somewhat high in Treatment 2, which is in line with the predictions of Table 4.1.

Because we used a within-subject design for each session, we analyzed the data for time and treatment order effects. The detailed regression results are discussed in Section 5.3.1.

We next use the data of the experiment to estimate the behavioral parameters γ and β before proceeding to test our hypotheses.

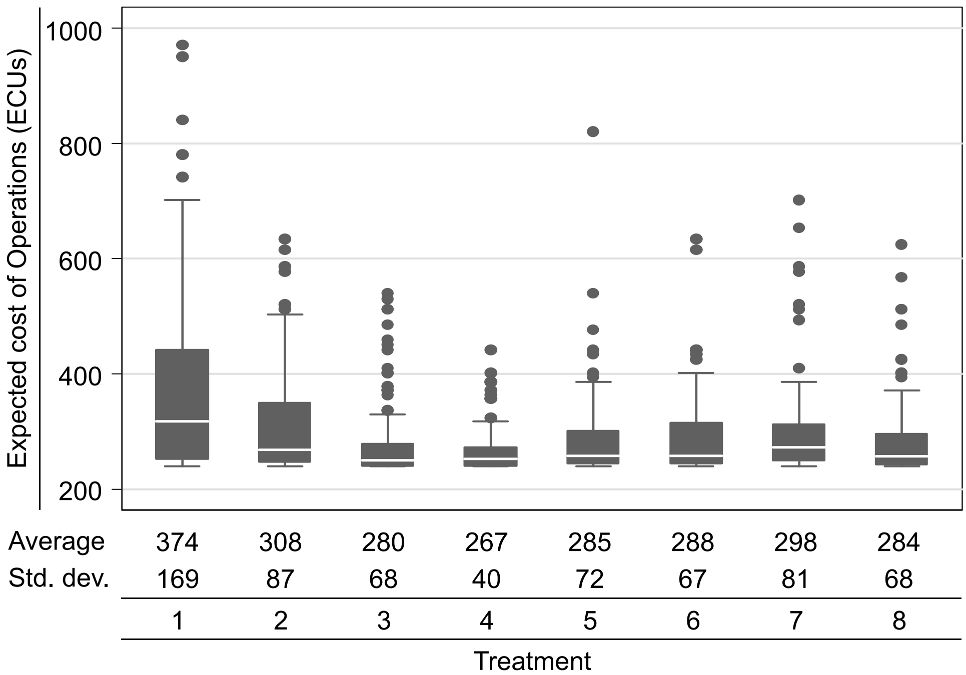
4.1.4. Estimation of Behavioral Parameters

Since we do not know whether a separating equilibrium exists for Treatments 1 and 2 (see Table 4.1), we use the data of Treatments 3–8 to estimate the forecast error penalty factor γ and the lying aversion factor β by maximum likelihood (ML) estimation. Based on Theorem



Note: Expectations are taken with respect to E. Light grey boxes mark treatments with $b \neq 10$.

Figure 4.2.: Expected profit of Sales in the main experiment



Note: Excluding fixed compensation C_O. Expectations are taken with respect to E.

Figure 4.3.: Expected cost of Operations in the main experiment

1, we formulate a nonlinear mixed effects model

$$\delta_{it} = u_{it} + \begin{cases} G^{-1}\left(\frac{b_{it}(1-\alpha)+\gamma_i p_{u,it}-\beta_i}{\gamma_i(p_{u,it}+p_{o,it})}\right) & \text{for } \frac{b_{it}(1-\alpha)-\beta_i}{\gamma_i} < p_{o,it} < p_{u,it} + 2\frac{b_{it}(1-\alpha)-\beta_i}{\gamma_i}, \\ 0 & \text{for } p_{u,it} + 2\frac{b_{it}(1-\alpha)-\beta_i}{\gamma_i} \leq p_{o,it} \leq p_{u,it} + 2\frac{b_{it}(1-\alpha)+\beta_i}{\gamma_i}, \\ G^{-1}\left(\frac{b_{it}(1-\alpha)+\gamma_i p_{u,it}+\beta_i}{\gamma_i(p_{u,it}+p_{o,it})}\right) & \text{for } p_{o,it} > p_{u,it} + 2\frac{b_{it}(1-\alpha)+\beta_i}{\gamma_i}, \end{cases}$$

where subjects are indexed by i and periods are indexed by t. The within-subject errors u_{it} are independently distributed as $\mathcal{N}(0, \sigma_u^2)$. We model the parameters γ and β as random coefficients to allow for time-constant unobserved heterogeneity among subjects. We separate the fixed effects from the random effects of the parameters by defining $\gamma_i = \mu_\gamma + r_{\gamma,i}$ and $\beta_i = \mu_\beta + r_{\beta,i}$, where $\boldsymbol{\mu} = \binom{\mu_\gamma}{\mu_\beta}$ are the fixed effects of the parameters and $\boldsymbol{r_i} = \binom{r_{\gamma,i}}{r_{\beta,i}}$ are the subject-specific random effects, which follow a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$ with variance-covariance matrix $\boldsymbol{\Psi} = \begin{pmatrix} \sigma_\gamma^2 & \rho\sigma_\gamma\sigma_\beta \\ \rho\sigma_\gamma\sigma_\beta & \sigma_\beta^2 \end{pmatrix}$. Since the observations between subjects are independent, but those within subjects are jointly conditional upon the random effects $\boldsymbol{r_i}$, we use the likelihood function

$$\begin{aligned} \mathcal{L}\left(\boldsymbol{\mu}, \boldsymbol{\Psi}, \sigma_u^2 \mid \boldsymbol{\delta}\right) =& \prod_i \int p\left(\boldsymbol{\delta_i} \mid \boldsymbol{\mu}, \boldsymbol{r_i}, \sigma_u^2\right) p\left(\boldsymbol{r_i} \mid \boldsymbol{\Psi}\right) d\boldsymbol{r_i} \\ =& \prod_i \int \frac{1}{\left(2\pi\sigma_u^2\right)^{T_i/2}} \cdot \exp\left(-\frac{1}{2\sigma_u^2}\left\|\boldsymbol{\delta_i} - \boldsymbol{\delta_i^{sep}}\right\|^2\right) \\ & \cdot \frac{1}{2\pi\sigma_\gamma\sigma_\beta\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{r_{\gamma,i}}{\sigma_\gamma^2} + \frac{r_{\beta,i}}{\sigma_\beta^2} - \frac{2r_{\gamma,i}r_{\beta,i}\rho}{\sigma_\gamma\sigma_\beta}\right)\right) d\boldsymbol{r_i}, \end{aligned} \tag{4.1}$$

where $p(\boldsymbol{\delta_i} \mid \cdot)$ and $p(\boldsymbol{r_i} \mid \cdot)$ are the densities of the corresponding multivariate normal distributions, T_i is the number of observations for subject i, and $\boldsymbol{\delta_i^{sep}}$ is a vector with the equilibrium distortions according to Equation (3.6) given the respective values of the subject-specific parameters γ_i, β_i and covariate vectors $\boldsymbol{b_i}, \boldsymbol{p_{o,i}}$, and $\boldsymbol{p_{u,i}}$.

To estimate the parameters, we use the Lindstrom and Bates (1990) algorithm implemented in the nlme package for R (Pinheiro et al., 2017). By reparameterizing $\tilde{\gamma}_i = \log(\gamma_i), \tilde{\beta}_i = \log(\beta_i)$ and setting $\gamma_i = \exp(\tilde{\gamma}_i), \beta_i = \exp(\tilde{\beta}_i)$ in Equation (4.1), we restrict the optimization to positive values for γ and β while keeping the optimization problem unconstrained (Pinheiro and Bates, 2000, p. 351). Following up on the assumptions above, $\tilde{\gamma}$ and $\tilde{\beta}$ follow a multivariate normal distribution and γ and β follow a multivariate log-normal distribution. We use $\mu_{\tilde{\gamma}}, \mu_{\tilde{\beta}}, \sigma_{\tilde{\gamma}}, \sigma_{\tilde{\beta}}, \tilde{\rho}$ to denote the distributional characteristics of the transformed variables.

			Model 1	Model 2	Model 3
Model description			▪ γ random ▪ β random	▪ γ fixed ▪ β random	▪ γ random ▪ β fixed
Estimation results	$\mu_{\tilde{\gamma}}$		1.074 (0.127)	1.042 (0.043)	1.084 (0.125)
	$\sigma_{\tilde{\gamma}}$		0.623		0.614
	$\mu_{\tilde{\beta}}$		0.500 (0.141)	0.678 (0.129)	0.591 (0.133)
	$\sigma_{\tilde{\beta}}$		0.141	0.000	
	σ_u		11.874	14.225	11.876
Distribution of γ and β	γ	Mode	1.986	2.834	2.028
		Mean	3.556	2.834	3.571
		Std. dev.	2.449		2.416
	β	Mode	1.616	1.971	1.805
		Mean	1.666	1.971	1.805
		Std. dev.	0.237	0.001	
Goodness of fit	$\log(\mathcal{L})$		-3,023	-3,129	-3,023
	BIC		6,086	6,284	6,073

Note: Standard errors are reported in parentheses. All fixed effects are significant at the 1% level (t-test).

Table 4.3.: Estimation results of behavioral models with γ and β

Tables 4.3 and 4.4 summarize the estimation results. We use the logarithm of the likelihood ($\log(\mathcal{L})$) and the Bayesian information criterion (BIC), defined as $\text{BIC} = -2\log(\mathcal{L}) + k\log(n)$, where n is the number of observations and k is the number of parameters, to assess the fit of alternative models.

The random effects estimated in a model where both $\tilde{\gamma}$ and $\tilde{\beta}$ are treated as random coefficients (Model 1) are highly correlated ($\tilde{\rho} = -0.998$). This result can be explained by the fact that in the majority of treatments (Treatments 3–6) both behavioral parameters γ and β affect forecast decisions in the same direction, i.e., they reduce the forecast inflation compared to the standard model. In particular for subjects in Session 1, where we only use Treatments 3 and 4 for the estimation procedure, the standard model predicts overforecasting in both treatments and either values of $\gamma > 1$ or values of $\beta > 0$ could be the reason for the more truthful forecast-

			Model 4	Model 5	Model 6
Model description			▪ γ random ▪ $\beta = 0$	▪ $\gamma = 1$ ▪ β random	▪ $\gamma = 2$ ▪ β random
Estimation results	$\mu_{\tilde{\gamma}}$		1.157 (0.120)		
	$\sigma_{\tilde{\gamma}}$		0.596		
	$\mu_{\tilde{\beta}}$			1.509 (0.063)	0.697 (0.183)
	$\sigma_{\tilde{\beta}}$			0.151	0.808
	σ_u		12.558	17.000	14.095
Distribution of γ and β	γ	Mode	2.230		
		Mean	3.799		
		Std. dev.	2.480		
	β	Mode		4.419	1.045
		Mean		4.572	2.783
		Std. dev.		0.693	2.671
Goodness of fit	$\log(\mathcal{L})$		-3,067	-3,269	-3,140
	BIC		6,154	6,558	6,301

Note: Standard errors are reported in parentheses. All fixed effects are significant at the 1% level (t-test).

Table 4.4.: Estimation results of reduced behavioral models

ing behavior that we observe in the experiment. We therefore follow the guidelines of Pinheiro and Bates (2000) for the selection of random effects and test two alternative models with $\tilde{\beta}$ random only (Model 2) and $\tilde{\gamma}$ random only (Model 3) (Pinheiro and Bates, 2000, pp. 282–284 and pp. 359–360).

In Model 2, the standard deviation of β is approximately zero and the model provides a significantly worse fit than Model 1 (likelihood ratio test, $p < 0.001$). In Model 3, we do not lose explanatory power (likelihood ratio test, $p = 0.8402$) and obtain a lower BIC than for Model 1. We conclude that Model 3 fits our data better than Models 1 and 2.

In Model 3, the forecast error penalty factor γ is log-normally distributed with a mean of 3.571 and the lying aversion factor β is 1.805. The interpretation is that subjects evaluate forecast error penalties 3.571 times higher on average than the equivalent sales bonus payments and experience a negative utility of 1.805 ECUs for every unit of forecast distortion.

We further test the specification of our model by comparing Model 3 with three reduced models, where only γ (Model 4) or β (Models 5 and 6) is included as an explanatory variable

(Table 4.4). In Model 4, we set $\beta = 0$ and in Model 5, we set $\gamma = 1$, i.e., we use the parameter values of the standard model. In Model 6, we set $\gamma = 2$ to test whether a "typical" loss aversion factor of prospect theory is sufficient to explain the observed behavior (for a discussion on the values of loss aversion factors, see Abdellaoui et al., 2007). The log-likelihood of Model 3 is significantly higher than that of the reduced models (likelihood-ratio test, $p < 0.001$ for all comparisons) and the BIC of Model 3 is lower than the BIC of the reduced models. These results show that neither the overproportional weighting of forecast error penalties nor lying aversion alone can explain the observed behavior as well as the combination of both effects.

The worse fit of Model 6 compared to Models 2 and 3 (likelihood-ratio test, $p < 0.001$ for both comparisons) further indicates that the estimated value of the parameter γ is significantly greater than two. This supports the presumption that the forecast error penalty factor γ includes various behavioral drivers that influence behavior in the same direction rather than loss aversion alone. We can further see from Figure 4.4 that the in-sample predicted forecasts based on the parameter estimates of Model 3 are well aligned with the actual forecast decisions made in each of the treatments.

The estimation procedure that we follow above builds upon the assumption that the within-subject errors u_{it} are independent and identically normally distributed with mean zero and variance σ_u^2. We use the within-subject residuals of the ML estimation to verify this assumption. The histogram and quantile-normal plot in Figure 4.5 indicate that the distribution of residuals is slightly leptokurtic. A Kolmogorov–Smirnov test confirms a significant deviation from a normal distribution ($p < 0.001$) due to a kurtosis of 4.747. However, the residuals are symmetrically distributed (skewness = 0.004) such that we can expect consistent, albeit slightly inefficient estimates (Pinheiro and Bates, 2000, p. 180).

4.1.5. Tests of Hypotheses

To test our hypotheses, we must distinguish between treatments where a separating equilibrium exists from those where we do not know if a separating equilibrium exists. Based on Corollary 1, we can conclude that an equilibrium along the lines of Theorem 1 exists for Treatments 3–8. With respect to the sales-bonus-only incentive scheme of Treatment 1, the results of the ML

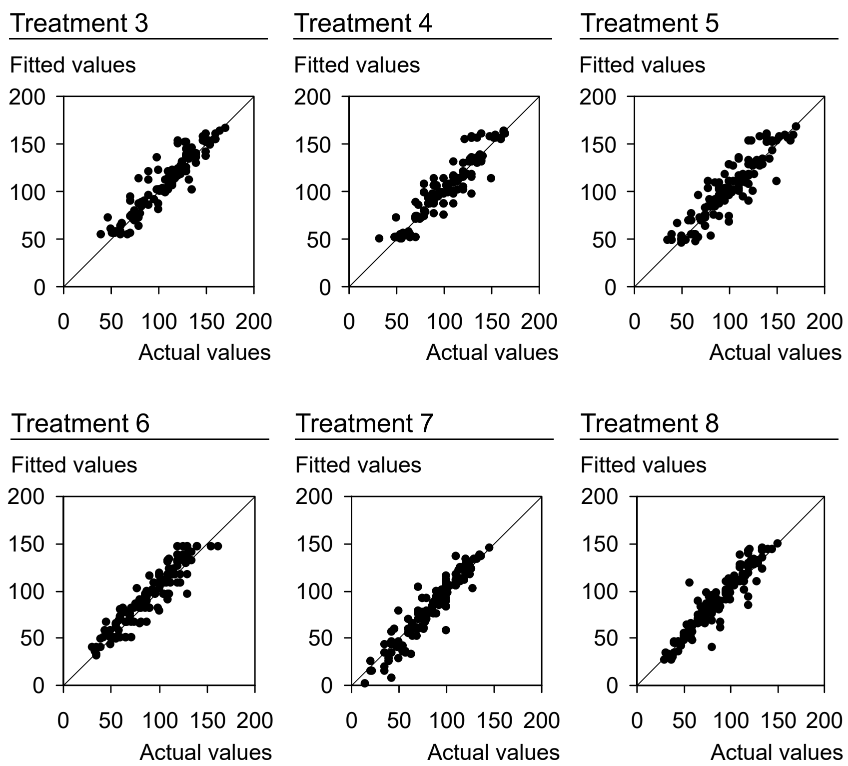


Figure 4.4.: Fitted versus actual forecasts of the main experiment

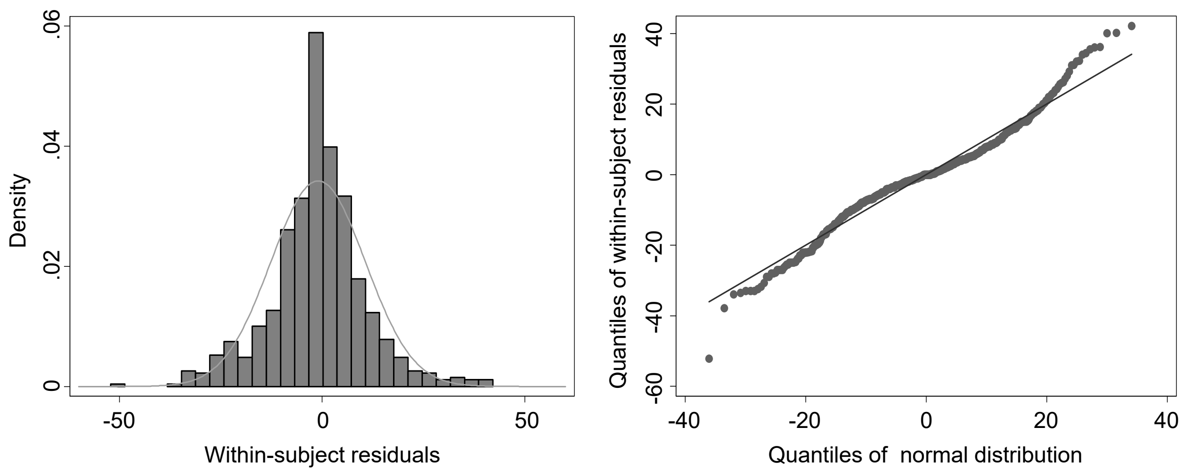


Figure 4.5.: Diagnostic plots of within-subject residuals of the main experiment

estimation (Model 3 in Table 4.3) indicate that there is no such equilibrium. In order for a separating equilibrium to exist, the condition $p_o > (b(1-\alpha) - \beta)/\gamma$ must hold. For Treatment 1, this requires $\beta > b(1-\alpha) = 8$, which is far above our estimate of $\beta = 1.805$. We further corroborate this finding by analyzing subject-level ordinary least squares (OLS) estimates of β. There is only one subject with an estimate of $\beta = 8.69 > 8$. For Treatment 2, it is unclear whether a separating equilibrium exists. In order for a separating equilibrium to exist, we require $\gamma > (b(1-\alpha) - \beta)/p_o = 1.732$. The behavioral parameter γ is a random coefficient and there is a nonnegligible probability that it is smaller than this value ($P(\gamma < 1.732) = 10.6\,\%$). We will therefore only use Treatment 1 to compare the behavior of Sales and Operations in a setting where no separating equilibrium exists to the behavior in Treatments 3–8, where a separating equilibrium does exist. Unless stated otherwise, we use two-sided Wilcoxon signed-rank tests and Mann–Whitney U tests based on subject-level averages by treatment to establish statistical validity.

Hypothesis 1 predicts that a model with $\gamma > 1$ and $\beta > 0$ better explains actual behavior than the standard model. The estimation results support this hypothesis. The mean estimate of γ is significantly different from 1 and the estimate of β is significantly different from 0 (t-test, $p < 0.001$ for both comparisons). Also, the explanatory power of Model 3 is significantly better (likelihood ratio tests, $p < 0.001$) than that of Models 4 and 5, where γ and β are set to 1 and 0 respectively.

Hypothesis 2 states that forecast distortions are smaller and forecast corrections are greater than predicted by the standard model. Figure 4.1 shows that the actual forecast distortions and forecast corrections deviate systematically from the solutions of the standard model (dashed lines) and Table 4.2 reports the corresponding averages. For Treatments 3–6, the standard model predicts positive forecast distortions and negative forecast corrections. In line with Hypothesis 2, the actual average forecast distortions are significantly smaller and the actual average forecast corrections are significantly greater than predicted by the standard model (one-sided Wilcoxon signed-rank test, $p < 0.001$ for all comparisons). For Treatments 7 and 8, the standard model predicts forecast distortions and corrections of zero. The actual average forecast distortions are significantly smaller than zero and the average forecast corrections are significantly greater than zero (one-sided Wilcoxon signed-rank test, $p < 0.001$ for all comparisons). We conclude that actual behavior is different from the behavior predicted by the standard model and that the behavioral model correctly describes the direction of deviation.

Hypothesis 3 predicts positive correlations between market conditions and forecasts, forecasts and order quantities, and order quantities and market conditions. The scatter plots in Figure 4.1 indicate positive correlations between all three pairs. The corresponding correlation coefficients are shown in Table 4.2. All correlations are significantly greater than zero (t-test after Fisher z-transformation, $p < 0.001$ for all treatments), which provides support for Hypothesis 3.

The positive correlations of Treatment 1 provide robustness to the experimental results of Özer et al. (2011, 2014), who also found that positive correlations under a sales-bonus-only incentive system exist. In contrast to Özer et al. (2011, 2014), we differentiate between treatments where a separating equilibrium exists (Treatments 3–8) and treatments where it does not (Treatment 1). *Hypothesis 4* states that the above correlations are higher in Treatments 3–8 than in Treatment 1. We can see from Figure 4.1 and Table 4.2 that the forecast distortions and forecast corrections vary more in Treatment 1 than in Treatments 3–8. In support of the hypothesis, the correlation coefficients $\rho(\hat{\phi}, \phi)$, $\rho(q, \hat{\phi})$ and $\rho(q, \phi)$ in Treatment 1 are significantly lower than in Treatments 3–8 (t-test after Fisher z-transformation, $p < 0.01$ for all comparisons).

Hypothesis 5 suggests that the expected cost of Operations is lower in Treatments 3–8 than in Treatment 1. Because the operational costs of the company are proportional to the overage and underage costs of Operations, this hypothesis is also indicative for the overall company profit. Table 4.2 shows that the average expected cost of Operations is significantly higher in Treatment 1 than in Treatments 3–8 (Wilcoxon signed-rank test for within-subject comparisons to Treatments 3 and 4, $p < 0.001$; Mann–Whitney U test for between-subject comparisons to Treatments 5–8, $p < 0.001$), which provides support for Hypothesis 5.

We conclude that the structural predictions of the behavioral model are supported by the experimental data. In particular, the sales-bonus-only incentive system of Treatment 1 leads to higher forecast inflations, lower correlations and higher expected operational costs than the incentive systems with forecast error penalties in Treatments 3–8.

Table 4.2 indicates that the actual average forecast distortions of Treatments 4 and 5 are close to zero. While the actual average forecast distortion is significantly different from zero in Treatment 4 (Wilcoxon signed-rank test, $p < 0.01$), it is not significantly different from zero in Treatment 5 (Wilcoxon signed-rank test, $p = 0.570$). Using the mean parameter estimates $\gamma = 3.571$ and $\beta = 1.805$, we can compute the in-sample predictions for Treatment 4 ($\delta^{sep} = 3.372$) and Treatment 5 ($\delta^{sep} = 0$). By coincidence, Treatment 5 covers the truthful forecast sharing case. While this finding is generally in line with Corollary 2, it is based on in-sample parameter estimates and hence has little external validity. To provide such a validation, we next analyze how well the model works with out-of-sample subjects and out-of-sample treatments.

4.2. Validation Experiment

4.2.1. Experimental Design and Results

In the validation experiment we used four treatments. Table 4.5 summarizes the design and results of the experiment. Treatments 9 and 10 were designed to incentivize truthful information sharing based on the behavioral model. To obtain the parameter values for over- and under-forecasting penalties, we used the mean values of the forecast error penalty factor ($\gamma = 3.571$) and the lying aversion factor ($\beta = 1.805$) that we estimated above and selected two treatments with integer parameter values $(p_o - p_u) \in [1.80; 3.81]$ that result in a distortion value of $\delta^{sep} = 0$. Treatments 11 and 12 were designed similarly based on the predictions of the standard model requiring $p_o - p_u = 10$.

We used the same experimental setup and protocol as in the main experiment and conducted the validation experiment with 32 new subjects. Figure 4.6 shows the individual decisions of the experiment. Table 4.5 summarizes the predictions of the standard model, the predictions of the behavioral model, the average forecast distortions and the average forecast corrections that we observed in the experiment. We use two-sided Wilcoxon signed-rank tests to compare actual averages with the predictions of both models.

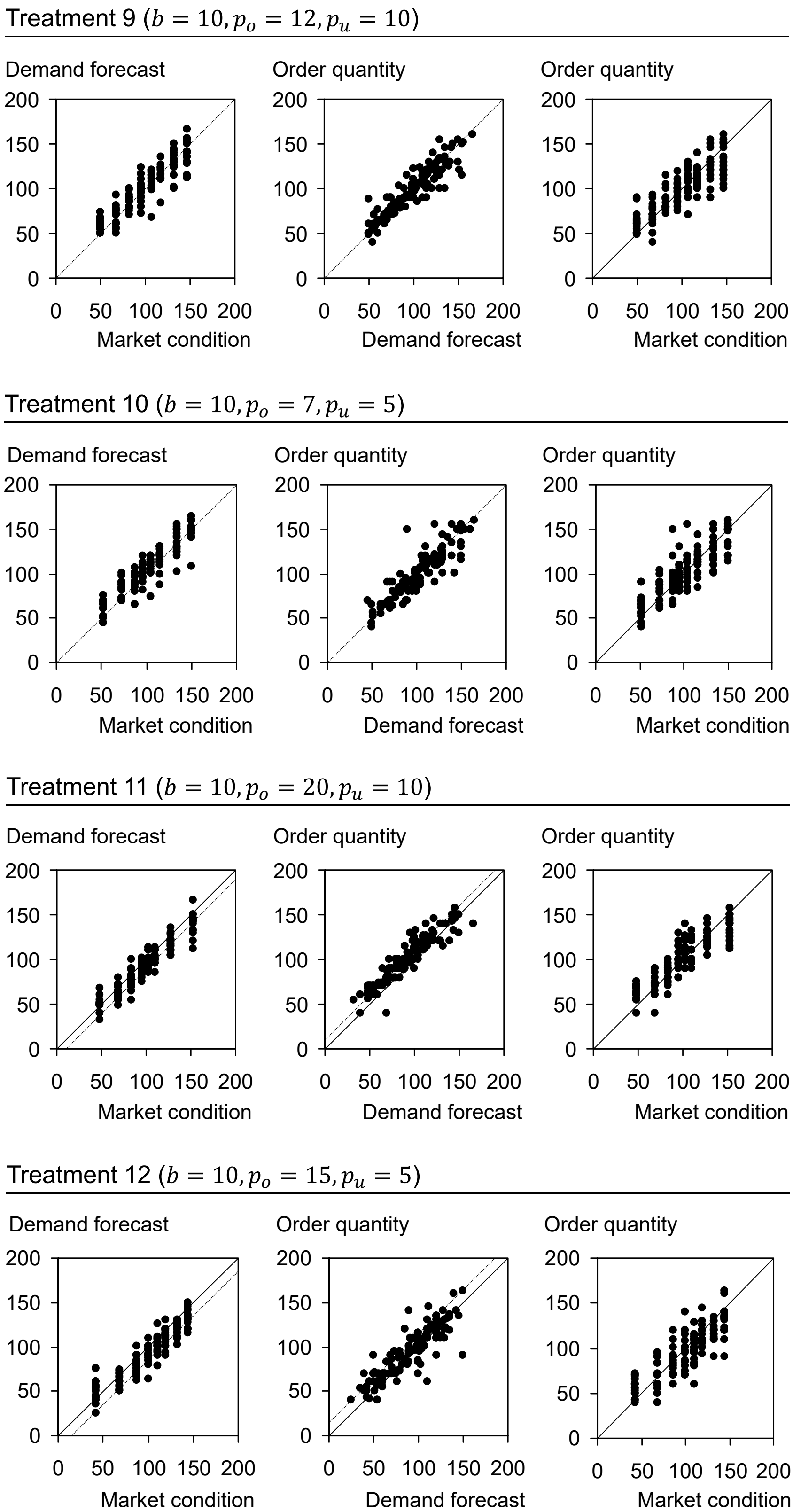


Figure 4.6.: Forecast and order decisions of the validation experiment

Treatment ($b/p_o/p_u$)	Forecast distortion $\hat{\phi} - \phi$ Std. model	Beh. model	Actual average	Corr. $(\hat{\phi}, \phi)$	Forecast correction $q - \hat{\phi}$ Std. model	Beh. model	Actual average	Corr. $(q, \hat{\phi})$	Expected cost of Operations
9 (10/12/10)	14.2	0.0	-0.2 (12.7)	0.91	-14.2	0.0	-1.6 (11.4)	0.92	273 (51.2)
10 (10/7/5)	29.0	0.0	3.6 (11.6)	0.92	-29.0	0.0	-2.9 (12.8)	0.90	273 (55.0)
11 (10/20/10)	0.0	-7.8	-6.9 (9.6)	0.95	0.0	7.8	7.8 (10.4)	0.93	266 (36.6)
12 (10/15/5)	0.0	-11.9	-8.4 (11.4)	0.93	0.0	11.9	2.8 (14.6)	0.87	280 (52.1)

Note: Standard deviations are reported in parentheses. Correlation is the Pearson correlation coefficient. Expected cost of Operations are reported in ECUs, excluding fixed compensation C_O. Expectations are taken with respect to E.

Table 4.5.: Design and results of the validation experiment

Treatments 9 and 10, which were designed with the *behavioral model*, have actual average forecast distortions that are not significantly (Treatment 9, $p = 0.836$) or only weakly significantly (Treatment 10, $p = 0.093$) different from zero. The standard model predicts much greater distortions that are highly significantly different from the actual averages ($p < 0.001$ for both comparisons). The results are similar for the forecast corrections.

Treatments 11 and 12, which were designed with the *standard model*, have actual average forecast distortions that are significantly below the prediction of the standard model ($p < 0.003$ for both comparisons). The behavioral model correctly predicts the underforecasting that we observe and the actual forecast distortions are not significantly (Treatment 11, $p = 0.642$) or only weakly significantly (Treatment 12, $p = 0.079$) different from these predictions. For the forecast corrections, we obtain mixed results. In Treatment 11, the actual average correction differs significantly from the prediction of the standard model ($p < 0.001$), but not from the prediction of the behavioral model ($p = 0.877$). In Treatment 12, the actual average correction does not differ significantly from the prediction of the standard model ($p = 0.146$), but differs significantly from the prediction of the behavioral model ($p < 0.002$).

In addition to the statistical tests above, Figure 4.7 compares the individual decisions of both Sales and Operations to the predicted values of the behavioral model. The visualization supports the finding that actual forecast and order decisions are generally in line with the predictions of the behavioral model. Note that, in contrast to the main experiment (Figure 4.4), the predicted values for forecast distortions and corrections are based on the mean parameter

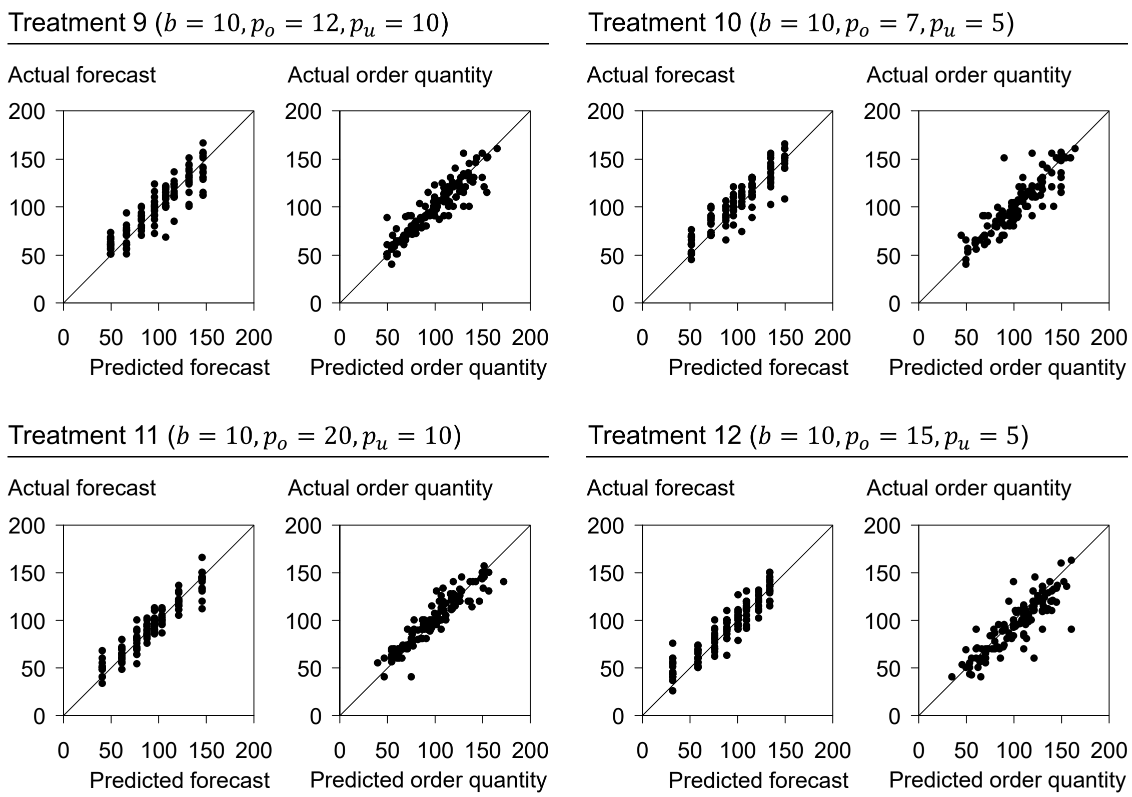


Figure 4.7.: Predicted versus actual decisions of the validation experiment

estimates of the main experiment, and are hence the same for all subjects.

The results of the validation experiment show that the behavioral model generalizes well and provides good estimates of the forecasting and ordering behavior of out-of-sample subjects with out-of-sample treatments. The treatments that we designed with the behavioral model perform well and deliver what they were designed for: They incentivize demand forecasts that are only minimally distorted by Sales and minimally corrected by Operations.

4.2.2. Sensitivity Analysis

The sensitivity of the equilibrium behavior under different truth-telling incentive schemes depends on the choice of forecast error incentives p_o and p_u. For a normally distributed market uncertainty E, it can be shown that the equilibrium forecast distortion δ^{sep} (see Equation 3.6) is less sensitive to changes in the behavioral parameters γ and β for high unit forecast error penalties than for low ones if the difference $p_o - p_u$ is the same. More formally, for any two sets of sales incentive parameters (b, p_o, p_u) and $(b, \bar{p}_o, \bar{p}_u)$ with $p_o - p_u = \bar{p}_o - \bar{p}_u$ and $\bar{p}_o > p_o$, $\bar{p}_u > p_u$, the following proposition holds:

Proposition 1. *If an equilibrium forecast distortion δ^{sep} exists, then it is weakly less sensitive to changes in the behavioral parameters γ and β under the incentive scheme $(b, \bar{p}_o, \bar{p}_u)$ compared to the incentive scheme (b, p_o, p_u).*

We use the same difference of $p_o - p_u = 2$ in both Treatment 9 and 10, but the unit forecast error penalties p_o and p_u are lower in Treatment 10 than in Treatment 9. Hence, estimation errors of the behavioral parameters and subject-specific deviations from the mean values of γ and β can lead to higher forecast distortions in Treatment 10 than in Treatment 9. A similar observation holds for Treatments 11 and 12, where the actual average forecast distortions of Treatment 12 differ more strongly from the prediction of the behavioral model than the forecast distortions of Treatment 11.

To visualize this property for the incentive systems of Treatments 9 and 10, Figure 4.8 shows the equilibrium forecast distortion for different values of γ (horizontal axis) and β (vertical axis). The white zone indicates parameter combinations for which the behavioral model predicts truthful information sharing. The lines indicate parameter combinations of equal forecast distortion with increasing distortion values (shown in white circles) towards $\gamma = 1$ and $\beta = 0$. The points mark our estimates $\gamma = 3.571$ and $\beta = 1.805$ as well as deviations of plus/minus 50 % from the mean estimates. We can observe that larger values for γ and β do not change the equilibrium prediction of zero inflation within the parameter space that we consider. If γ and β become smaller, however, the prediction changes from truthful information sharing to overforecasting. If both parameters deviate by -50 % from the mean, for example, our model predicts forecast inflations of 4.4 units in Treatment 9 and 8.2 units in Treatment 10.

The results of the sensitivity analysis can also explain the comparably low expected cost of Operations (267 ECUs) in the absolute forecast error incentive Treatment 4 of the main experiment. Compared to all other treatments of the main experiment, Treatment 4 has the highest unit forecast error penalties p_o and p_u, which could be the reason why we observe the lowest variances in forecast and order decisions and the lowest expected cost of Operations

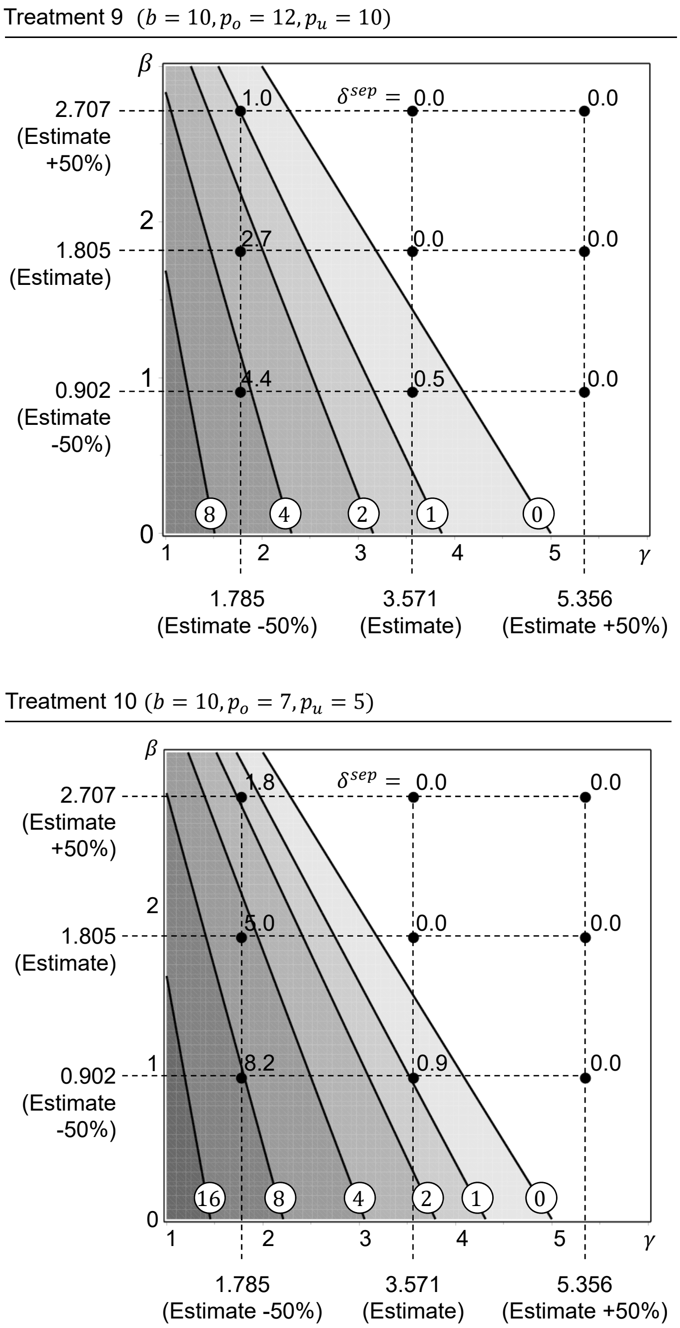


Figure 4.8.: Sensitivity of forecast distortions depending on behavioral parameter values

in this treatment. However, as discussed in Section 4.1.5, the forecasts of Treatment 4 are still significantly biased. We can use Treatment 9 of the validation experiment to verify that a differentiated forecast error incentive system with a similar magnitude of forecast error penalties incentivizes truthful forecasts, while leading to similar expected cost of Operations (273 ECUs) that are not significantly different from the cost in Treatment 4 (Mann–Whitney U test, $p = 0.366$).

5. Discussion

While the behavioral model that we specified in Section 3.1 is grounded in theory and fits the data well (both in-sample as well as out-of-sample), there are other potential explanations for the behavior that we observe in the experiments. In the following, we therefore present alternative models and their fit to our experimental data (Section 5.1), additional experimental data (Section 5.2), and supplementary analyses of the data of the main experiment (Section 5.3) that provide robustness to our previous results.

5.1. Alternative Models

5.1.1. Non-Bayesian Belief Updating

To derive the separating equilibria of our behavioral model, we use the concept of PBE, which assumes that signal receivers update their belief based on Bayes' rule whenever possible (see Section 3.2.1). Previous research has shown that this assumption can be violated in practice (e.g., Charness and Levin, 2005). To see if the Bayesian concept captures belief updating in our setting well, we test two alternative models that relax this assumption.

Trust model.

In the specific context of forecast information sharing, Özer et al. (2011) identify trust on the part of the forecast receiver as the driving factor of belief updating. To test whether such an approach explains the behavior better than the approach we used, we follow Özer et al. (2011) and formulate a model where Operations applies a simple updating rule by forming a weighted average of the observed forecast $\hat{\phi}$ and the prior distribution of the market condition Φ. We formulate the new belief

$$\mu_\xi(\phi \mid \hat{\phi}) = f_\xi(\phi \mid \hat{\phi}),$$

where f_ξ is the density function of $\xi\hat{\phi} + (1-\xi)\Phi$ and $\xi \in (0;1]$ is a trust factor. According to the definition of the trust factor, equilibria in this model cannot be separating, i.e., Operations cannot infer the true market condition, except for the special case of $\xi = 1$ and truthful forecasts $\hat{\phi} = \phi$. We describe a set of equilibrium strategies for Sales and Operations under trust-based belief updating in the following proposition:

Proposition 2. *If Operations uses a trust-based belief updating rule $\mu_\xi(\phi \mid \hat{\phi}) = f_\xi(\phi \mid \hat{\phi})$, and if at least $p_o > 0$ or $\beta > 0$, there exists the following equilibrium:*

(a) *The best response of Operations to a forecast $\hat{\phi}$ is the quantity*

$$q_\xi(\hat{\phi}) = \xi\hat{\phi} + F_Z^{-1}(\alpha),$$

where F_Z is the distribution function of $Z = (1-\xi)\Phi + E$.

(b) *The best response of Sales is the forecast $\hat{\phi}_\xi$ that solves the first-order condition*

$$b\xi\left(1 - G\left(\xi\hat{\phi}_\xi + F_Z^{-1}(\alpha) - \phi\right)\right)$$
$$= \gamma(p_o + p_u)G\left(\hat{\phi}_\xi - \phi\right) - \gamma p_u \begin{cases} +\beta \text{ for } p_o - p_u < \frac{2}{\gamma}\left[b\xi\left(1 - G\left((\xi-1)\phi + F_Z^{-1}(\alpha)\right)\right) - \beta\right], \\ -\beta \text{ for } p_o - p_u > \frac{2}{\gamma}\left[b\xi\left(1 - G\left((\xi-1)\phi + F_Z^{-1}(\alpha)\right)\right) + \beta\right]. \end{cases}$$

For $\frac{2}{\gamma}\left[b\xi\left(1 - G\left((\xi-1)\phi + F_Z^{-1}(\alpha)\right)\right) - \beta\right] \leq p_o - p_u \leq \frac{2}{\gamma}\left[b\xi\left(1 - G\left((\xi-1)\phi + F_Z^{-1}(\alpha)\right)\right) + \beta\right]$, the best response of Sales is the truthful forecast, i.e., $\hat{\phi}_\xi = \phi$.

Note that q_ξ is independent of the incentive parameters. Operations interprets a forecast $\hat{\phi}$ in the same way across all experimental treatments. For our experimental data with $\Phi \sim \mathcal{N}(100, 30^2)$, $E \sim \mathcal{N}(0, 30^2)$ and $\alpha = 0.5$, we have Z normally distributed with mean $(1-\xi)\cdot 100$ and hence $F_Z^{-1}(\alpha) = (1-\xi)\cdot 100$. Hence, for all $\xi < 1$ Operations corrects big forecasts ($\hat{\phi} > 100$) downwards and small forecasts ($\hat{\phi} < 100$) upwards. Both predictions of the model are at odds with what we observe in the experiments. For example, forecast corrections in Figure 4.1 and

Table 4.2 differ depending on the treatment and the direction of forecast correction does not depend on the magnitude of the forecast.

Since we do not have a closed-form solution for the forecasting behavior of Sales in the trust model, we use subject-level OLS in order to estimate the behavioral parameters of the trust model. We compare the model fit based on the BIC, which we calculate as BIC $= \log(n)k + n\log(\text{RSS}/n)$, where n is the number of observations, k is the number of parameters and RSS is the sum of squared residuals. Table 5.1 compares the fit of alternative belief models with that of our original behavioral model (Bayesian model) using the simplified estimation procedure. Note that the estimates for γ and β of the Bayesian model in Table 5.1 are slightly different from the estimates in Table 4.3 due to different assumptions regarding the distribution of γ and β and the different estimation procedures. The higher BIC of the trust model indicates that the Bayesian model provides a better fit to our data than the trust model. Note that the median of the estimates for ξ is 1, implying that the majority of Sales subjects (25 out of 32 subjects) do not expect Operations to follow a strategy where the direction of the forecast correction depends on the size of the forecast.

		Bayesian Model	Trust Model	Naïveté Model
Estimates	γ	4.059 [2.227] (4.599)	3.728 [1.857] (4.754)	3.954 [1.857] (4.664)
	β	2.610 [1.566] (2.600)	1.429 [0.476] (1.778)	1.949 [0.952] (2.238)
	ξ		0.871 [1.000] (0.335)	
	θ			0.543 [0.500] (0.434)
BIC		4,159	4,348	4,343

Note: Reported numbers are the mean, [median] and (standard deviation) of the subject-level estimates.

Table 5.1.: Estimation results of models with alternative belief updating

Naïveté model.

An alternative approach to behavioral belief updating in signaling games is grounded in the idea of sophisticated and naïve (or credulous) responses on the part of the signal receiver (Crawford, 2003; Kartik et al., 2007). Suppose Sales played some invertible strategy $s(\phi)$. A sophisticated response would be if Operations correctly inferred the forecast distortion and updated the belief to $\mu(\phi \mid \hat{\phi}) = 1$ for $\phi = s^{-1}(\hat{\phi})$ (and 0 otherwise). A naïve response would be if Operations fully trusted the forecast and updated the belief to $\mu(\phi \mid \hat{\phi}) = 1$ for $\phi = \hat{\phi}$ (and 0 otherwise). Our experimental data suggests that subjects seem to be quite sophisticated since the average value of forecast corrections approximately offsets the average value of forecast distortions (see Table 4.2). However, it is possible that some subjects do not fully adjust to the sophisticated response but are more credulous than others. We use a factor $\theta \in (0; 1]$ to indicate the "degree of naïveté" and we model the belief of Operations as

$$\mu_\theta(\phi \mid \hat{\phi}) = \begin{cases} 1 \text{ for } \phi = \theta\hat{\phi} + (1-\theta)s^{-1}(\hat{\phi}), \\ 0 \text{ otherwise.} \end{cases}$$

Note that the naïveté-based belief is similar to the trust-based belief because both place some weight on the actual forecast. The advantage of the naïveté model is that it takes into account different incentive parameterizations because the remaining weight $(1-\theta)$ is not attributed to the prior distribution Φ but to the inverted signal $s^{-1}(\hat{\phi})$.

Based on the knowledge that, for a given set of incentive parameters, subjects seem to distort forecasts by a fixed value that is independent of the market condition, we limit attention to linear signaling strategies of the form $s_\theta(\phi) = \phi + \delta$ with $s'_\theta(\phi) = 1$. Then, the following proposition describes a set of equilibrium strategies for Sales and Operations:

Proposition 3. *If Operations uses a naïveté-based belief updating rule μ_θ, and if at least $p_o > 0$ or $\beta > 0$, there exists the following equilibrium:*

(a) *The best response of Operations to a forecast $\hat{\phi}$ is the quantity*

$$q_\theta(\hat{\phi}) = \theta\hat{\phi} + (1-\theta)s^{-1}(\hat{\phi}) + G^{-1}(\alpha).$$

(b) *The best response of Sales is the strategy $s_\theta(\phi) = \phi + \delta_\theta$, where δ_θ solves the first-order* *condition*

$$b\left[1 - G\left(\theta\delta_\theta + G^{-1}(\alpha)\right)\right] = \gamma\left[(p_o + p_u)G(\delta_\theta) - p_u\right] \begin{cases} +\beta \text{ for } p_o - p_u < \frac{2}{\gamma}\left(b(1-\alpha) - \beta\right), \\ -\beta \text{ for } p_o - p_u > \frac{2}{\gamma}\left(b(1-\alpha) + \beta\right). \end{cases}$$

For $\frac{2}{\gamma}(b(1-\alpha) - \beta) \leq p_o - p_u \leq \frac{2}{\gamma}(b(1-\alpha) + \beta)$, *the best response of Sales is the truthful forecast, i.e.,* $\delta_\theta = 0$.

Table 5.1 reports the results of subject-level OLS estimations. The estimates for θ suggest that Sales expects Operations to be partly naïve and partly sophisticated. However, the BIC of the naïveté model is bigger than the BIC of the base model, indicating that the base model provides a better fit to our data due to a more economical use of parameters.

5.1.2. Expectations-Based Reference Points

In the formulation of the utility function of Sales (Section 3.1.2), we assume a status quo-based reference point in the evaluation of forecast errors. As a result, all forecast error penalties are perceived as losses and the effect of loss aversion is captured by the forecast error penalty factor γ. Because we do not know if a status quo-based reference point of zero adequately captures reference-dependent valuations in our setting, we next analyze alternative reference points. Since there is no externally given reference point in our experiments, such as a goal or target bonus level (e.g., Heath et al., 1999; Ockenfels et al., 2015), we explore the possibility of expectations-based reference points in more depth.

Kahneman and Tversky (1979) had already acknowledged that "there are situations in which gains and losses are coded relative to an expectation or aspiration level that differs from the status quo" (Kahneman and Tversky, 1979, p. 286). According to Kőszegi and Rabin (2006), expectations-based reference points are "determined endogenously" by the decision maker (Kőszegi and Rabin, 2006, p. 1133) and as Ho et al. (2010) argue, these reference points are likely to be formed based on salient pieces of information that the decision maker receives or expects to receive.

With regard to the forecast decision in our setting, the two most salient pieces of information are the market condition ϕ and the demand realization $d = \phi + e$. Hence, Sales could form reference forecast error penalties by imagining a situation where the forecast matches the market condition (the truthful forecast) or where the forecast matches actual demand and as such causes no forecast error penalty. To test these two reference points formally, we follow the approach of Kőszegi and Rabin (2006) and split the forecast error related part of the utility function into "consumption utilities"

$$U_S^P(\hat{\phi} \mid \phi) = -\mathbb{E}_E\left(p_o[\hat{\phi} - (\phi + E)]^+\right),$$
$$U_S^N(\hat{\phi} \mid \phi) = -\mathbb{E}_E\left(p_u[(\phi + E) - \hat{\phi}]^+\right),$$

and "gain-loss utilities"

$$\bar{v}\left(p_o\mathbb{E}_E\left([\hat{\phi}^R - (\phi + E)]^+ - [\hat{\phi} - (\phi + E)]^+\right)\right),$$
$$\bar{v}\left(p_u\mathbb{E}_E\left([(\phi + E) - \hat{\phi}^R]^+ - [(\phi + E) - \hat{\phi}]^+\right)\right),$$

where $\hat{\phi}^R$ is the reference forecast and $\bar{v}$ is a value function. Compared to our original model, we use a slightly modified version of the value function $\bar{v}(x) = \tau[x]^+ - \lambda\tau[-x]^+$, where τ is the weight on the gain-loss utility and $\lambda \geq 1$ is a loss aversion factor.

If the reference forecast is the *realized demand* ($\hat{\phi}^R = d$), Sales evaluates deviations from the expected forecast error that would result if Sales forecasted realized demand correctly. Then we have $[\hat{\phi}^R - (\phi + E)]^+ = [(\phi + E) - \hat{\phi}^R]^+ = 0$ and the value function is applied to the negative consumption utilities $U_S^P(\hat{\phi} \mid \phi)$, $U_S^N(\hat{\phi} \mid \phi)$, which results in a total reference-dependent forecast error utility of $(1 + \lambda\tau)\left(U_S^P(\hat{\phi} \mid \phi) + U_S^N(\hat{\phi} \mid \phi)\right)$. This model is equivalent to our original model where the reference forecast error penalty is also equal to zero, with the only difference that we have $\gamma = 1 + \lambda\tau$. Hence, we do not analyze this model further.

If the reference forecast is the *market condition* ($\hat{\phi}^R = \phi$), Sales evaluates deviations from the expected forecast error that would result from a truthful forecast. Note that the resulting reference point, i.e., the reference forecast error penalty from setting $\hat{\phi}^R = \phi$, is still stochastic due to the remaining uncertainty in E. Adding the respective gain-loss utilities to the utility function of Sales gives

$$\begin{aligned} U_S^R(q,\hat{\phi} \mid \phi) =& C_S + b\mathbb{E}_E\left(\min(\phi+E),q\right) - \mathbb{E}_E\left(p_o[\hat{\phi}-(\phi+E)]^+ + p_u[(\phi+E)-\hat{\phi}]^+\right) \\ &+ \tau\left(p_o\mathbb{E}_E\left[[-E]^+ - [\hat{\phi}-(\phi+E)]^+\right]^+ + p_u\mathbb{E}_E\left[[E]^+ - [(\phi+E)-\hat{\phi}]^+\right]^+\right) \\ &- \lambda\tau\left(p_o\mathbb{E}_E\left[[\hat{\phi}-(\phi+E)]^+ - [-E]^+\right]^+ + p_u\mathbb{E}_E\left[[(\phi+E)-\hat{\phi}]^+ - [E]^+\right]^+\right) \\ &- \beta\left|\hat{\phi}-\phi\right|. \end{aligned} \tag{5.1}$$

The following proposition describes an equilibrium, which is structurally equivalent to the equilibrium of Theorem 1, but based on the utility function of Equation (5.1):

Proposition 4. *If Sales evaluates forecast error penalties relative to an expectations-based reference point of* $\hat{\phi}^R = \phi$*, and if* $p_o > (b(1-\alpha)-\beta)/(1+\lambda\tau)$*, there exists a separating equilibrium* $(s_R^{sep}, q_R^{sep}, \mu_R^{sep})$*:*

(i) *For all* ϕ*:* $s_R^{sep}(\phi) = \phi + \delta_R^{sep}$,

(ii) *for all* $\hat{\phi}$*:* $q_R^{sep}(\hat{\phi}) = \hat{\phi} - \delta_R^{sep} + G^{-1}(\alpha)$,

(iii) *for all* $\phi, \hat{\phi}$*:* $\mu_R^{sep}(\phi \mid \hat{\phi}) = \begin{cases} 1 & \text{for } \phi = \hat{\phi} - \delta_R^{sep}, \\ 0 & \text{otherwise,} \end{cases}$

with distortion value

$$\delta_R^{sep} = \begin{cases} G^{-1}\left(\frac{b(1-\alpha)+(1+\tau)p_u-\beta}{(1+\tau)p_u+(1+\lambda\tau)p_o}\right) & \text{for } p_o < \frac{(1+\tau)}{(1+\lambda\tau)}p_u + 2\frac{b(1-\alpha)-\beta}{(1+\lambda\tau)}, \\ 0 & \text{for } \frac{(1+\tau)}{(1+\lambda\tau)}p_u + 2\frac{b(1-\alpha)-\beta}{(1+\lambda\tau)} \le p_o \le \frac{(1+\lambda\tau)}{(1+\tau)}p_u + 2\frac{b(1-\alpha)+\beta}{(1+\tau)}, \\ G^{-1}\left(\frac{b(1-\alpha)+(1+\lambda\tau)p_u+\beta}{(1+\lambda\tau)p_u+(1+\tau)p_o}\right) & \text{for } p_o > \frac{(1+\lambda\tau)}{(1+\tau)}p_u + 2\frac{b(1-\alpha)+\beta}{(1+\tau)}. \end{cases} \tag{5.2}$$

The equilibrium of this model is structurally equivalent to the equilibrium of the original behavioral model (see Theorem 1). The main difference is that γ is now replaced by either

$(1+\tau)$ or $(1+\lambda\tau)$. For example, in the domain of inflated forecasts (first case of δ_R^{sep}), the expected overforecasting penalty is bigger than the expected overforecasting penalty from reporting the true market condition. Hence p_o is weighed by $(1+\lambda\tau)$. On the contrary, the expected underforecasting penalty is smaller than the expected underforecasting penalty from reporting the true market condition. Hence p_u is weighed by $(1+\tau)$. Note that we cannot keep the original parameter γ as a factor that scales the forecast error penalties and captures the possible effects of ex-post forecast error minimization or ambiguity aversion because the model would be overparameterized.

We use the equilibrium prediction δ_R^{sep} in order to estimate the parameters τ, λ, and β by ML estimation with mixed effects. Table 5.2 summarizes the estimation results of the original behavioral model (Model 3 of Table 4.3) and two reference point models. Similar to the estimation in Section 4.1.4, we assume that τ and λ are log-normally distributed random coefficients and that β is a log-transformed fixed coefficient (Reference Point Model 1). We find that the explanatory power of this model is slightly better than that of the original model but at the cost of an additional parameter, which results in a slightly higher information criterion (a BIC of 6,084 versus a BIC of 6,073). The random coefficients in Reference Point Model 1 are highly correlated ($\rho = -0.987$), indicating that the random effects of both parameters are redundant. Besides, the standard deviation of λ is estimated to be zero, which suggests that the random effect of the parameter does not yield additional value to the model.

We therefore reduce Reference Point Model 1 to a model where only τ is assumed to be random (Reference Point Model 2). The BIC improves compared to Reference Point Model 1 and the removal of one parameter does not lead to a worse model fit (likelihood ratio test, $p > 0.05$). However, the reduced model also falls slightly behind the goodness of fit of the original model (a BIC of 6,076 versus a BIC of 6,073), indicating that the expectations-based reference point that we consider does not provide a better explanation of the data than the status quo-based reference point. Additionally, the interpretation of the estimates of Reference Point Model 2 is similar to the original model: The estimate for λ is 0.976 and not significantly different from 1 (t-test, $p = 0.75$). This implies that there is not a clear change in preferences between the domains of gains and losses at the assumed reference point. Given $\lambda \approx 1$, the

			Original Model	Reference Point Model 1	Reference Point Model 2
Estimation results	$\mu_{\tilde{\gamma}}$		1.084 (0.125)		
	$\sigma_{\tilde{\gamma}}$		0.614		
	$\mu_{\tilde{\beta}}$		0.591 (0.133)	0.661 (0.157)	0.563 (0.164)
	$\mu_{\tilde{\tau}}$			0.698 (0.239)	0.640 (0.222)
	$\sigma_{\tilde{\tau}}$			1.048	0.975
	$\mu_{\tilde{\lambda}}$			-0.079 (0.074)	-0.025 (0.077)
	$\sigma_{\tilde{\lambda}}$			0.000	
	σ_u		11.876	11.849	11.887
Distribution of $\gamma, \beta, \tau, \lambda$	γ	Mode	2.028		
		Mean	3.571		
		Std. dev.	2.416		
	β	Mean	1.805	1.936	1.756
	τ	Mode		0.671	0.733
		Mean		3.480	3.053
		Std. dev.		4.919	3.849
	λ	Mode		0.924	0.976
		Mean		0.924	0.976
		Std. dev.		0.000	
Goodness of fit	$\log(\mathcal{L})$		-3,023	-3,019	-3,021
	BIC		6,073	6,084	6,076

Note: Standard errors are reported in parentheses.

Table 5.2.: Estimation results of reference point models

only difference between δ^{sep} and δ^{sep}_R is that the forecast error penalty factor γ of the original model is replaced by $(1+\tau)$ in the expectations-based reference point model. The estimates of $\gamma = 3.571$ and $(1+\tau) = 4.053$ are similar in value as well. The same is true for the estimates of the lying aversion factor β in both models ($\beta = 1.805$ in the original model and $\beta = 1.756$ in Reference Point Model 2). Hence, we conclude that a model with a reference forecast $\hat{\phi}^R = \phi$ does not deliver a better explanation of our data than the original model and that the interpretation of the parameters supports a reference point of zero.

We acknowledge that the discussion above only scratches the surface of the complex issue of reference-dependent utility. Subjects might form all kinds of (potentially different and time-varying) reference points based on their individual experience and subjective expectations. We have taken the two most salient pieces of information in our setting and used them to form a reference point with regard to the forecast decision. The analyses indicate that a status quo-based reference point provides a good explanation of the behavior that we observe in our experiments.

5.2. Additional Experiments

5.2.1. Computerized Order Decisions

To further support our choice of behavioral parameters, we tested the extent to which their effects are driven by the interaction between two human beings. We ran an experiment where 16 Sales subjects interacted with a computer instead of a human counterpart, repeating Treatments 5–8 of the main experiment with the same choice of incentive parameters for Sales. Again, half of the subjects played the treatments in the order shown in Table 4.1, and the other half in reverse order. The design of the experiment and the results are summarized in Table 5.3. Subjects received the same briefing as in the main experiment, with the only difference that their forecast now served as data input for a production planning system. We used the rounded standard model predictions for the behavior of Operations as the decision rule for the planning system. For example, in Treatment 5C the system planned to produce the forecast minus 38 units. Subjects were informed about these automated decision rules. They could use a decision support table, which showed that their expected profits were maximized when choosing exactly the same amount of inflation that the system would subsequently deduct from the forecast.

The best response of Sales in the computerized setting varies from the original setting because the order decision of the computer is now a deterministic function $q_c(\hat{\phi}) = \hat{\phi} - c$, where c is the constant that the computer deducts from the forecast. The following proposition summarizes the optimal response of Sales to such an ordering strategy:

Proposition 5. *If the (automated) order decision is* $q_c(\hat{\phi}) = \hat{\phi} - c$ *and if at least* $p_o > 0$ *or*

$\beta > 0$, the best response of Sales is the forecast $\hat{\phi}_c$ that solves the first-order condition

$$
\begin{aligned}
& b\left(1 - G\left(\hat{\phi}_c - \phi - c\right)\right) \\
& \quad = \gamma(p_o + p_u)G\left(\hat{\phi}_c - \phi\right) - \gamma p_u \begin{cases} +\beta \text{ for } p_o - p_u < \frac{2}{\gamma}\left[b(1 - G(-c)) - \beta\right], \\ -\beta \text{ for } p_o - p_u > \frac{2}{\gamma}\left[b(1 - G(-c)) + \beta\right]. \end{cases}
\end{aligned}
\tag{5.3}
$$

For $\frac{2}{\gamma}\left[b(1 - G(-c)) - \beta\right] \leq p_o - p_u \leq \frac{2}{\gamma}\left[b(1 - G(-c)) + \beta\right]$, the best response of Sales is the truthful forecast $\hat{\phi}_c = \phi$.

Treatment ($b/p_o/p_u$)	Forecast correction by computer	Forecast distortion $\hat{\phi} - \phi$: Standard model	Actual average (main experiment)	Actual average (comp. experiment)	Expected cost of Operations
5C (10/6/4)	-38.0	38.1	1.2 (14.4)	29.8 (17.5)	283 (55.3)
6C (10/8/2)	-16.0	15.9	-8.4 (14.2)	10.5 (13.3)	264 (33.2)
7C (10/10/0)	0.0	0.0	-16.7 (17.1)	-3.5 (12.4)	259 (25.2)
8C (10/12/2)	0.0	0.0	-14.8 (14.0)	-2.4 (10.4)	254 (15.4)

Note: Standard deviations are reported in parentheses. Expected cost of Operations are reported in ECUs, excluding fixed compensation C_O. Expectations are taken with respect to E.

Table 5.3.: Design and results of the computerized order decisions experiment

As argued in Section 3.1.2, payoffs in the incentive system of Sales differ in their level of uncertainty when interacting with a human Operations player. While payoffs from forecast error penalties depend on the known uncertainty in the market error E, the sales bonus is ambiguous because it also depends on the (individual) quantity choice of Operations. Even though we set $\alpha = 0.5$ in the main experiment to minimize the potential biases in Operations' order decision, Sales subjects cannot be sure that Operations players update their beliefs as assumed and translate this belief into a rational order decision. Letting a computer make this decision based on a known algorithm removes this ambiguity. In other words, for a given forecast decision, Sales knows the order quantity with certainty and can hence compute the expected sales bonus based on the known uncertainty in the market error E. Consequently, forecast decisions in the computerized experiment should not be biased towards optimizing the forecast error penalty, which we expect to result in a smaller parameter value γ. Note that the other drivers of γ (loss aversion and ex-post forecast error minimization) are not related to the strategic nature of the game and should therefore still be reflected in the forecast error penalty factor γ. Hence, we have the following hypothesis:

Hypothesis 6. *The forecast error penalty factor γ in the computerized experiment is smaller than in the main experiment, but larger than in the standard model, i.e., $1 < \gamma < 3.571$.*

In addition to removing the effect of ambiguity aversion, the modifications in the computerized experiment should also eliminate social concerns (Katok and Wu, 2009). As discussed in Section 3.1.2, we expect Sales to experience a disutility when lying to Operations. In the existing literature, lying aversion has only been studied in economic settings with social interactions. We are not aware of any studies that test whether lying aversion prevails in interactions with nonhuman players. Instead, all concurrent explanations for the phenomenon of lying aversion, ranging from "guilt aversion" (Charness and Dufwenberg, 2006) to people having a "preference for promise-keeping per se" (Vanberg, 2008), suggest that it is rooted in the interaction with another human being. Hence, we do not expect subjects in the computerized experiment to feel a disutility from reporting untruthful forecasts, which is formulated in the next hypothesis:

Hypothesis 7. *The lying aversion factor β in the computerized experiment*

(a) *is smaller than in the main experiment, i.e., $\beta < 1.805$, and*

(b) *does not improve the fit of the model compared to a model without a lying aversion factor.*

Figure 5.1 shows the results of the experiment, i.e., the individual forecast decisions and the resulting order quantities. The dashed lines mark the equilibrium of Proposition 5 based on the standard model assumptions ($\gamma = 1, \beta = 0$). Table 5.3 summarizes the predicted forecast distortions under standard model assumptions, the actual average forecast distortions of the main experiment and the actual average forecast distortions of the computerized experiment. The average forecast distortions of the computerized experiment differ considerably from the forecast distortions of the main experiment (Mann–Whitney U test, $p < 0.001$ for all comparisons). Furthermore, they are not significantly different from the predictions of the standard model (Wilcoxon signed-rank test, $p > 0.140$ for all treatments).

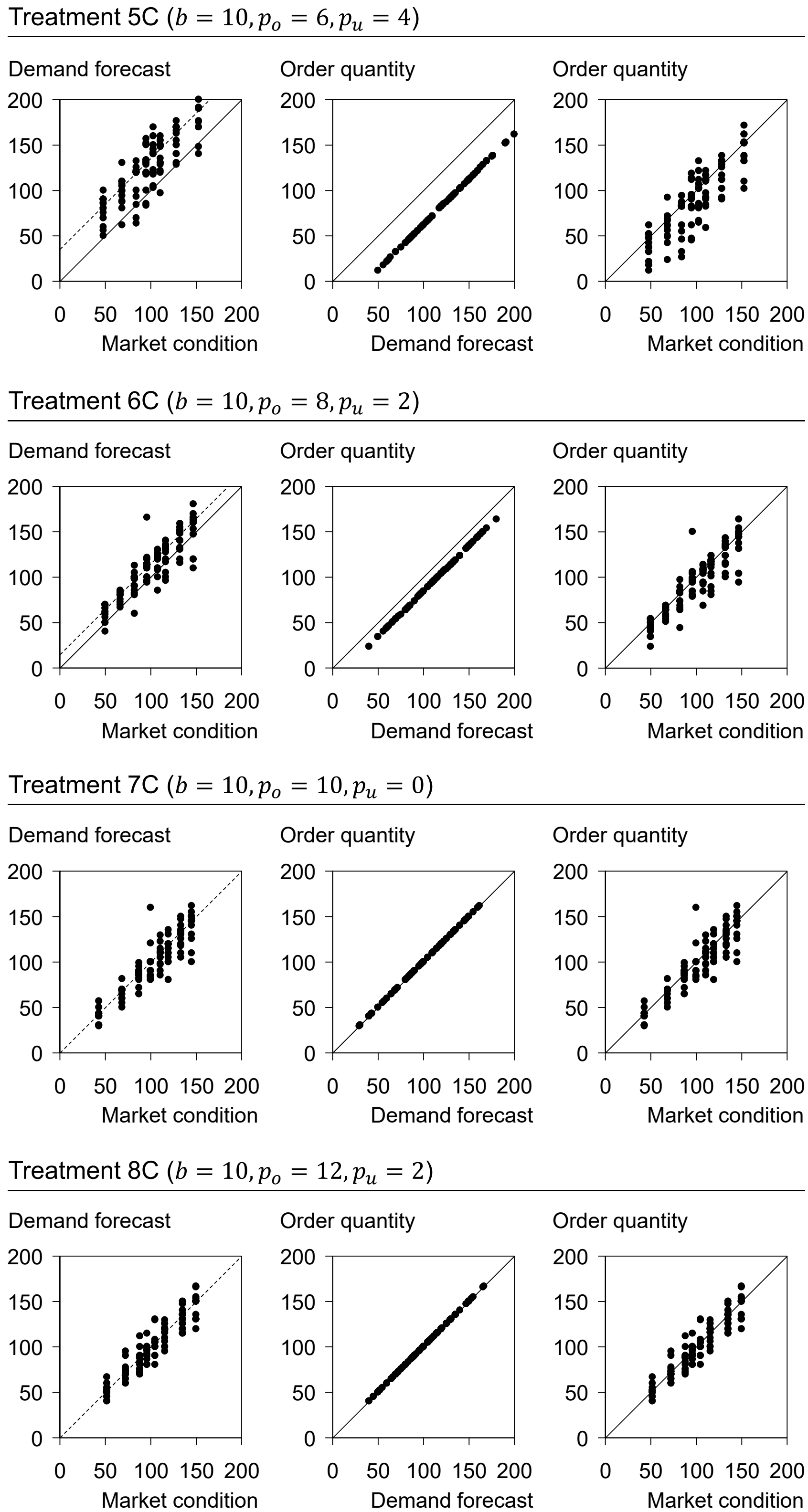


Figure 5.1.: Forecast and order decisions of the computerized order decisions experiment

Since we do not have a closed-form solution for the forecasting behavior of Sales (Equation 5.3), we use subject-level OLS to generate estimates for γ and β (see Computer Model 1 in Table 5.4). In support of Hypothesis 6, the average estimate of $\gamma = 1.585$ is significantly different from both the average ML estimate ($\gamma = 3.571$) and the average OLS estimate ($\gamma = 4.059$) of the main experiment (Mann–Whitney U test, $p < 0.01$ for both comparisons) but significantly greater than one (Wilcoxon signed-rank test, $p = 0.005$). Similarly, in support of Hypothesis 7(a), the average estimate of $\beta = 0.770$ is significantly different from the ML estimate ($\beta = 1.805$, Wilcoxon signed-rank test, $p = 0.008$) and the average OLS estimate ($\beta = 2.610$, Mann–Whitney U test, $p = 0.006$) of the main experiment.

Due to a restriction of $\beta \geq 0$ in the estimation procedure to ensure existence of equilibria, the average estimate of $\beta = 0.770$ is significantly greater than zero (Wilcoxon signed-rank test, $p = 0.008$). Note, however, that the median estimate of $\beta = 0.051$ is close to zero. We therefore estimate a restricted model (Computer Model 2) with $\beta = 0$. Computer Model 2 results in a slightly better aggregate model fit (BIC = 2,538) than Computer Model 1 (BIC = 2,542), which supports Hypothesis 7(b). Moreover, on the level of the individual subject, the unrestricted model (Computer Model 1) delivers a significantly better fit to the data for only two out of 16 subjects (F-test, $p > 0.163$ for 14 subjects, $p < 0.05$ for two subjects). Note that we do not use likelihood ratio tests for model comparison as in the main experiment due to the different estimation procedures.

		Computer Model 1	Computer Model 2
Estimates	γ	1.585 [1.091] (0.881)	1.508 [1.090] (0.804)
	β	0.770 [0.051] (1.698)	
BIC		2,542	2,538

Note: Reported numbers are the mean, [median], and (standard deviation) of the subject-level estimates.

Table 5.4.: Estimation results of the computerized order decisions experiment

While the experiment with automated order decisions does not exclude alternative behavioral factors as drivers for the forecast and order decisions that we observe in the main experiment, it does provide support for our choice of behavioral parameters: Both γ and β change with respect to the direction predicted by their underlying theories when we replace the human Operations player with a computerized one.

5.2.2. Repeated Interactions

In our behavioral model and in the main experiment, we assume that Sales and Operations do not know each other and interact only once. Their decisions are assumed to be independent of any past experience or future expectations regarding the behavior of their counterpart. Interactions between departments of one company, however, are likely to occur repeatedly. In the case of forecast information sharing, this may happen on a monthly basis as shown in the case example of PharmaCo (Chapter2). We therefore ran two additional experimental sessions, where Sales subjects interacted with the same Operations player over the course of 30 periods. The design of the experiment and the results are summarized in Table 5.5. Figure 5.2 shows the individual forecast and order decisions of all subjects. Dashed lines mark the predictions of the one-shot behavioral model ($\gamma = 3.571, \beta = 1.805$).

We chose one treatment of the main experiment with inflated forecasts to model the situation that we frequently observed in practice (Treatment 2) and one treatment with truthful forecasts (Treatment 5). Each treatment was played with a different group of subjects (30 subjects in Treatment 2R and 22 subjects in Treatment 5R). Subjects received the same briefing as in the main experiment, except that they were informed about the fact that they would interact with the same partner throughout the entire experiment. Also, they were told that the experiment would last between 20 and 40 periods in order to avoid endgame behavior. We excluded one observation from the data where a Sales subject accidentally entered an order quantity of 1,000 instead of 100 in period 28, which (s)he reported to the experimenter when collecting the payoff. Excluding all observations of this subject and his/her fellow player does not change the results of our analyses. The analyses below are therefore based on all but this single observation.

Previous research has shown that repeated interactions and their resulting effect on reputation foster cooperation (Dal Bó, 2005). Unlike cooperative actions that are based on reciprocity or fairness (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), cooperative behavior in repeated interactions is mainly due to the fact that the actors “expect future material benefits

Treatment ($b/p_o/p_u$)	Forecast distortion $\hat{\phi} - \phi$ Beh. model	Act. avg. (one-shot)	Act. avg. (repeated)	Forecast correction $q - \hat{\phi}$ Beh. model	Act. avg. (one-shot)	Act. avg. (repeated)	Expected cost of Operations Act. avg. (one-shot)	Act. avg. (repeated)
2R (14/3/3)	19.5	13.3 (16.0)	8.0 (12.3)	-19.5	-15.4 (20.3)	-6.8 (16.2)	308 (87.4)	285 (30.4)
5R (10/6/4)	0.0	1.2 (14.4)	-0.5 (15.3)	0.0	3.2 (13.2)	-4.0 (19.1)	285 (72.1)	291 (30.4)

Note: Standard deviations are reported in parentheses. Expected cost of Operations are reported in ECUs, excluding fixed compensation C_O. Expectations are taken with respect to E.

Table 5.5.: Design and results of the repeated interactions experiment

from their actions" (Fehr and Falk, 2002, p. 689). For example, in the case of signaling games, the use of trigger or review strategies by the signal receiver (in our case Operations) can help to uncover untruthful forecasting behavior (for a discussion and further references, see Ren et al., 2010).

For their game of forecast information sharing with sales-bonus-only incentives, the experimental data of Özer et al. (2011) suggests that forecasts are more truthful in a repeated interaction than a one-time interaction but that some forecast distortion remains if Operations cannot perfectly verify the amount of forecast distortion ex-post. Hence, we formulate the following hypothesis:

Hypothesis 8. *The absolute values of forecast distortions and corrections in a repeated interaction are*

(a) *smaller than those predicted by the one-shot behavioral model, and*

(b) *smaller than those observed in the one-shot experiment.*

For incentive systems where the behavioral model predicts inflated forecasts and where we observe inflated forecasts in the main experiment (Treatment 2R), Hypothesis 8 implies that forecasts will be less inflated in a repeated interaction. For incentive systems where the behavioral model predicts truthful forecasting and where we observe truthful forecasts in the main experiment (Treatment 5R), the hypothesis implies that forecasts will also be truthful in a repeated interaction.

Table 5.5 shows the predictions of the behavioral model ($\gamma = 3.571, \beta = 1.805$) under one-shot assumptions next to the results of the one-shot and repeated interactions experiment. We

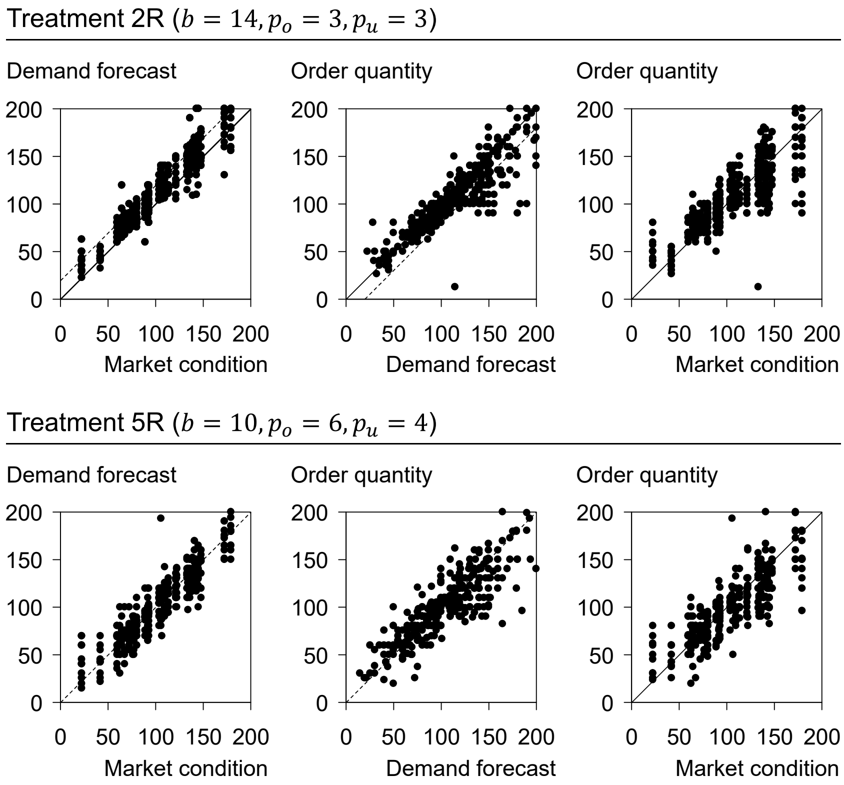


Figure 5.2.: Forecast and order decisions of the repeated interactions experiment

find support for Hypothesis 8(a) in both treatments of the experiment. In Treatment 2R, the actual average forecast inflation in the experiment is 8.0. It is significantly smaller than the behavioral model prediction (Wilcoxon signed-rank test, $p = 0.001$), but significantly greater than zero (Wilcoxon signed-rank test, $p = 0.002$). Similarly, the absolute average forecast correction of 6.8 is significantly smaller than predicted by the behavioral model (Wilcoxon signed-rank test, $p = 0.001$), but significantly greater than zero (Wilcoxon signed-rank test, $p = 0.004$). For Treatment 5R, the one-shot behavioral model predicts truthful information sharing and the results are in line with the prediction: The average forecast distortions and corrections are not significantly different from zero (Wilcoxon signed-rank test, $p > 0.5$ in both cases).

In order to test Hypothesis 8(b), we run GLS random effects regressions based on the data of Treatments 2 and 5 of the main experiment and Treatments 2R and 5R of the repeated interactions experiment. We regress the forecast $\hat{\phi}$ on the market condition ϕ, on a dummy variable $T2$ that differentiates Treatments 2 and 2R ($T2 = 1$) from Treatments 5 and 5R ($T2 = 0$) and on two dummy variables that mark the repeated interaction treatments ($T2R$ and $T5R$). Note that we add two dummy variables for the repeated interaction setting because we expect different effects of the repeated interaction in Treatment 2R compared to Treatment 5R. Due to the different experimental setups of the main experiment (subjects played four different treatments over eight periods each) and the repeated interactions experiment (subjects played only one treatment over 30 periods), there could be other factors than the repeated interaction that affect behavior. We try to control for part of these factors by adding the period t, and the period within a treatment (t_T) as variables that represent the experience of a subject in the regression. We run a similar regression for the order quantity q, except that we use the forecast $\hat{\phi}$ as an explanatory variable instead of the market condition ϕ. The regression functions are as follows:

$$\hat{\phi}_{it} = \text{Intercept} + \eta_{\phi} \cdot \phi_{it} + \eta_{T2}^{FC} \cdot T2_{it} + \eta_{T2R}^{FC} \cdot T2R_{it} + \eta_{T5R}^{FC} \cdot T5R_{it} + \eta_{t}^{FC} \cdot t + \eta_{t_T}^{FC} \cdot t_T + v_i^{FC} + u_{it}^{FC},$$

and

$$q_{it} = \text{Intercept} + \eta_{\hat{\phi}} \cdot \hat{\phi}_{it} + \eta_{T2}^{Q} \cdot T2_{it} + \eta_{T2R}^{Q} \cdot T2R_{it} + \eta_{T5R}^{Q} \cdot T5R_{it} + \eta_{t}^{Q} \cdot t + \eta_{t_T}^{Q} \cdot t_T + v_i^{Q} + u_{it}^{Q},$$

where subjects are indexed by i and periods by t. The error terms are split into a subject-specific part ($v_i^{FC} \sim \mathcal{N}(0, \sigma^2_{v,FC}), v_i^Q \sim \mathcal{N}(0, \sigma^2_{v,Q})$) and a part that is independent across all observations ($u_{it}^{FC} \sim \mathcal{N}(0, \sigma^2_{u,FC}), u_{it}^Q \sim \mathcal{N}(0, \sigma^2_{u,Q})$) to account for the grouped structure of the data.

Table 5.6 summarizes the results of both regressions. Because the overall experience t and the experience within a treatment t_T are highly correlated, including both variables causes a multicollinearity problem. Therefore, we included only the time variable with the higher explanatory power in each of the regressions. Confirming previous results, the treatment dummy variable $T2$ is highly significant and indicates that, on average, subjects forecast 12.254 units more and order 15.744 units less in Treatment 2 (14/3/3) than in Treatment 5 (10/6/4). In line with Hypothesis 8(b), we find that when playing Treatment 2 repeatedly with the

Variable	Forecast $\hat{\phi}$			Quantity q		
Intercept	8.016	***	(2.869)	19.890	***	(3.021)
ϕ (market condition)	0.928	***	(0.009)			
$\hat{\phi}$ (forecast)				0.805	***	(0.012)
$T2$ (14/3/3)	12.254	***	(3.836)	-15.744	***	(3.679)
$T2R$ (14/3/3 repeated)	-5.758		(3.843)	8.573	**	(3.581)
$T5R$ (10/6/4 repeated)	-2.184		(4.169)	-6.119		(3.877)
t (period)				0.168	***	(0.054)
t_T (period within treatment)	0.069	*	(0.043)			

Note: Significance of estimates (z-test): *** p < 0.01, ** p < 0.05, * p < 0.1.
Standard errors are reported in parentheses.

Table 5.6.: Regression results of repeated interaction effects

same partner ($T2R$), subjects forecast less (-5.758 units) and order more (8.573 units) than in the one-shot interaction. This effect is significant for the order quantities ($p = 0.017$), but it does not attain conventional levels of statistical significance for the forecasts ($p = 0.134$). When playing Treatment 5 repeatedly ($T5R$), we do not expect to find behavior that differs from the one-shot interaction and the regressions show that neither forecasts ($p = 0.600$) nor order quantities ($p = 0.115$) are significantly different from those in a one-shot interaction. We conclude that we find directional support for Hypothesis 8(b).

Figure 5.3 shows the average forecast distortions over time for the two treatments. We analyzed the data for time and learning effects by conducting a random effects GLS regression of the forecasts on the market conditions and periods. Period is not a significant explanatory variable (z-test, $p > 0.2$ for both treatments), which indicates that forecasting behavior does not change significantly over time.

Figure 5.4 shows the average forecast corrections over time. The ordering behavior is more volatile than the forecasting behavior, especially in the first half of the experiment. A random effects GLS regression of the order quantities on the forecasts and periods reveals that period is a significant explanatory variable in at least one of the treatments (z-test, $p = 0.001$ in Treatment 2R and $p = 0.104$ in Treatment 5R). In both treatments, order quantities increase over time (0.23 units per period in Treatments 2R and 0.15 units per period in Treatment 5R). This observation is consistent with the regression results in Table 5.6 that Operations subjects

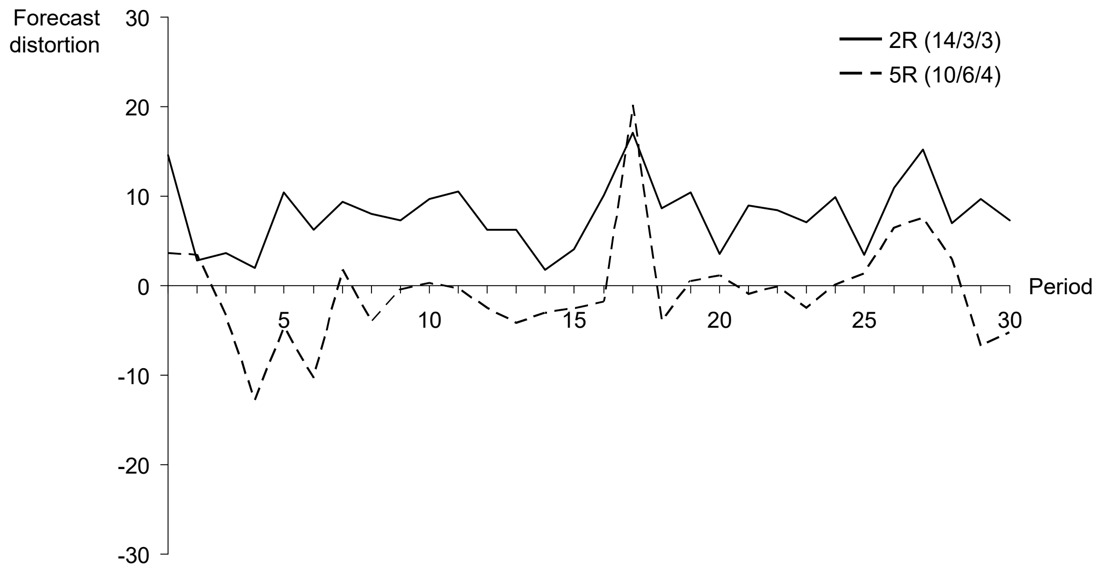


Figure 5.3.: Average forecast distortions over time in the repeated interactions experiment

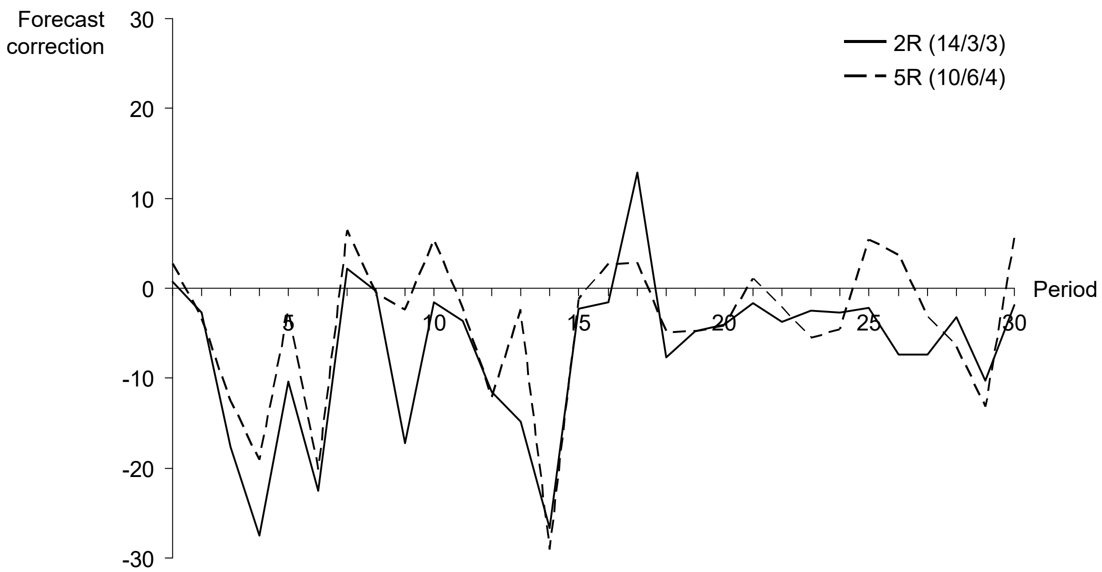


Figure 5.4.: Average forecast corrections over time in the repeated interactions experiment

tend to become slightly more cooperative over time. We will further investigate this finding in the one-shot setting of the main experiment in Section 5.3.1.

We conclude that repeated interactions alleviate the problem of distorted forecast information to some extent, but do not entirely remedy forecast biases in the case of ill-defined forecast incentives. This finding is in line with our observations at PharmaCo, where some bias remained in the forecasts after the implementation of forecast accuracy incentives (see Section 2.3). Additionally, the above analyses suggest that incentive schemes which lead to truthful information sharing in one-time interactions also support truthful forecasts in a repeated game setting. Hence, companies can use the behavioral model of Chapters 3 and 4 to design incentive systems for truthful forecast information sharing despite a potential violation of the one-time interaction assumption.

5.3. Supporting Data Analyses

5.3.1. Analysis of Time and Order Effects

In our main experiment, each subject played multiple periods under multiple treatment conditions. We hence run a set of random effects GLS regressions to identify possible time trends over the course of 32 periods as well as potential treatment order effects caused by the within-subject design used in each of the two sessions (for a detailed discussion on between- and within-subject designs, see Charness et al., 2012).

Time effects.

To test if the behavior of subjects in our experiment changes systematically over time, we regress the forecast $\hat{\phi}$ on the market condition ϕ, dummy variables for seven out of eight treatments ($T2$ to $T8$) and the period t ($t = 1..32$). We run a similar regression for the order quantity q, except that we use the forecast $\hat{\phi}$ as an explanatory variable instead of the market condition ϕ. We formulate the regression functions as

$$\hat{\phi}_{it} = \text{Intercept} + \eta_{\phi} \cdot \phi_{it} + \eta_{T2}^{FC} \cdot T2_{it} + ... + \eta_{T8}^{FC} \cdot T8_{it} + \eta_{t}^{FC} \cdot t + v_{i}^{FC} + u_{it}^{FC}, \qquad (5.4)$$

and

$$q_{it} = \text{Intercept} + \eta_{\hat{\phi}} \cdot \hat{\phi}_{it} + \eta^{Q}_{T2} \cdot T2_{it} + ... + \eta^{Q}_{T8} \cdot T8_{it} + \eta^{Q}_{t} \cdot t + v^{Q}_{i} + u^{Q}_{it}, \tag{5.5}$$

where subjects are indexed by i and periods by t. The error terms are split into a subject-specific part ($v_i^{FC} \sim \mathcal{N}(0, \sigma^2_{v,FC}), v_i^Q \sim \mathcal{N}(0, \sigma^2_{v,Q})$) and a part that is independent across all observations ($u_{it}^{FC} \sim \mathcal{N}(0, \sigma^2_{u,FC}), u_{it}^Q \sim \mathcal{N}(0, \sigma^2_{u,Q})$) to account for the grouped structure of the data. We also tested for learning effects within a treatment by adding another variable to Equations (5.4) and (5.5) that marks the period within a treatment, $t_T = 1..8$, however, the estimates were not significantly different from zero in the forecast or the order quantity regression (z-test, $p > 0.5$).

Variable	Forecast $\hat{\phi}$	Quantity q
Intercept	56.282 *** (3.602)	10.258 *** (3.499)
ϕ (market condition)	0.872 *** (0.017)	
$\hat{\phi}$ (forecast)		0.643 *** (0.014)
$T2$ (14/3/3)	-27.495 *** (2.052)	13.398 *** (1.938)
$T3$ (12/7/7)	-34.061 *** (2.052)	22.116 *** (1.963)
$T4$ (10/10/10)	-37.157 *** (2.052)	25.628 *** (1.979)
$T5$ (10/6/4)	-39.868 *** (4.332)	26.831 *** (3.874)
$T6$ (10/8/2)	-49.373 *** (4.332)	27.740 *** (3.895)
$T7$ (10/10/0)	-57.565 *** (4.332)	30.289 *** (3.916)
$T8$ (10/12/2)	-55.461 *** (4.332)	30.610 *** (3.907)
t (period)	-0.154 *** (0.099)	0.111 ** (0.051)

Note: Significance of estimates (z-test): *** p < 0.01, ** p < 0.05, * p < 0.1.
Standard errors are reported in parentheses.

Table 5.7.: Regression results of time effects in the main experiment

Table 5.7 shows the coefficient estimates for both regressions. The significantly positive estimates for η_ϕ and $\eta_{\hat{\phi}}$ support our previous finding that (a) forecasts depend on the market condition and (b) order decisions are driven by the forecast (compare Section 4.1.5). The comparably low estimate $\eta_{\hat{\phi}} = 0.643$ in the order quantity regression is considerably driven by the sales-bonus-only treatment (Treatment 1). Individual regressions by treatment result in estimates $\eta_{\hat{\phi}} = 0.278$ for Treatment 1 and in estimates $\eta_{\hat{\phi}} \in (0.700; 0.889)$ for all other treatments. The coefficients of the treatment dummies are all significantly different from zero, confirming that the use of forecast error incentive schemes leads to different levels of forecasting and ordering than a sales-bonus-only incentive scheme.

We find time effects in both the forecast decisions of Sales and in the order decisions of Operations. The size of these effects is moderate: On average, Sales inflates the forecast 4.9 units less and Operations orders 3.6 units more at the end of each session compared to the beginning. Both trends indicate an increased level of cooperation over time. However, since both parties adjust their decisions over time, this adaptation is not reflected in the overall operational performance. When regressing the expected cost of Operations on the treatment dummies and the period, the time trend is not significant (z-test, $p = 0.250$).

Interestingly, the time trends in both the forecast and order decisions are at odds with the time trends that Özer et al. (2011) find in their data. Under their sales-bonus-only incentive system, forecast inflation increases and order decisions decrease over time, hence cooperation seems to deteriorate over time. These different observations could be an indicator that the more truthful forecasting behavior under forecast error incentive systems has a positive effect on cooperation over time. However, additional experimental data would be needed to test if this reversal in time effects is truly a result of the different incentive systems or rather a result of the different experimental setups, e.g., due to the use of multiple treatments per subject in our setting versus one treatment per subject in the experiment of Özer et al. (2011).

To analyze whether the time trends found in the above regressions are driven by our experimental design, we run separate regressions by session. We regress the forecast $\hat{\phi}$ on the market condition ϕ, dummy variables for three out of four treatments ($T2$ to $T4$ for Session 1 and $T6$ to $T8$ for Session 2), the period t and a dummy SG that indicates whether a subject played the treatments in the order shown in Table 4.1 ($SG = 0$) or in reverse order ($SG = 1$). We also tested for interaction effects between period and subgroup, but did not find any of significance (z-tests, $p > 0.1$ in all regressions). Again, we run a similar regression for the order quantity q, except that we use the forecast $\hat{\phi}$ as an explanatory variable instead of the market condition ϕ.

Table 5.8 shows the coefficient estimates for both sessions. The time coefficients η_t^{FC} and η_t^{Q} have the same signs as in the regression on the entire data set. With respect to the forecast decisions, the time effect is significant only in Session 2. With respect to the order decisions, the time effect is significant only in Session 1. We conclude that the time trends reported in Table 5.7 are prevalent (though not necessarily significant anymore) in both sessions of our experiment. However, the overall effect is moderate and we limit its impact on behavior in each of the treatments by the use of a reverse order design for half of the subjects. Hence, we do not investigate this issue any further, but acknowledge that it is a by-product of our repeated one-shot interaction design.

Session 1

Variable	Forecast $\hat{\phi}$			Quantity q		
Intercept	54.343	***	(5.863)	14.586	***	(5.617)
ϕ (market condition)	0.902	***	(0.030)			
$\hat{\phi}$ (forecast)				0.569	***	(0.022)
$T2$ (14/3/3)	-27.469	***	(2.581)	11.303	***	(2.304)
$T3$ (12/7/7)	-34.001	***	(2.581)	19.462	***	(2.353)
$T4$ (10/10/10)	-37.068	***	(2.582)	22.679	***	(2.382)
t (period)	-0.141		(0.099)	0.180	**	(0.085)
SG (subgroup)	-2.566		(6.407)	10.182	*	(6.001)

Session 2

Variable	Forecast $\hat{\phi}$			Quantity q		
Intercept	19.321	***	(4.021)	30.216	***	(2.967)
ϕ (market condition)	0.842	***	(0.015)			
$\hat{\phi}$ (forecast)				0.756	***	(0.017)
$T6$ (10/8/2)	-9.475	***	(1.321)	1.883		(1.397)
$T7$ (10/10/0)	-17.634	***	(1.322)	5.247	***	(1.414)
$T8$ (10/12/2)	-15.504	***	(1.323)	5.244	***	(1.406)
t (period)	-0.166	***	(0.051)	0.045		(0.053)
SG (subgroup)	0.453		(5.035)	-6.700	**	(2.997)

Note: Significance of estimates (z-test): *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Standard errors are reported in parentheses.

Table 5.8.: Regression results of time and order effects by session of the main experiment

Order effects.

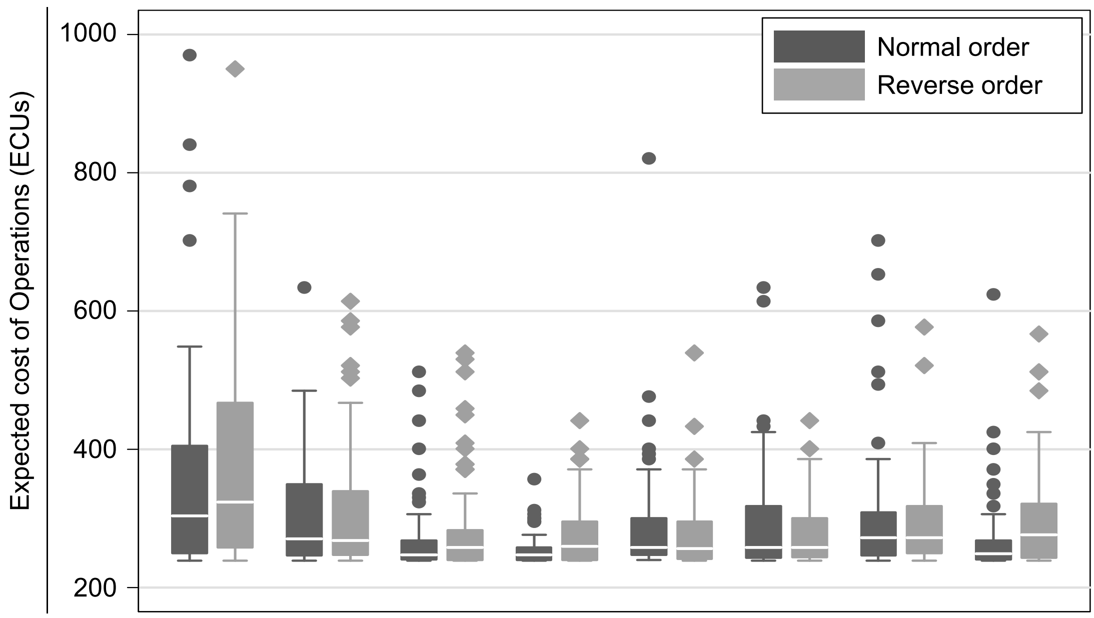
While the regressions on the session level (see Table 5.8) do not point to any systematic differences in the behavior of the subgroups over time, they do show evidence of order effects. The estimates of η^{FC}_{SG} and η^{Q}_{SG} indicate that the order of treatments does not seem to matter for the Sales subjects, but that it plays a significant role in the decisions of the Operations subjects. We use the subgroup-specific statistics reported in Table 5.9 and Figure 5.5 to verify this pattern. In Session 1, Operations subjects who experienced rather truthful forecasting behavior in Treatment 4 (10/10/10) at the beginning of the experiment ordered $\eta^{Q}_{SG} = 10.182$ units more on average than those who experienced high levels of inflation in the sales-bonus-only treatment (Treatment 1 (16/0/0)) first. When regressing the data of each treatment separately, this effect is significant at the 5 % confidence level in Treatment 4 only ($\eta^{Q}_{SG} = 5.163$) even though the estimates in Treatments 1–3 are all positive ($\eta^{Q}_{SG} \in (10.675; 13.000)$). A similar logic applies to Session 2: Operations subjects who were first exposed to rather truthful forecasting behavior in Treatment 5 (10/6/4) place higher orders than the subjects who are confronted with considerable forecast distortions in Treatments 7 (10/10/0) and 8 (10/12/2) at the beginning of the game ($\eta^{Q}_{SG} = -6.700$). Separate regressions by treatment of Session 2 show that the subgroup is a significant determinant of order decisions at the 5 % level in Treatments 5 and 6 ($\eta^{Q}_{SG} = -5.865$ and $\eta^{Q}_{SG} = -6.803$ respectively). While not significant, the estimates in Treatments 7 and 8 have similar values ($\eta^{Q}_{SG} = -6.209$ and $\eta^{Q}_{SG} = -7.479$ respectively). It thus seems that subjects consider the history of the game and the experience of previous treatments even though they play anonymously with different partners in each period.

A possible explanation for the systematically different behavior of the subgroups lies in the concept of reciprocity. Reciprocal people "reward kind actions and punish unkind ones" (Falk and Fischbacher, 2006, p. 293). The mechanism of reciprocity was originally studied in the context of bilateral interactions (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). However, further research in this area has shown that a person A tends to reciprocate the behavior experienced in an interaction with person B to a third person C who was not part of the original interaction (for the concept of "indirect reciprocity," see Nowak and Sigmund, 2005). Hence, Operations subjects who experienced truthful forecasts at the beginning of the

Treatment ($b/p_o/p_u$)	Forecast distortion $\hat{\phi} - \phi$: All subjects	Normal order	Reverse order	Forecast correction $q - \hat{\phi}$: All subjects	Normal order	Reverse order
1 (16/0/0)	40.7 (41.6)	44.4 (53.3)	37.0 (25.0)	-38.9 (46.0)	-48.0 (55.0)	-29.7 (33.1)
2 (14/3/3)	13.3 (16.0)	14.3 (19.4)	12.4 (11.7)	-15.4 (20.3)	-21.0 (16.9)	-9.7 (21.9)
3 (12/7/7)	6.9 (12.3)	7.0 (13.0)	6.8 (11.7)	-3.9 (14.4)	-9.1 (10.2)	1.2 (16.1)
4 (10/10/10)	3.9 (11.8)	4.3 (9.7)	3.6 (13.6)	1.0 (9.7)	-1.6 (6.3)	3.6 (11.7)
5 (10/6/4)	1.2 (14.4)	3.5 (15.5)	-1.0 (12.9)	3.2 (13.2)	5.7 (13.8)	0.6 (12.3)
6 (10/8/2)	-8.4 (14.2)	-8.0 (15.2)	-8.8 (13.4)	7.1 (14.9)	10.4 (16.9)	3.9 (11.9)
7 (10/10/0)	-16.7 (17.1)	-19.1 (20.1)	-14.4 (13.2)	12.3 (15.0)	16.0 (15.1)	8.6 (14.2)
8 (10/12/2)	-14.8 (14.0)	-16.0 (11.8)	-13.5 (15.9)	11.6 (15.2)	15.6 (14.8)	7.5 (14.5)

Note: Standard deviations are reported in parentheses.

Table 5.9.: Summary statistics by order of play of the main experiment



Treatment	1		2		3		4		5		6		7		8	
Average	360	386	304	312	271	290	255	278	293	277	295	281	305	292	270	298
Std. dev.	151	186	77	97	57	77	23	49	86	54	81	48	95	64	59	74

Note: Excluding fixed compensation C_O. Expectations are taken with respect to E.

Figure 5.5.: Expected cost of Operations by order of play of the main experiment

game might reciprocate this seemingly cooperative behavior by choosing higher order quantities in later periods.

While there could be other reasons for why subjects behave differently in the two subgroups, our main interest is whether the existence of order effects biases the results of our main experiment in Section 4.1. Thanks to the reverse order design between the two subgroups in each session, we cover both the case of positive and negative social experience at the beginning of the game in the data of our main experiment. Hence, the spread of decisions in our data might be comparably high, but the data should not be systematically distorted due to experimental order effects. To further substantiate this presumption, we analyze the experimental data based on only the first treatment that a subject played. For subjects in Session 1, we use the data of Treatments 1 and 4, depending on the order of play, and for subjects in Session 2, we use the data of Treatments 5 and 8.

To estimate the behavioral parameters γ and β based on the reduced data set, we run the nonlinear mixed effects ML regression specified in Section 4.1.4 based on the data of Treatments 4, 5 and 8 only. As in the original estimation, we exclude Treatment 1 because we do not have an equilibrium prediction for it. In analogy to Table 4.4, we compare the estimation results of the full data set (Model 3) to the new estimation results based on the reduced data set (Model 7) in Table 5.10. In Model 7, the random effects are degenerated ($\sigma_{\tilde{\gamma}} \approx 0$), which can be attributed to the small sample size (24 subjects, 8 observations per subject). The fixed effect of γ is highly significant (t-test, $p < 0.001$), whereas the fixed effect of β is not significantly different from zero ($p = 0.154$). To test whether both behavioral parameters are needed to explain the data, we run two additional regressions, where only γ (Model 8) or only β (Model 9) are included as (random) explanatory variables. By comparison of the BIC, we find that Model 8 fits the data better than Model 7, while the fit of Model 9 is worse than that of Model 7. Hence, based on the reduced data set, a lying aversion factor $\beta > 0$ does not seem to be needed, whereas the forecast error penalty factor γ improves the explanatory power of the model.

With respect to our hypotheses (see Sections 3.3 and 4.1.5), the reduced data set supports nearly all previous findings. For Hypotheses 2–5, the analyses of the reduced data set deliver

			Model 3	Model 7	Model 8	Model 9
Model description			▪ γ random ▪ β fixed ▪ $n = 768$	▪ γ random ▪ β fixed ▪ $n = 192$	▪ γ random ▪ $\beta = 0$ ▪ $n = 192$	▪ $\gamma = 1$ ▪ β random ▪ $n = 192$
Estimation results	$\mu_{\tilde{\gamma}}$		1.084 (0.125)	0.937 (0.098)	1.154 (0.191)	
	$\sigma_{\tilde{\gamma}}$		0.614	0.000	0.736	
	$\mu_{\tilde{\beta}}$		0.591 (0.133)	0.431 (0.301)		1.300 (0.068)
	$\sigma_{\tilde{\beta}}$					0.008
	σ_u		11.876	14.930	12.901	16.853
Distribution of γ and β	γ	Mode	2.028	2.553	1.845	
		Mean	3.571	2.553	4.160	
		Std. dev.	2.416	0.000	3.529	
	β	Mode	1.805	1.539		3.668
		Mean	1.805	1.539		3.668
		Std. dev.				0.030
Goodness of fit	$\log(\mathcal{L})$		-3,023	-791	-779	-815
	BIC		6,073	1,604	1,575	1,645

Note: Standard errors are reported in parentheses.

Table 5.10.: Estimation results based on the first treatments of the main experiment

similar significant results as the analyses of the full data set. For Hypothesis 1, the findings are mixed: Based on the estimation results above, we find support for the hypothesis that the behavioral model explains actual decisions better than the standard model. However, only the forecast error penalty factor γ seems to be relevant in this context and we do not find evidence for the existence of a lying aversion factor $\beta > 0$. Given the reduced set of treatments, however, this deviation from previous results can be explained. As indicated in Section 4.1.4, there is an overlap in the directional effects of γ and β: In the behavioral model, both parameters predict a reduction in forecast inflation if the standard model predicts forecasts that are higher than the actual market condition. This is the case in Treatments 4 and 5 and the experimental observation is in line with this prediction. In Treatment 8, the standard model predicts truthful forecasts, but average forecasts in the experiment are deflated. This effect can only be explained by a forecast error penalty factor $\gamma > 1$ (a lying aversion factor $\beta > 0$ would support the prediction of the standard model). Hence, only for one third of observations (and subjects) of the reduced data set do the behavioral effects that we investigate influence decisions in different directions. Given that subjects play only one treatment in the reduced data set, the decisions of Treatment 8 could be attributed to comparably low individual values of γ for these subjects, rather than average values of γ whose "downward effect" is partially offset by an "upward effect" of β. Therefore, the forecast error penalty factor γ is sufficient to explain all of the observed behavior in this subset of treatments and the lying aversion factor β does not increase the explanatory power of the model.

To summarize, we acknowledge that there are order effects caused by the within-subject design of our experiment. However, we limited the impact of order effects on our results by the reverse order of play for half of the subject pool. As a result, the spread of decisions in our data set is comparably wide, but we can confirm that almost all of our previous results hold when analyzing a reduced data set which eliminates potential order effects.

5.3.2. Estimation of Treatment-Level Parameters

The parameter estimation in Section 4.1.4 is based on the data of all treatments of the main experiment, where we know that a separating equilibrium exists based on Corollary 1. To analyze, if the magnitude of the behavioral parameters γ and β depends on the incentive parameters, we estimate the behavioral parameters separately for each treatment of the main experiments. Since the small sample size of only eight observations per subject and treatment causes convergence problems in the ML procedure that we use in Section 4.1.4, we estimate the treatment-level parameters based on OLS by subject.

Table 5.11 summarizes the results. The mean estimate of the forecast error penalty factor γ varies from 3.04 in Treatment 3 to a value of 5.98 in Treatment 6. The mean estimate of the lying aversion factor β varies from 0.75 in Treatment 5 to a value of 4.84 in Treatment 7. This comparably wide spread of estimates could be explained by (a) the small sample size of eight data points per regression and (b) the directional overlap of the behavioral parameters that we discuss in Section 4.1.4, which makes it difficult to differentiate between the effects of γ and β in some of the treatments. The high standard deviations of the estimates indicate that the estimates are not very precise. In fact, for the estimates of the forecast error penalty factor γ, we only find a significant difference when comparing Treatment 6 with Treatments 3 (Mann–

Treatment ($b/p_o/p_u$)	Forecast error penalty factor γ Mean	Median	Std. dev.	Lying aversion factor β Mean	Median	Std. dev.
3 (12/7/7)	3.04	2.73	2.19	2.65	3.03	2.09
4 (10/10/10)	3.54	3.27	2.16	2.47	1.62	1.72
5 (10/6/4)	5.13	2.73	3.93	0.75	0.61	0.97
6 (10/8/2)	5.98	6.50	3.59	3.09	1.52	3.52
7 (10/10/0)	4.14	3.18	2.59	4.84	4.34	3.31
8 (10/12/2)	5.35	5.36	3.37	3.52	3.99	3.34
OLS estimates by treatment	4.53	3.36	3.15	2.89	2.52	2.49
OLS estimates across all treatments	4.06	2.23	4.60	2.61	1.57	2.60
Random effects ML estimates	3.57	2.96	2.42	1.80	1.80	n. a.

Table 5.11.: Estimation results by treatment of the main experiment

Whitney U test, $p = 0.036$) and 7 (Wilcoxon signed-rank test, $p = 0.013$). For the estimates of the lying aversion factor β, only the estimates of Treatment 5 differ significantly from the estimates of the other treatments (Mann–Whitney U test for between-subject comparisons, $p \in (0.001; 0.011)$; Wilcoxon signed-rank test for within-subject comparisons, $p \in (0.001; 0.043)$). The last three lines of Table 5.11 provide a comparison of the average treatment-level OLS estimates to average subject-level OLS estimates across all treatments and to the random effects ML estimates of Section 4.1.4. The comparison shows that the average results of the different estimation approaches are similar.

5.3.3. Measures of Quality Control

Due to the complexity of the main experiment, we took different measures to ensure that subjects had a good understanding of the game. In particular, we used the pre-experiment quiz and the post-experiment questionnaire to validate the understanding. Additionally, we provided decision support tables that relieved subjects of calculating expected values, but which allowed them to concentrate on the dynamic decision problem.

Pre-Experiment Quiz.

Before the experiment started, subjects answered a multiple choice quiz on screen to test their understanding of the sequence of events, the basic rules of the probability distribution, and the calculation of profits (see Section A.3.2 of the Appendix). We clustered nine multiple-choice questions in blocks of three on the screen and subjects could only advance to the next screen after having answered all three questions correctly. If one question was answered incorrectly, subjects were only told that there was an error, i.e., the incorrectly answered question was not revealed in order to avoid a "trial-and-error" strategy. They were incentivized by a payment of 1,000 ECUs (= EUR 1) for each block of questions that they answered correctly in the first try. The experiment did not start before all subjects had completed the quiz and answered all questions correctly. To test if subjects understood the rules and logic of the game well, we analyzed the number of errors they made in the pre-experiment quiz. An error occurred if at least one of three questions on the screen was answered incorrectly. The histogram in Figure 5.6 summarizes the results. One third of all subjects (31 %) answered the quiz without any errors. The vast majority of subjects (88 %) had a maximum of three errors. Given that subjects did not receive feedback on which one of three questions they had answered incorrectly, this shows that they had a fairly good understanding of the game.

Hypothetical Order Quantities.

In Section 4.1.1, we argue that a critical ratio of $\alpha = 0.5$ for Operations should lead to comparably unbiased order decisions. We validated this assumption by means of a question in the post-experiment questionnaire (see Section A.3.5 of the Appendix). 31 out of 32 Operations subjects in both sessions answered the questionnaire. We asked them which order quantity they would choose if they knew that the true market condition was 75 (100, 125). The majority of subjects, i.e., 80.6 % (83.9 %, 80.6 %) chose the market condition and hence the optimal order quantity. The remaining subjects chose both too low and too high order quantities, resulting in average order quantities of 76.03 $(101.23, 124.35)$ that are close to and not significantly different from the true market condition (t-test, $p > 0.05$). The histograms in Figure 5.7 show the distribution of choices for each of the market conditions.

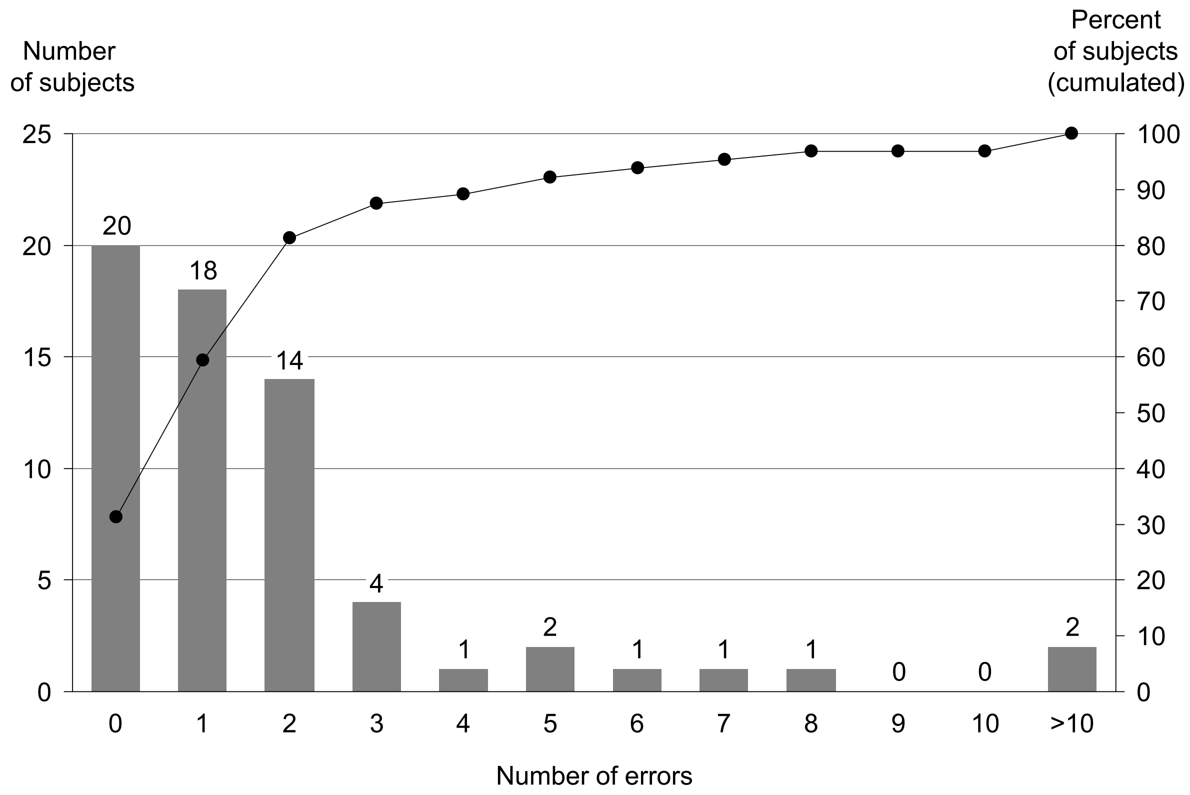


Figure 5.6.: Distribution of errors in the pre-experiment quiz of the main experiment

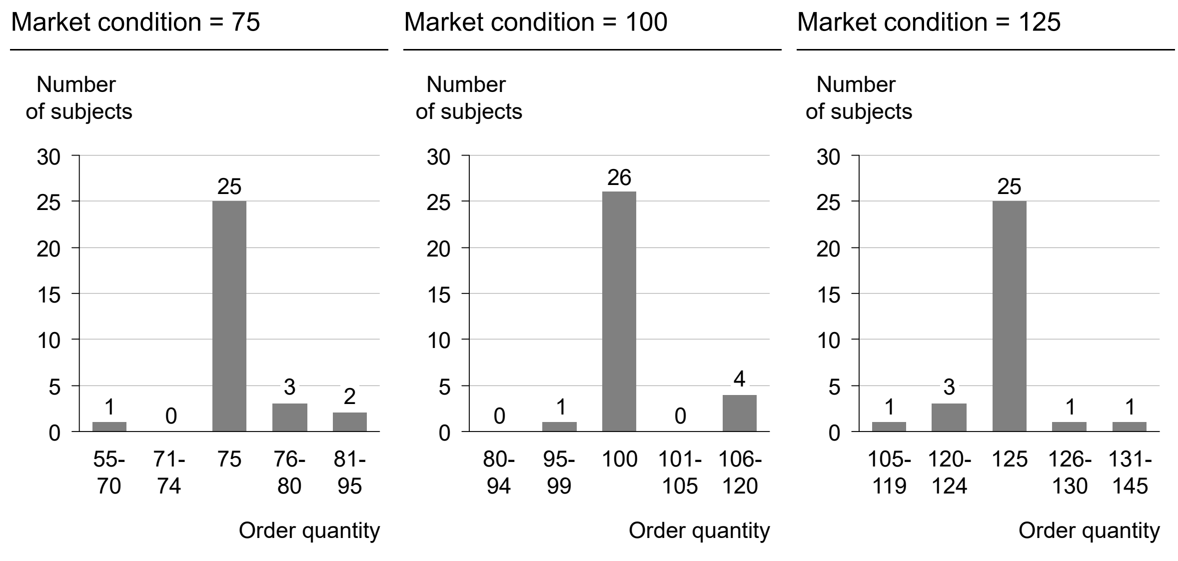


Figure 5.7.: Distribution of hypothetical order quantities of the main experiment

Decision support.

In each period, both Sales and Operations subjects could use a decision support table (see screenshots in Figures A.4 and A.5 of the Appendix). We tracked the use of the button that triggered the display of the table in order to see how intensively the decision support was used. Figure 5.8 displays the share of subjects across periods that looked at the decision support tables.

As expected, the majority (72 %) of both Sales and Operations subjects displayed the table in the first test period. In the remaining test periods, 30 % of Sales and 19 % of Operations subjects on average used the decision support. During the 32 actual periods of the game, 17 % of Sales and 7 % of Operations subjects on average made use of the tables. Given that their decision matrix was static across the entire game, it is not surprising that Operations subjects used the decision support significantly less than Sales subjects (Mann–Whitney U test based on averages by period, $p < 0.001$), whose decision matrix changed based on the incentive system and market condition. Overall, the pattern of use suggests that decision support was needed at the beginning of the game, but that most subjects became confident in their decisions and developed their own strategies over time.

5.3.4. Analysis of Risk Attitudes

After the 32 regular rounds of the main experiment were finished, we gave subjects a set of lottery decisions to elicit their risk attitudes. We followed the basic design of Holt and Laury (2002) and adapted the monetary values to our payout system. Table 5.12 shows the ten decision tasks that we used. In each decision task, subjects had the choice between a comparably safe option (Option A) and a more risky option (Option B). In the end, a random number was drawn to choose which lottery would be played. The respective profit was added to the profit that had previously been achieved in the experiment.

The decision tasks are designed such that the expected payouts of both options are increasing from decision task 1 to 10. However, the expected payout of Option A is increasing slowly, whereas the expected payout of Option B is increasing quickly. In decision task 1, the expected payout of Option A (820 ECUs) is considerably higher than that of Option B (238 ECUs). In

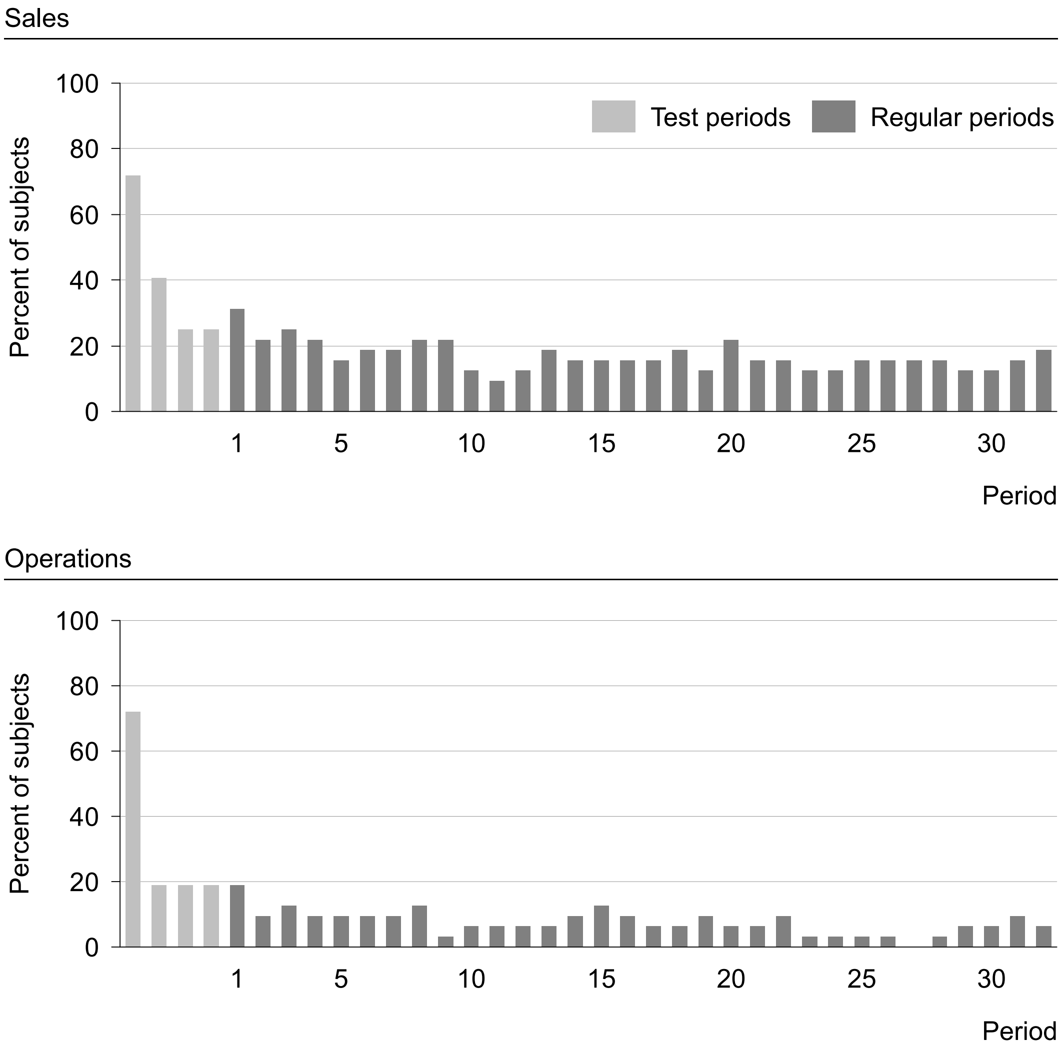


Figure 5.8.: Use of decision support in the main experiment

decision task 9, the expected payout of Option B (1,738 ECUs) is considerably higher than that of Option A (980 ECUs). A risk-neutral decision maker would choose the safe option (A) in decision tasks 1 to 4 and the risky option (B) in decision tasks 5 to 9. Decision task 10 serves as a sanity check since both options are safe payouts, i.e., all subjects, independent of their risk attitudes, should choose Option B with the higher payout. If people assess risks consistently, they should only switch once from Option A to Option B over the course of the ten decisions.

Figure 5.9 shows the distribution of safe choices, i.e., the choices of Option A, across the ten

	Option A		Option B	
Decision	1,000 ECUs	800 ECUs	1,925 ECUs	50 ECUs
1	10%	90%	10%	90%
2	20%	80%	20%	80%
3	30%	70%	30%	70%
4	40%	60%	40%	60%
5	50%	50%	50%	50%
6	60%	40%	60%	40%
7	70%	30%	70%	30%
8	80%	20%	80%	20%
9	90%	10%	90%	10%
10	100%	0%	100%	0%

Table 5.12.: Design of the risk aversion task

decision tasks. The graph is based on the responses of 58 out of 64 subjects. We excluded four subjects due to preference reversals in the sense that they switched multiple times between Options A and B starting with decision task 1 through to 10. We excluded an additional two subjects because they chose the option with the lower safe payout in decision task 10.

Due to the switching point of a risk-neutral decision maker between decisions 4 and 5, we classify all subjects with five or more safe choices as risk-averse (Holt and Laury, 2002). Across all subjects, 82.8 % are risk-averse with slightly more prevalence among Sales subjects (89.7 %) than Operations subjects (75.9 %). The average number of safe choices is 6.2, which also reflects the more risk-averse attitude of Sales subjects with an average value of 6.4 safe choices compared to 5.7 safe choices among Operations subjects. Risk preferences in our subject pool are generally in line with findings in previous research, which show that most people behave in a risk-averse manner (Holt and Laury, 2002; Eckel and Grossman, 2008).

Extending the analyses of Section 5.3.1, we run a random effects GLS regression of the forecast $\hat{\phi}$ on the market condition ϕ, dummy variables for seven out of eight treatments ($T2$ to $T8$), the period t ($t = 1..32$) and the number of safe choices ($SAFE$). We run a comparable regression for the order quantity q, except that we use the forecast $\hat{\phi}$ as an explanatory variable instead of the market condition ϕ. Each regression is based on the data of 29 out of 32 subjects.

Table 5.13 summarizes the regression results. We do not find evidence that the risk attitude

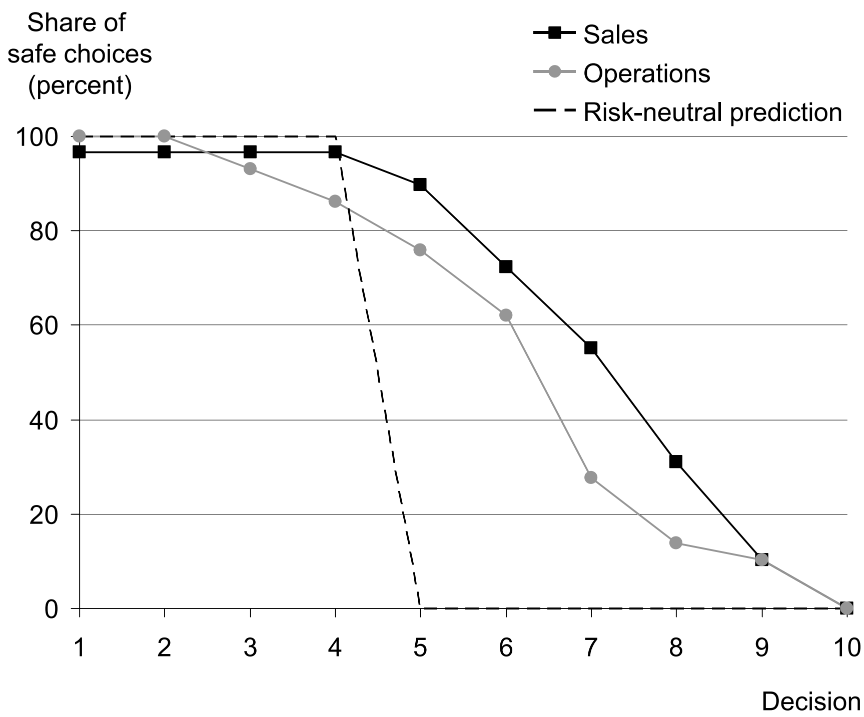


Figure 5.9.: Distribution of safe choices in the main experiment

Variable	Forecast $\hat{\phi}$		Quantity q	
Intercept	51.781 ***	(7.983)	3.247	(5.255)
ϕ (market condition)	0.881 ***	(0.018)		
$\hat{\phi}$ (forecast)			0.628 ***	(0.015)
$T2$ (14/3/3)	-27.102 ***	(2.179)	13.174 ***	(2.032)
$T3$ (12/7/7)	-33.692 ***	(2.180)	22.559 ***	(2.061)
$T4$ (10/10/10)	-36.451 ***	(2.182)	27.375 ***	(2.078)
$T5$ (10/6/4)	-38.128 ***	(4.731)	30.592 ***	(3.123)
$T6$ (10/8/2)	-48.670 ***	(4.731)	31.605 ***	(3.151)
$T7$ (10/10/0)	-57.447 ***	(4.731)	34.142 ***	(3.181)
$T8$ (10/12/2)	-54.941 ***	(4.731)	34.389 ***	(3.170)
t (period)	-0.176 ***	(0.060)	0.051	(0.055)
$SAFE$ (number of safe choices)	0.557	(1.157)	1.090	(0.679)

Note: Significance of estimates (z-test): *** p < 0.01, ** p < 0.05, * p < 0.1.
Standard errors are reported in parentheses.

Table 5.13.: Regression results of risk aversion effects

is a significant explanatory variable in either the forecast decision of Sales (z-test, $p = 0.630$) or in the order decision of Operations ($p = 0.108$). This finding is generally in line with the underlying theories of our behavioral model and the design of the main experiment. As discussed in Section 3.1.2, the elements of the expected payoff of Sales differ in their levels of uncertainty. The forecast error penalty is subject to a quantifiable risk given by the uncertainty in the random market error E. The sales bonus is not only subject to the uncertainty in the market error but also to an unknown uncertainty in the order quantity decision of Operations. In other words, the sales bonus is the result of a compound lottery with an ambiguous stage (the order decision of Operations) followed by a risky stage (the realization of the market error). In our experiments, we never vary the market risk, which is always $E \sim \mathcal{N}(0; 30)$. Hence, it is not surprising that the concept of risk aversion cannot explain the different valuations of the two payoff streams. Instead, as argued before, ambiguity aversion could be a driving force for the increased weight on the forecast error penalty compared to the sales bonus, which is represented by the forecast error penalty factor γ in the utility function of Sales.

6. Conclusion

6.1. Summary of Results

In this thesis, we investigated the impact of forecast-based incentive systems on the accuracy of demand forecasts within a firm. We motivated and contextualized our research by a *case study* of a global pharmaceutical company that introduced forecast accuracy incentives in the target agreements of their sales and marketing personnel. The qualitative results showed that forecast accuracy incentives increased the overall awareness for the importance of good forecasts and improved the collaboration between sales and operations departments. The quantitative analyses showed an improvement in forecast accuracy and a reduction in forecast bias.

Due to the high number of uncontrollable effects in a single case field study, we transferred the forecast information exchange to a *game-theoretic model*. In this model, we considered a company where Sales is responsible for demand forecasting and Operations is responsible for ordering. Our interest was to examine the effect of different forecast error incentive systems for Sales on demand forecasting and order decisions. We developed utility functions of Sales and Operations and included behavioral factors that capture an overproportional reaction to forecast error penalties and an aversion to lying to other people. We modeled the demand forecasting and ordering process as a signaling game and derived the Pareto-dominant separating equilibria of the game.

In a next step, we tested the model in a *laboratory experiment* and showed that actual forecast and order decisions deviate significantly from those of an expected-payoff-maximizing decision maker. We estimated the behavioral parameters to design incentive systems for truthful forecast information sharing. We conducted a validation experiment, which showed that (close to) truthful information sharing can be achieved and that the incentive systems designed with the behavioral model outperform those that were designed with the standard model. We further tested the design of our behavior model by constructing alternative models with non-Bayesian belief updating and expectations-based reference points, but found that our behavioral model provides the better fit to the data of our experiments. We also showed that actual behavior in nonhuman and repeated interactions changes as expected based on the behavioral theories underlying our model.

Our research shows that forecast-based incentives can improve the accuracy of demand forecasts. It further shows that it does not suffice to consider the monetary elements of an incentive system to model forecasting behavior accurately, but that behavioral factors are also important. For incentive system designers, it is critical that they acknowledge the relevance of these factors because incentive systems that ignore them may perform poorly. In settings like the one we described, a natural human aversion to lying and repeated interactions between decision makers can cause forecasts to be more truthful than standard theory would suggest, even without formal incentives. To improve forecast accuracy further, managers can include a forecast-based component in the incentive system for their sales personnel.

6.2. Design Implications

We can use the behavioral model of Chapter 3 and the experimental results of Chapter 4 to derive implications for the design of sales incentive systems in practice. We argued that it is in the best interest of a company to choose an incentive system that enables truthful forecast information sharing. As our analyses show, there are multiple design options that fulfill this condition. In the following, we briefly discuss their differences in order to provide guidance to managers who are interested in reducing forecast biases and increasing the efficiency of their supply chain.

First, an incentive system designer needs to choose the class of incentive system. Both theory (Corollary 2(a)) and experimental results (Treatment 1) suggest that a *sales-bonus-only* incentive system leads to distorted forecasts. Unless the sales bonus is small (such that the natural aversion to lying offsets the incentive to inflate the forecast), means such as using forecast error incentive systems are required to achieve truthful forecasting. With regard to *absolute forecast error* incentive systems, theory predicts that forecasts will be inflated unless the sales bonus is small. In contrast to the sales-bonus-only incentive system, this inflation trends to zero as the forecast error incentives become large (Corollary 2(b) and Treatment 4). A *differentiated forecast error* incentive system can be parameterized for truthful forecast information sharing (Corollary 2(c) and Treatments 5, 9 and 10). In particular, it is more economical in terms of requiring smaller forecast error penalties than an absolute forecast error incentive system to achieve the same level of (potentially small) forecast distortion (Corollary 3).

Our analyses indicate that only differentiated forecast error incentive systems incentivize truthful forecast information sharing. However, there are practical considerations that may favor an absolute forecast error incentive system. For example, an absolute forecast error incentive system is less complex due to the smaller number of parameters. It also has a more intuitive interpretation that forecasts should neither be too high nor too low, which might be a desirable message when introducing forecast error incentives. Finally, if the forecast error incentives are sufficiently high, the remaining bias in the forecasts might be small. Hence, depending on the context, an absolute forecast error incentive system can be the preferred option in practice even if a small forecast bias remains.

Second, an incentive system designer must choose the parameters b, p_o, and p_u. For a differentiated forecast error incentive system, Theorem 1 shows that all parameter combinations with $(p_o - p_u) \in [2(b(1-\alpha) - \beta)/\gamma; 2(b(1-\alpha) + \beta)/\gamma]$ fulfill the condition of zero forecast inflation. For a given unit sales bonus b, the choice of the unit forecast error penalties p_o and p_u depends on the objectives of the company. On the one hand, it may be necessary to limit the impact of forecast error incentives on the incentive system of Sales by setting low unit forecast error penalties. For example, at PharmaCo, the sales and marketing division used the variable compensation budget of their employees to incentivize typical sales targets and was hesitant to include variable compensation elements that were intended to improve operational performance. The smaller p_o and p_u, the smaller the share of the variable compensation budget that is required to support these incentives. On the other hand, as the analysis in Section 4.2.2 shows, there is an argument for robustness that prefers incentive systems with high unit forecast error penalties: The higher p_o and p_u, the less sensitive an equilibrium prediction of (close to) zero inflation is to deviations in the behavioral parameters γ and β. Hence, there is no unanimous recommendation regarding the size of the forecast error penalties. Incentive system designers need to balance both arguments when choosing the parameters for their respective setting.

Even if the particular setup in a company differs from the one that we investigated in our model and experiments, we believe that our general results hold across a range of applications for several reasons: (i) As long as the demand forecast is used to trigger supply, the inherent conflict between sales maximization and forecast error minimization remains. Hence, forecast error penalties can be used to incentivize Sales to forecast truthfully. (ii) With respect to the choice of incentive parameters, the conditions that we derived for a truth-telling incentive scheme allow for a range of truth-telling parameter combinations. A truth-telling solution is therefore quite robust to small changes in the values of both incentive parameters and behavioral parameters. Incentive system designers can furthermore increase the robustness of a truth-telling parameter combination by choosing higher forecast error penalties. (iii) If the interaction between Sales and Operations is repeated over time, the truth-telling parameter combinations of the one-shot model still hold and deviations from the truth-telling solution are even more unlikely. (iv) Beyond the conditions of a truth-telling solution, monotonicity of the distortion function further implies that the direction of change in human behavior is predictable. Hence, in practice, the parameters of an incentive scheme can be calibrated over time to fit a particular setting.

6.3. Boundaries and Future Research

Our research takes a multimethod approach to the question of how to incentivize truthful forecast information sharing within a firm. We have shown that forecast-based incentives have the potential to improve the forecasting performance in theory as well as in practice. The combination of field research and game-theoretic modeling covers two ends of a spectrum. On the one hand, the case study at PharmaCo is complex, rich in information and subject to a large number of uncontrollable factors. On the other hand, our behavioral model is a highly simplified representation of reality that captures a small number of factors only. Both approaches offer valuable insights into the effect of forecast-based incentives on demand forecasts, but are also limited in their very own ways.

First, it is a natural limitation of a *case study* that it is highly specific to the object of investigation and its outcomes might hence not be externally valid (for an extended discussion, see Yin, 2009). For example, the organizational setup, the forecasting processes, and the existing incentive landscape at PharmaCo are probably unique and influence the way forecasts are created and interpreted. Also, the highly regulated market environment of a pharmaceutical company is special and different from that of other industries. Hence, our field research would benefit from replications in other organizations to substantiate the findings of our particular case and to help identify moderating effects that influence the way forecast-based incentives affect forecasting decisions. Another limitation of our case study is that the implementation of forecast accuracy incentives at PharmaCo did not follow a top-down plan but was the result of a process that was shaped by various stakeholders. As researchers, we were primarily observers rather than designers of this process. A fruitful avenue of future research in this respect could be field experiments that allow for more control over the implementation of new incentives. For example, one could test a differentiated forecast error incentive system in one treatment group versus an absolute forecast error incentive system in another treatment group to investigate if the efficiency advantage of our experimental analyses holds in practice.

Second, it is an inherent property of a *theoretical model* that it is a reduced mapping of the real world (see, e.g., the model definition of Stachowiak, 1973). We have designed a model that is reduced to the incentives under investigation and that captures selected behavioral elements. By choosing a high level of reduction, we kept our model analytically tractable and could derive equilibria with the potential to predict human behavior. However, there are many elements of a real world setting that our model does not capture. For example, personal relationships between Sales and Operations, career concerns or other nonmonetary incentives could naturally dampen forecast distortions. This becomes obvious in the case study of PharmaCo, where forecasts are not as distorted as one might expect in the absence of forecast-related incentives. An interesting pursuit in the future would be to test how these factors influence forecast accuracy beyond the effect of forecast-based incentives.

Besides adding factors that our model does not capture, there is also potential to better understand the current parameters of the model. We have used two behavioral parameters to capture various human biases that could be relevant in our setting. Our estimation results combined with additional experiments have shown that these two factors explain actual behavior well and can be used to design incentive systems for truthful forecast information sharing in our setting. However, it would be interesting to analyze the underlying behavioral drivers in more detail to make even better behavioral predictions under different conditions. This could be achieved by decomposing the forecast error penalty factor γ into its basic elements and testing their individual effects in targeted experiments, e.g., by using priming techniques or post-experiment survey questions.

Moreover, incentive systems in organizations can differ from the simple one that we considered and implementations could require adaptations of our incentive system design. For instance, if a company uses a particular performance measure to quantify forecast accuracy (such as the mean absolute percentage error that was used at PharmaCo), it might want to link incentives to this performance measure as opposed to using the penalty function that we used. Given what is known about mental accounting and reference-dependent preferences (see Section 3.1.2), changes in the structure of the incentive system are likely to change human behavior. For example, forecast accuracy targets might provide external reference points that change the evaluation of gains and losses. Previous research also suggests that it matters whether incentives in an organization are common knowledge or private information of the respective actor. The disclosure of incentive conflicts may lead to perverse effects in the sense that it increases the bias in information exchange between an informed and uninformed party (Cain et al., 2005). Hence, future research could investigate the effect of different variants of forecast-based incentives on forecasting decisions and how the level of incentive information matters.

Lastly, there is reason to believe that the effect of economic incentives on behavior does not always follow a monotonic relationship. One line of research in economics suggests that monetary incentives may backfire, i.e., they may reduce intrinsic motivation and crowd out socially desired behavior, such as trust and lying aversion (e.g., Fehr and Fischbacher, 2002; Malhotra and Murnighan, 2002; Bowles, 2009). There is, however, no unanimous answer with regard to the conditions under which this is the case. While some research suggests that incentives must be sufficiently large to avoid such negative effects (Gneezy and Rustichini, 2000), other studies suggest that even small monetary incentives motivate people to be honest when they could obtain a much larger amount by lying (Wang and Murnighan, 2017). It would therefore be interesting to investigate further how the size of forecast-based incentives affects behavior and whether, e.g., small incentives bear the risk of increasing forecast distortions instead of reducing them.

To summarize, there are multiple directions for future research in the field of behavioral OM that emerge from this thesis. Our limited understanding of the complex mechanisms in human decision making offer ample potential for more research on the effect of (forecast-based) incentives under different conditions. A promising and insightful way forward seems to be the combination of theory-based laboratory experiments and field research. On the one hand, controlled experiments allow for a rigorous and comparably simple test of the effect of individual treatment variables. On the other hand, field research ensures that the practically relevant variables are tested and that the experimental results are replicable in practice.

A. Appendix

A.1. Proofs

A.1.1. Proof of Theorem 1

We will first show that $(s^{sep}, q^{sep}, \mu^{sep})$ is a separating equilibrium of the forecast information sharing game if $p_o > (b(1-\alpha)-\beta)/\gamma$. We will then show that this equilibrium Pareto-dominates all other differentiable separating equilibria. We conclude the proof by showing that no differentiable separating equilibrium exists if $p_o \leq (b(1-\alpha)-\beta)/\gamma$. Supporting phase line diagrams and vector field plots are contained in Appendix A.2.

It is straightforward to verify that μ^{sep} satisfies Bayesian consistency of beliefs. Also, $q^{sep}(\hat{\phi}) = \hat{\phi} - \delta^{sep} + G^{-1}(\alpha)$ is a sequentially rational response for Operations to a signal $\hat{\phi}$ given the updated belief $\mu^{sep}(\phi \mid \hat{\phi})$. To see this, rewrite U_O for a given ϕ as

$$U_O(q \mid \phi) = C_O - \gamma_O \left(c_o \int_{-\infty}^{q-\phi} G(x)dx + c_u \left(\mathbb{E}_E(E) + \phi - q + \int_{-\infty}^{q-\phi} G(x)dx \right) \right)$$

and take the first derivative with respect to q: $\frac{\partial U_O(q|\phi)}{\partial q} = -\gamma_O \left((c_u + c_o)G(q-\phi) - c_u\right)$. Setting $\frac{\partial U_O(q|\phi)}{\partial q} = 0$, solving for q and setting $\phi = \hat{\phi} - \delta^{sep}$ according to μ^{sep} yields $q^{sep}(\hat{\phi}) = \hat{\phi} - \delta^{sep} + G^{-1}(\alpha)$. The second derivative $\frac{\partial^2 U_O(q|\phi)}{\partial q^2} = -\gamma_O(c_u + c_o)g(q-\phi)$ confirms that U_O is strictly concave over the support of g and hence q^{sep} is the unique maximizer of U_O given s^{sep} and μ^{sep}. We proceed by showing that s^{sep} is a sequentially rational strategy for Sales in a separating equilibrium.

Suppose that, in equilibrium, Sales adopts some invertible, differentiable signaling function $s : \mathbb{R} \to \mathbb{R}$ (see assumption in Section 3.2). Let $s(\phi)$ be the signal sent by type ϕ. A separating equilibrium requires Operations to be able to infer the type that sent the signal and to react sequentially rationally. Hence, Operations updates the belief to $\mu(\phi \mid s(\phi)) = 1$ if $\phi = s^{-1}(s(\phi))$ and 0 otherwise. Based on this belief, Operations sets $q(s(\phi)) = s^{-1}(s(\phi)) + G^{-1}(\alpha) = \phi + G^{-1}(\alpha)$. For s to be a sequentially rational strategy for Sales given the belief system μ and response function q, there must be no incentive for a type ϕ to mimic some other type $\bar{\phi} \neq \phi$. Hence, in order to verify that the particular strategy s^{sep} satisfies sequential rationality in a separating equilibrium, we will show that the first- and second-order conditions

$$\left.\frac{\partial U_S(q(s(\bar{\phi})), s(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}}\right|_{\bar{\phi}=\phi} = 0, \text{ and}$$

$$\frac{\partial^2 U_S(q(s(\bar{\phi})), s(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}^2} \leq 0 \text{ for all } \bar{\phi},$$

hold for $s = s^{sep}$ at every point where U_S and s^{sep} are differentiable. We first rewrite U_S as

$$\begin{aligned} U_S(q, \hat{\phi} \mid \phi) = C_S + b\left(q - \int_{-\infty}^{q-\phi} G(x)dx\right) - \gamma p_o \int_{-\infty}^{\hat{\phi}-\phi} G(x)dx \\ - \gamma p_u \left(\int_{-\infty}^{\hat{\phi}-\phi} G(x)dx + \mathbb{E}_E(E) - (\hat{\phi} - \phi)\right) - \beta\left|\hat{\phi} - \phi\right|, \end{aligned} \tag{A.1}$$

making use of $\mathbb{E}_E \min(\phi + E, q) = \int_{-\infty}^{q-\phi}(\phi + x)g(x)dx + \int_{q-\phi}^{\infty} qg(x)dx = q - \int_{-\infty}^{q-\phi} G(x)dx$, $\mathbb{E}_E\left[\hat{\phi} - (\phi + E)\right]^+ = \int_{-\infty}^{\hat{\phi}-\phi}((\hat{\phi} - \phi) - x)g(x)dx = \int_{-\infty}^{\hat{\phi}-\phi} G(x)dx$, and $\mathbb{E}_E\left[(\phi + E) - \hat{\phi}\right]^+ = \int_{\hat{\phi}-\phi}^{\infty}(x - (\hat{\phi} - \phi))g(x)dx = \int_{-\infty}^{\hat{\phi}-\phi} G(x)dx + \mathbb{E}_E(E) - (\hat{\phi} - \phi)$. We express the expected utility of Sales in equilibrium for a market condition of ϕ when sending a signal $s(\bar{\phi})$ that triggers reaction $q(s(\bar{\phi}))$ as

$$\begin{aligned} U_S(q(s(\bar{\phi})), s(\bar{\phi}) \mid \phi) = & C_S + b\left(\bar{\phi} + G^{-1}(\alpha) - \int_{-\infty}^{\bar{\phi}-\phi+G^{-1}(\alpha)} G(x)dx\right) - \gamma p_o \int_{-\infty}^{s(\bar{\phi})-\phi} G(x)dx \\ & - \gamma p_u \left(\int_{-\infty}^{s(\bar{\phi})-\phi} G(x)dx + \mathbb{E}_E(E) - (s(\bar{\phi}) - \phi)\right) - \beta\left|s(\bar{\phi}) - \phi\right|. \end{aligned} \tag{A.2}$$

Differentiating Equation (A.2) with respect to $\bar{\phi}$ gives

$$\frac{\partial U_S(q(s(\bar{\phi})), s(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}} = b\left(1 - G\left(\bar{\phi} + G^{-1}(\alpha) - \phi\right)\right) - \gamma\left((p_o + p_u)G(s(\bar{\phi}) - \phi) - p_u\right) s'(\bar{\phi}) \begin{cases} -\beta s'(\bar{\phi}) \text{ for } s(\bar{\phi}) > \phi, \\ +\beta s'(\bar{\phi}) \text{ for } s(\bar{\phi}) < \phi, \end{cases} \tag{A.3}$$

where $s'(\bar{\phi})$ denotes the first derivative of s with respect to $\bar{\phi}$. Evaluating Equation (A.3) at $\bar{\phi} = \phi$ and setting equal to 0 gives the first-order condition as a differential equation:

$$s'(\phi) = \begin{cases} \frac{b(1-\alpha)}{\gamma[(p_o+p_u)G(s(\phi)-\phi)-p_u]+\beta} \text{ for } s(\phi) > \phi, \\ \frac{b(1-\alpha)}{\gamma[(p_o+p_u)G(s(\phi)-\phi)-p_u]-\beta} \text{ for } s(\phi) < \phi. \end{cases} \tag{A.4}$$

Setting $s = s^{sep}$, with $(s^{sep})'(\phi) = 1$, Equation (A.4) holds for $\frac{b(1-\alpha)-\beta}{\gamma} < p_o < 2\frac{b(1-\alpha)-\beta}{\gamma} + p_u$ and for $p_o > 2\frac{b(1-\alpha)+\beta}{\gamma} + p_u$. For $2\frac{b(1-\alpha)-\beta}{\gamma} + p_u \leq p_o \leq 2\frac{b(1-\alpha)+\beta}{\gamma} + p_u$, we have $s^{sep}(\phi) = \phi$ and Equation (A.4) is not defined, but we can verify for this case that the right hand side of Equation (A.3) evaluated at $\bar{\phi} = \phi$ is > 0 for $s(\phi) < \phi$ and < 0 for $s(\phi) > \phi$.

The second derivative of Equation (A.2) is

$$\frac{\partial^2 U_S(q(s(\bar{\phi})), s(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}^2} = -\, bg\left(\bar{\phi} + G^{-1}(\alpha) - \phi\right) - \gamma(p_o + p_u)g(s(\bar{\phi}) - \phi)s'(\bar{\phi})s'(\bar{\phi}) - \gamma\left((p_o + p_u)G(s(\bar{\phi}) - \phi) - p_u\right) s''(\bar{\phi}) \begin{cases} -\beta s''(\bar{\phi}) \text{ for } s(\bar{\phi}) > \phi, \\ +\beta s''(\bar{\phi}) \text{ for } s(\bar{\phi}) < \phi. \end{cases} \tag{A.5}$$

Setting $s = s^{sep}$, with $(s^{sep})'(\bar{\phi}) = 1$ and $(s^{sep})''(\bar{\phi}) = 0$, the right hand side of Equation (A.5) simplifies to $-bg\left(\bar{\phi} + G^{-1}(\alpha) - \phi\right) - \gamma(p_o + p_u)g(\bar{\phi} + \delta^{sep} - \phi) < 0$ for $s^{sep}(\bar{\phi}) \neq \phi$, i.e., U_S is strictly concave in $\bar{\phi}$ for signals $s^{sep}(\bar{\phi})$ to the left and to the right of ϕ. Since we have $\beta \geq 0$ by assumption, concavity extends to $s^{sep}(\bar{\phi}) = \phi$. Hence we can conclude that s^{sep} is a sequentially rational strategy for Sales and that $(s^{sep}, q^{sep}, \mu^{sep})$ is a separating equilibrium of the game.

To prove that $(s^{sep}, q^{sep}, \mu^{sep})$ Pareto-dominates all other differentiable separating equilibria, we analyze all additional candidate equilibrium signaling strategies of Sales in a differentiable separating equilibrium. A candidate equilibrium signaling strategy s_c must be continuous, invertible, differentiable everywhere except for a countable number of points and where it is differentiable it must satisfy Equation (A.4). Where Equation (A.4) is not defined ($s_c(\phi) = \phi$), it must satisfy

$$\lim_{s_c(\phi)\to\phi^-} \left.\frac{\partial U_S(q(s_c(\bar{\phi})), s_c(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}}\right|_{\bar{\phi}=\phi} \geq 0, \text{ and}$$

$$\lim_{s_c(\phi)\to\phi^+} \left.\frac{\partial U_S(q(s_c(\bar{\phi})), s_c(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}}\right|_{\bar{\phi}=\phi} \leq 0.$$

In the following, we first develop some general arguments for the evaluation of potential candidate equilibrium signaling strategies, before differentiating several cases based on the values of b, p_o and p_u. For ease of analysis, we substitute $u(\phi) = s(\phi) - \phi$ and $u'(\phi) = s'(\phi) - 1$ to transform Equation (A.4) into an autonomous differential equation

$$u'(\phi) = \begin{cases} \frac{b(1-\alpha)}{\gamma[(p_o+p_u)G(u(\phi))-p_u]+\beta} - 1 \text{ for } u(\phi) > 0, \\ \frac{b(1-\alpha)}{\gamma[(p_o+p_u)G(u(\phi))-p_u]-\beta} - 1 \text{ for } u(\phi) < 0, \end{cases} \tag{A.6}$$

where u expresses the forecast distortion as a function of the type ϕ. Any candidate equilibrium signaling strategy s_c has a corresponding candidate equilibrium distortion strategy u_c defined by $u_c(\phi) = s_c(\phi) - \phi$.

We start by making use of the requirement that a candidate equilibrium signaling strategy s_c and hence a candidate equilibrium distortion strategy u_c must be continuous.

Lemma 1. *Any candidate equilibrium distortion strategy u_c must have a single interval co-domain $C \subseteq \mathbb{R}$.*

Proof. Proof of Lemma 1. The proof follows directly from the requirement of continuity, i.e., $\lim_{\phi\to\phi_0} u_c(\phi) = u_c(\phi_0)$ for all $\phi_0 \in \mathbb{R}$. □

We next exploit the fact that Equation (A.6) is an autonomous differential equation.

Lemma 2. *Any candidate equilibrium distortion strategy u_c must be monotonic.*

Proof. Proof of Lemma 2. Let u_c be a candidate equilibrium distortion strategy, i.e., u_c is continuous and differentiable everywhere except for a countable number of points and, where it is differentiable, it satisfies Equation (A.6). Suppose u_c is nonmonotonic. Then we can find types $\phi_1 < \phi_2 < \phi_3$ such that we either have $u_c(\phi_2) < \min\{u_c(\phi_1), u_c(\phi_3)\}$ or $u_c(\phi_2) > \max\{u_c(\phi_1), u_c(\phi_3)\}$. We analyze the first case only, the second case is analogous. Also assume that $u_c(\phi_1) \geq u_c(\phi_3)$, the remaining subcase is analogous. Since u_c is nondifferentiable at a countable number of points only and $u_c(\phi_2) < u_c(\phi_3)$, there exists $\phi_4 \in (\phi_2, \phi_3)$ with $u_c(\phi_4) \in (u_c(\phi_2), u_c(\phi_3))$ and $u_c'(\phi_4) > 0$. Denote $u_0 = u_c(\phi_4)$. Let $A = \{\phi \in (\phi_1, \phi_2) \mid u_c(\phi) = u_0\}$. By the intermediate value theorem this set is nonempty and continuity of u_c implies that the maximum of this set is well defined. Denote this maximum by ϕ_5. With u_c being the solution to an autonomous differential equation, we have $u_c'(\phi_5) = u_c'(\phi_4) > 0$ implying there exists some $\epsilon > 0$ with $\epsilon < \phi_2 - \phi_5$ such that $u_c(\phi_5 + \epsilon) > u_c(\phi_5) = u_0$. The intermediate value theorem then implies the existence of $\phi_6 \in (\phi_5 + \epsilon, \phi_2)$ with $u_c(\phi_6) = u_0$. This contradicts ϕ_5 being the maximum of A. We conclude that u_c is monotonic. □

Next, we exploit the requirement that a candidate equilibrium signaling strategy s_c must be invertible. Note first, that the utility function of Sales is separable. Let $U_S^{SB}(q \mid \phi) = b\mathbb{E}_E \min((\phi + E), q)$ be the quantity-dependent and $U_S^{FC}(\hat{\phi} \mid \phi) = -\gamma \mathbb{E}_E(p_o[\hat{\phi} - (\phi + E)]^+ + p_u[(\phi + E) - \hat{\phi}]^+) - \beta \left|\hat{\phi} - \phi\right|$ be the forecast-dependent part of the expected utility function of Sales. By setting $\frac{\partial U_S^{FC}}{\partial \hat{\phi}}$ equal to 0, solving for $\hat{\phi}$ and verifying that $\frac{\partial^2 U_S^{FC}}{\partial \hat{\phi}^2} = -\gamma(p_o + p_u)g(\hat{\phi} - \phi) < 0$ for all $\hat{\phi} \neq \phi$, we find that Sales has a unique preferred forecast $\hat{\phi}^{pref}(\phi) = \phi + \delta^{pref}$ with distortion value

$$\delta^{pref} = \begin{cases} G^{-1}\left(\frac{\gamma p_u - \beta}{\gamma(p_u + p_o)}\right) & \text{for } p_o < p_u - \frac{2\beta}{\gamma}, \\ 0 & \text{for } p_u - \frac{2\beta}{\gamma} \leq p_o \leq p_u + \frac{2\beta}{\gamma}, \\ G^{-1}\left(\frac{\gamma p_u + \beta}{\gamma(p_u + p_o)}\right) & \text{for } p_o > p_u + \frac{2\beta}{\gamma}, \end{cases} \tag{A.7}$$

that maximizes U_S^{FC}.

Lemma 3. *Any candidate equilibrium distortion strategy u_c must take values $u_c(\phi) \geq \delta^{pref}$ for all ϕ.*

Proof. Proof of Lemma 3. Invertibility requires a continuous candidate signaling strategy s_c to be strictly monotonic. Hence (as $u_c(\phi) = s_c(\phi) - \phi$), in the analysis of the candidate distortion strategy u_c we require either $u_c'(\phi) \leq -1$ for all ϕ or $u_c'(\phi) \geq -1$ for all ϕ. Using Equation (A.6) it is straightforward to verify that we have $u_c'(\phi) < -1$ for all $\phi \in \left\{\phi \mid u_c(\phi) < \delta^{pref}\right\}$ and $u_c'(\phi) > -1$ for all $\phi \in \left\{\phi \mid u_c(\phi) > \delta^{pref}\right\}$. Since we have $u_c'(\phi) < -1$ but $u_c(\phi) < \delta^{pref}$ bound from above, there exists no candidate distortion strategy u_c that takes on values $u_c(\phi) < \delta^{pref}$ while covering the entire type space. We conclude that u_c can only take on distortion values $u_c(\phi) \geq \delta^{pref}$. □

We continue by ruling out linear signaling strategies of the form $s_c(\phi) = \phi + \delta$ for $\delta \neq \delta^{sep}$ and hence constant distortion strategies $u_c(\phi) = \delta$ for $\delta \neq \delta^{sep}$. We use the following lemma to further restrict cases where Equation (A.6) is not defined ($u_c(\phi) = 0$ and $u_c(\phi) = \delta^{pref}$).

Lemma 4. *A candidate equilibrium distortion strategy u_c can only take values $u_c(\phi) = \delta$ for all ϕ in some interval $I \subseteq \mathbb{R}$ if $\delta = \delta^{sep}$.*

Proof. Proof of Lemma 4. Assuming that Sales plays strategy $u_c(\phi) = \delta$ for all $\phi \in I$, the sequentially rational response for Operations is $q(\hat{\phi}) = \hat{\phi} - \delta + G^{-1}(\alpha)$ to all signals $\hat{\phi}$ associated with types in I. As defined above, a candidate equilibrium distortion strategy u_c must satisfy Equation (A.6). It is straightforward to verify for $\delta > 0$ (i.e., $u_c(\phi) > 0$) that u_c satisfies Equation (A.6) if and only if $\delta = \delta^{sep} = G^{-1}\left(\frac{b(1-\alpha)+\gamma p_u - \beta}{\gamma(p_u+p_o)}\right)$. Similarly, for $\delta < 0$ (i.e., $u_c(\phi) < 0$) u_c satisfies Equation (A.6) if and only if $\delta = \delta^{sep} = G^{-1}\left(\frac{b(1-\alpha)+\gamma p_u + \beta}{\gamma(p_u+p_o)}\right)$. For $\delta = 0$ we require

$$\lim_{u_c(\phi)\to 0^+} \left.\frac{\partial U_S(q(\bar{\phi} + u_c(\bar{\phi})), \bar{\phi} + u_c(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}}\right|_{\bar{\phi}=\phi} \leq 0, \text{ and}$$

$$\lim_{u_c(\phi)\to 0^-} \left.\frac{\partial U_S(q(\bar{\phi} + u_c(\bar{\phi})), \bar{\phi} + u_c(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}}\right|_{\bar{\phi}=\phi} \geq 0.$$

This holds if and only if $2\frac{b(1-\alpha)-\beta}{\gamma} + p_u \leq p_o \leq 2\frac{b(1-\alpha)+\beta}{\gamma} + p_u$, i.e., if $\delta = \delta^{sep} = 0$. □

Finally, we exploit the requirement of Pareto dominance. In a separating equilibrium, Operations always chooses $q^{FB}(\phi)$. Hence the expected utility of Operations $U_O(q^{FB}(\phi))$ and the quantity-dependent expected utility of Sales $U_S^{SB}(q^{FB}(\phi) \mid \phi)$ are the same in all separating equilibria. It follows that a separating equilibrium with signaling strategy s_1 Pareto dominates another separating equilibrium with signaling strategy s_2 if $U_S^{FC}(s_1(\phi) \mid \phi) \geq U_S^{FC}(s_2(\phi) \mid \phi)$ for all ϕ and $U_S^{FC}(s_1(\phi) \mid \phi) > U_S^{FC}(s_2(\phi) \mid \phi)$ for at least one ϕ. We will also say that one signaling (distortion) strategy Pareto dominates another if the corresponding separating equilibrium Pareto dominates the other.

Lemma 5. *A candidate equilibrium distortion strategy u_c that takes on values $u_c(\phi) > \delta^{sep}$ for all ϕ in some interval $I \subseteq \mathbb{R}$ and takes on values $u_c(\phi) = \delta^{sep}$ for all $\phi \in \mathbb{R} \setminus I$ is Pareto dominated by the strategy u^{sep} that takes on distortion values $u^{sep}(\phi) = \delta^{sep}$ for all $\phi \in \mathbb{R}$.*

Proof. Proof of Lemma 5. We can see from the definitions of δ^{sep} (Equation 3.6) and δ^{pref} (Equation A.7) that we have $\delta^{sep} \geq \delta^{pref}$ for any combination of parameter values. If Sales plays strategy $u_c(\phi)$ then we have $U_S^{FC}(\phi + u_c(\phi) \mid \phi) = U_S^{FC}(\phi + \delta^{sep} \mid \phi)$ for all $\phi \in \mathbb{R} \setminus I$. For all $\phi \in I$, we have $U_S^{FC}(\phi + u_c(\phi) \mid \phi) < U_S^{FC}(\phi + \delta^{sep} \mid \phi)$ because $u_c(\phi) > \delta^{sep} \geq \delta^{pref}$ for all $\phi \in I$ and δ^{pref} is the optimum of the strictly concave function U_S^{FC}. Since Φ is a continuous random variable with domain $\mathbb{R}$, there is a positive probability of having $\phi \in I$ and we conclude that u^{sep} Pareto dominates u_c. □

We now differentiate four cases based on the values of b, p_o, p_u and the properties of Equation (A.6). We use the function w given by

$$w(u) = \begin{cases} \frac{b(1-\alpha)}{\gamma[(p_o+p_u)G(u)-p_u]+\beta} - 1 \text{ for } u > 0, \\ \frac{b(1-\alpha)}{\gamma[(p_o+p_u)G(u)-p_u]-\beta} - 1 \text{ for } u < 0, \end{cases} \tag{A.8}$$

to examine the first derivative of u as a function of u only and hence ease the following expositions. It is straightforward to verify that we have $w'(u) < 0$ always. The following cases are supported by phase line diagrams and example vector field plots in Appendix A.2.

Case 1: $\left(p_o > 2\frac{b(1-\alpha)+\beta}{\gamma} + p_u\right)$. We have $\delta^{sep} = G^{-1}\left(\frac{b(1-\alpha)+\gamma p_u+\beta}{\gamma(p_u+p_o)}\right) < 0$ and $\delta^{pref} = G^{-1}\left(\frac{\gamma p_u+\beta}{\gamma(p_u+p_o)}\right) < 0$ with $\delta^{pref} < \delta^{sep}$ and distinguish seven regions of forecast distortions:

(1.i) For $u > 0$, we have $-1 < w(u) < 0$ with $\lim_{u \to 0^+} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) + \beta} - 1$ and $\lim_{u \to \infty} w(u) = \frac{b(1-\alpha)}{\gamma p_o + \beta} - 1$.

(1.ii) For $u = 0$, $w(u)$ is not defined.

(1.iii) For $\delta^{sep} < u < 0$, we have $-1 < w(u) < 0$ with $\lim_{u \to \delta^{sep+}} w(u) = 0$ and $\lim_{u \to 0^-} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) - \beta} - 1$.

(1.iv) For $u = \delta^{sep}$, we have $w(u) = 0$.

(1.v) For $\delta^{pref} < u < \delta^{sep}$, we have $w(u) > 0$, $w'(u)$ with $\lim_{u \to \delta^{pref+}} w(u) = \infty$ and $\lim_{u \to \delta^{sep-}} w(u) = 0$.

(1.vi) For $u = \delta^{pref}$, $w(u)$ is not defined.

(1.vii) For $u < \delta^{pref}$ we have $w(u) < -1$ with $\lim_{u \to -\infty} w(u) = -\frac{b(1-\alpha)}{\gamma p_u + \beta} - 1$ and $\lim_{u \to \delta^{pref-}} w(u) = -\infty$.

Note the following for this and all subsequent cases: First, given Lemma 1, a candidate equilibrium distortion strategy u_c must take values in one region only or in several adjacent regions. Second, given Lemma 2, in these adjacent regions a strategy u_c must be either all increasing ($w(u_c) \geq 0$) or all decreasing ($w(u_c) \leq 0$).

In case 1, region (1.iv) covers the Pareto-dominant equilibrium distortion strategy $u^{sep}(\phi) = \delta^{sep}$. We rule out all distortion strategies that take values in region (1.vii) by Lemma 3. We further rule out distortion strategies that take more than an isolated value in region (1.vi) by Lemma 4. Also, no candidate equilibrium distortion strategy u_c exists with values in region (1.v) because u_c must be concave ($w(u_c) > 0$ and $w'(u_c) < 0$) in this region and therefore cannot stay in the region for $\phi \to -\infty$. Finally, any candidate equilibrium distortion strategy in regions (1.i) to (1.iii) is Pareto dominated by u^{sep} (see Lemma 5).

Case 2 $\left(2\frac{b(1-\alpha)-\beta}{\gamma} + p_u \leq p_o \leq 2\frac{b(1-\alpha)+\beta}{\gamma} + p_u\right)$. We have $\delta^{sep} = 0$. Based on the value of δ^{pref}, we differentiate two subcases:

Case 2a $\left(p_o > p_u + \frac{2\beta}{\gamma}\right)$. We have $\delta^{pref} = G^{-1}\left(\frac{\gamma p_u + \beta}{\gamma(p_u + p_o)}\right) < 0$ and distinguish five regions of forecast distortions:

(2a.i) For $u > 0$, we have $-1 < w(u) < 0$ with $\lim_{u \to 0^+} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) + \beta} - 1$ and $\lim_{u \to \infty} w(u) = \frac{b(1-\alpha)}{\gamma p_o + \beta} - 1$.

(2a.ii) For $u = \delta^{sep} = 0$, $w(u)$ is not defined.

(2a.iii) For $\delta^{pref} < u < 0$, we have $w(u) > 0$ with $\lim_{u \to \delta^{pref}+} w(u) = \infty$ and $\lim_{u \to 0^-} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) - \beta} - 1$.

(2a.iv) For $u = \delta^{pref}$, $w(u)$ is not defined.

(2a.v) For $u < \delta^{pref}$, we have $w(u) < -1$ with $\lim_{u \to \delta^{pref}-} w(u) = -\infty$ and $\lim_{u \to -\infty} w(u) = -\frac{b(1-\alpha)}{\gamma p_u + \beta} - 1$.

Region (2a.ii) covers the Pareto-dominant equilibrium distortion strategy $u^{sep}(\phi) = \delta^{sep} = 0$. Similar to the line of arguments in case 1, we rule out all distortion strategies that take values in region (2a.v) by Lemma 3. We further exclude distortion strategies that take more than an isolated value in region (2a.iv) by Lemma 4. Also, there does not exist any candidate equilibrium distortion strategy with values in region (2a.iii) because such a strategy must be concave in this region and can hence not stay in the region for all ϕ. Finally, any candidate equilibrium distortion strategy that takes values in region (2a.i) is Pareto dominated according to Lemma 5.

Case 2b $\left(p_o \le p_u + \frac{2\beta}{\gamma}\right)$. We have $\delta^{pref} = 0$ and distinguish three regions of forecast distortions:

(2b.i) For $u > 0$, we have $-1 < w(u) < 0$ with $\lim_{u \to 0^+} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) + \beta} - 1$ and $\lim_{u \to \infty} w(u) = \frac{b(1-\alpha)}{\gamma p_o + \beta} - 1$.

(2b.ii) For $u = \delta^{sep} = \delta^{pref} = 0$, $w(u)$ is not defined.

(2b.iii) For $u < 0$, we have $w(u) < -1$ with $\lim_{u \to 0-} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) - \beta} - 1$ and $\lim_{u \to -\infty} w(u) = -\frac{b(1-\alpha)}{\gamma p_u + \beta} - 1$.

Region (2b.ii) covers the Pareto-dominant equilibrium distortion strategy $u^{sep}(\phi) = \delta^{sep} = \delta^{pref} = 0$. Distortion strategies with values in region (2b.iii) can be ruled out based on Lemma 3 and distortion strategies with values in region (2b.i) are Pareto dominated by u^{sep} (see Lemma 5).

Case 3 $\left(\frac{b(1-\alpha)-\beta}{\gamma} < p_o < 2\frac{b(1-\alpha)-\beta}{\gamma} + p_u\right)$. We have $\delta^{sep} = G^{-1}\left(\frac{b(1-\alpha)+\gamma p_u - \beta}{\gamma(p_u + p_o)}\right) > 0$ and differentiate three subcases based on the value of δ^{pref}:

Case 3a $\left(p_o > p_u + \frac{2\beta}{\gamma}\right)$. We have $\delta^{pref} = G^{-1}\left(\frac{\gamma p_u + \beta}{\gamma(p_u + p_o)}\right) < 0$ and distinguish seven regions of forecast distortions:

(3a.i) For $u > \delta^{sep}$, we have $-1 < w(u) < 0$ with $\lim_{u \to \delta^{sep+}} w(u) = 0$ and $\lim_{u \to \infty} w(u) = \frac{b(1-\alpha)}{\gamma p_o + \beta} - 1$.

(3a.ii) For $u = \delta^{sep}$, we have $w(u) = 0$.

(3a.iii) For $0 < u < \delta^{sep}$, we have $w(u) > 0$ with $\lim_{u \to \delta^{sep-}} w(u) = 0$ and $\lim_{u \to 0^+} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) + \beta} - 1$.

(3a.iv) For $u = 0$, $w(u)$ is not defined.

(3a.v) For $\delta^{pref} < u < 0$, we have $w(u) > 0$ with $\lim_{u \to \delta^{pref+}} w(u) = \infty$ and $\lim_{u \to 0^-} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) - \beta} - 1$.

(3a.vi) For $u = \delta^{pref}$, $w(u)$ is not defined.

(3a.vii) For $u < \delta^{pref}$ we have $w(u) < -1$ with $\lim_{u \to -\infty} w(u) = -\frac{b(1-\alpha)}{\gamma p_u + \beta} - 1$ and $\lim_{u \to \delta^{pref-}} w(u) = -\infty$.

Region (3a.ii) covers the Pareto-dominant equilibrium distortion strategy $u^{sep}(\phi) = \delta^{sep}$. Distortion strategies that take values in region (3a.vii) are not possible due to Lemma 3. We can also rule out distortion strategies that take more than an isolated value in regions (3a.iv) or (3a.vi) by Lemma 4. Further, no candidate equilibrium distortion strategy u_c exists with values in regions (3a.iii), (3a.v) or in a combination of regions (3a.iii)-(3a.v) because u_c must be concave in these regions and therefore cannot cover the entire type space while staying within these regions. Lastly, any candidate equilibrium distortion strategy with values in region (3a.i) is Pareto dominated by u^{sep} according to Lemma 5.

Case 3b $\left(p_u - \frac{2\beta}{\gamma} \leq p_o \leq p_u + \frac{2\beta}{\gamma}\right)$. We have $\delta^{pref} = 0$ and distinguish five regions of forecast distortions:

(3b.i) For $u > \delta^{sep}$, we have $-1 < w(u) < 0$ with $\lim_{u \to \delta^{sep+}} w(u) = 0$ and $\lim_{u \to \infty} w(u) = \frac{b(1-\alpha)}{\gamma p_o + \beta} - 1$.

(3b.ii) For $u = \delta^{sep}$, we have $w(u) = 0$.

(3b.iii) For $0 < u < \delta^{sep}$, we have $w(u) > 0$ with $\lim_{u \to \delta^{sep-}} w(u) = 0$ and $\lim_{u \to 0^+} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) + \beta} - 1$.

(3b.iv) For $u = 0$, $w(u)$ is not defined.

(3b.v) For $u < 0$, we have $w(u) < -1$ with $\lim_{u \to 0^-} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) - \beta} - 1$ and $\lim_{u \to -\infty} w(u) = -\frac{b(1-\alpha)}{\gamma p_u + \beta} - 1$.

Region (3b.ii) covers the Pareto-dominant equilibrium distortion strategy $u^{sep}(\phi) = \delta^{sep}$. Distortion strategies in region (3b.v) are not allowed due to Lemma 3. We can also exclude distortion strategies that take more than an isolated value in region (3b.iv) by Lemma 4. Since any candidate equilibrium distortion strategy with values in region (3b.iii) must be concave in this region, there is no such strategy that stays within the region for all ϕ. Finally, any candidate equilibrium distortion strategy that takes values in region (3b.i) is Pareto dominated by u^{sep} (see Lemma 5).

Case 3c $\left(p_o < p_u - \frac{2\beta}{\gamma}\right)$. We have $\delta^{pref} = G^{-1}\left(\frac{\gamma p_u - \beta}{\gamma(p_u + p_o)}\right) > 0$ and distinguish seven regions of forecast distortions:

(3c.i) For $u > \delta^{sep}$, we have $-1 < w(u) < 0$ with $\lim_{u \to \delta^{sep+}} w(u) = 0$ and $\lim_{u \to \infty} w(u) = \frac{b(1-\alpha)}{\gamma p_o + \beta} - 1$.

(3c.ii) For $u = \delta^{sep}$, we have $w(u) = 0$.

(3c.iii) For $\delta^{pref} < u < \delta^{sep}$, we have $w(u) > 0$ with $\lim_{u \to \delta^{sep-}} w(u) = 0$ and $\lim_{u \to \delta^{pref+}} w(u) = \infty$.

(3c.iv) For $u = \delta^{pref}$, $w(u)$ is not defined.

(3c.v) For $0 < u < \delta^{pref}$, we have $w(u) > 0$ with $\lim_{u \to \delta^{pref-}} w(u) = -\infty$ and $\lim_{u \to 0^+} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) + \beta} - 1$.

(3c.vi) For $u = 0$, $w(u)$ is not defined.

(3c.vii) For $u < 0$, we have $w(u) < -1$ with $\lim_{u \to 0^-} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) - \beta} - 1$ and $\lim_{u \to -\infty} w(u) = -\frac{b(1-\alpha)}{\gamma p_u + \beta} - 1$.

Region (3c.ii) covers the Pareto-dominant equilibrium distortion strategy $u^{sep}(\phi) = \delta^{sep}$. Values in regions (3c.v)-(3c.vii) cannot be part of a candidate equilibrium distortion strategy due to Lemma 3. Also, not more than an isolated point of region (3c.iv) can be part of a candidate equilibrium distortion strategy (see Lemma 4). Any candidate equilibrium distortion strategy with values in region (3c.iii) must be concave in this region and therefore cannot cover the entire type space without leaving the region. Lastly, any candidate equilibrium distortion strategy that takes values in region (3c.i) is Pareto dominated by u^{sep} (see Lemma 5).

Case 4 $\left(p_o \leq \frac{b(1-\alpha)-\beta}{\gamma}\right)$. δ^{sep} does not exist. We differentiate three subcases based on the value of δ^{pref}:

Case 4a $\left(p_o > p_u + \frac{2\beta}{\gamma}\right)$. We have $\delta^{pref} = G^{-1}\left(\frac{\gamma p_u+\beta}{\gamma(p_u+p_o)}\right) < 0$ and distinguish five regions of forecast distortions:

(4a.i) For $u > 0$, we have $w(u) > 0$ with $\lim_{u\to\infty} w(u) = \frac{b(1-\alpha)}{\gamma p_o+\beta} - 1$ and $\lim_{u\to 0^+} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o-p_u)+\beta} - 1$.

(4a.ii) For $u = 0$, $w(u)$ is not defined.

(4a.iii) For $\delta^{pref} < u < 0$, we have $w(u) > 0$ with $\lim_{u\to 0^-} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o-p_u)-\beta} - 1$ and $\lim_{u\to\delta^{pref}+} w(u) = \infty$.

(4a.iv) For $u = \delta^{pref}$, $w(u)$ is not defined.

(4a.v) For $u < \delta^{pref}$, we have $w(u) < -1$ with $\lim_{u\to\delta^{pref}-} w(u) = -\infty$ and $\lim_{u\to-\infty} w(u) = -\frac{b(1-\alpha)}{\gamma p_u+\beta} - 1$.

Distortion strategies with values in region (4a.v) are not allowed due to Lemma 3. Further, a candidate equilibrium distortion strategy can at most take an isolated value in regions (4a.ii) and (4a.iv) based on Lemma 4. Finally, any candidate equilibrium distortion strategy that goes through regions (4a.i) and (4a.iii) must be concave and therefore cannot stay within these regions while covering the entire type space. We conclude that no candidate equilibrium distortion strategy exists in case 4a.

Case 4b $\left(p_u - \frac{2\beta}{\gamma} \le p_o \le p_u + \frac{2\beta}{\gamma}\right)$. We have $\delta^{pref} = 0$ and distinguish three regions of forecast distortions:

(4b.i) For $u > 0$, we have $w(u) > 0$ with $\lim_{u\to\infty} w(u) = \frac{b(1-\alpha)}{\gamma p_o+\beta} - 1$ and $\lim_{u\to 0^+} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o-p_u)+\beta} - 1$.

(4b.ii) For $u = 0$, $w(u)$ is not defined.

(4b.iii) For $u < 0$, we have $w(u) < -1$ with $\lim_{u\to 0^-} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o-p_u)-\beta} - 1$ and $\lim_{u\to-\infty} w(u) = -\frac{b(1-\alpha)}{\gamma p_u+\beta} - 1$.

Distortion strategies in region (4b.iii) are excluded based on Lemma 3. In region (4b.ii), at most an isolated value can be part of a candidate equilibrium distortion strategy (see Lemma 4). Any candidate equilibrium distortion strategy in region (4b.i) must be concave and cannot therefore stay within the region for all ϕ. We conclude that there is no candidate equilibrium distortion strategy in case 4b.

Case 4c $\left(p_o < p_u - \frac{2\beta}{\gamma}\right)$. We have $\delta^{pref} = G^{-1}\left(\frac{\gamma p_u - \beta}{\gamma(p_u + p_o)}\right) > 0$ and distinguish five regions of forecast distortions:

(4c.i) For $u > \delta^{pref}$, we have $w(u) > 0$ with $\lim_{u \to \infty} w(u) = \frac{b(1-\alpha)}{\gamma p_o + \beta} - 1$ and $\lim_{u \to \delta^{pref}+} w(u) = \infty$.

(4c.ii) For $u = \delta^{pref}$, $w(u)$ is not defined.

(4c.iii) For $0 < u < \delta^{pref}$, we have $w(u) < -1$ with $\lim_{u \to \delta^{pref}-} w(u) = -\infty$ and $\lim_{u \to 0^+} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) + \beta} - 1$.

(4c.iv) For $u = 0$, $w(u)$ is not defined.

(4c.v) For $u < 0$, we have $w(u) < -1$ with $\lim_{u \to 0^-} w(u) = \frac{b(1-\alpha)}{\frac{1}{2}\gamma(p_o - p_u) + \beta} - 1$ and $\lim_{u \to -\infty} w(u) = -\frac{b(1-\alpha)}{\gamma p_u - \beta} - 1$.

We rule out distortion strategies with values in regions (4c.iii)-(4c.v) based on Lemma 3. Also, there cannot be a candidate equilibrium distortion strategy that takes more than an isolated value in region (4c.ii) based on Lemma 4. Lastly, any candidate equilibrium distortion strategy with values in region (4c.i) must be concave and therefore cannot cover the entire type space without leaving the region. We conclude that there is no candidate equilibrium distortion strategy in case 4c.

We can thus summarize that, given $p_o > \frac{b(1-\alpha)-\beta}{\gamma}$ (see cases 1-3), s^{sep} is the Pareto-dominant differentiable strategy for Sales in response to the ordering strategy q^{sep}, which is played by Operations in any separating equilibrium. We can further say that for $p_o \leq \frac{b(1-\alpha)-\beta}{\gamma}$ (see case 4), no candidate equilibrium signaling strategy and hence no differentiable separating equilibrium exists. We conclude that whenever differentiable separating equilibria of the general forecast error game exist, the equilibrium $(s^{sep}, q^{sep}, \mu^{sep})$ Pareto dominates all other differentiable separating equilibria. □

A.1.2. Proof of Proposition 1

Recall that the market uncertainty E is symmetrically distributed with mean zero (see Section 3.1.1). Suppose the distribution function of the market uncertainty $G(\cdot)$ is weakly concave on $\mathbb{R}_-$ and weakly convex on $\mathbb{R}_+$. This includes the normal distribution that we use in the experiments.

We use the partial derivatives of δ^{sep} with respect to the behavioral parameters γ and β to express the sensitivity of δ^{sep} to changes in γ and β:

$$\frac{\partial \delta^{sep}}{\partial \gamma} = \begin{cases} -\frac{b(1-\alpha)-\beta}{\gamma^2(p_u+p_o)}(G^{-1})'\left(\frac{b(1-\alpha)+\gamma p_u-\beta}{\gamma(p_u+p_o)}\right) \leq 0 & \text{for } p_o - p_u < 2\frac{b(1-\alpha)-\beta}{\gamma}, \\ 0 & \text{for } 2\frac{b(1-\alpha)-\beta}{\gamma} \leq p_o - p_u \leq 2\frac{b(1-\alpha)+\beta}{\gamma}, \\ -\frac{b(1-\alpha)+\beta}{\gamma^2(p_u+p_o)}(G^{-1})'\left(\frac{b(1-\alpha)+\gamma p_u+\beta}{\gamma(p_u+p_o)}\right) \leq 0 & \text{for } p_o - p_u > 2\frac{b(1-\alpha)+\beta}{\gamma}, \end{cases} \tag{A.9}$$

and

$$\frac{\partial \delta^{sep}}{\partial \beta} = \begin{cases} -\frac{1}{\gamma(p_u+p_o)}(G^{-1})'\left(\frac{b(1-\alpha)+\gamma p_u-\beta}{\gamma(p_u+p_o)}\right) \leq 0 & \text{for } p_o - p_u < 2\frac{b(1-\alpha)-\beta}{\gamma}, \\ 0 & \text{for } 2\frac{b(1-\alpha)-\beta}{\gamma} \leq p_o - p_u \leq 2\frac{b(1-\alpha)+\beta}{\gamma}, \\ +\frac{1}{\gamma(p_u+p_o)}(G^{-1})'\left(\frac{b(1-\alpha)+\gamma p_u+\beta}{\gamma(p_u+p_o)}\right) \geq 0 & \text{for } p_o - p_u > 2\frac{b(1-\alpha)+\beta}{\gamma}. \end{cases} \tag{A.10}$$

Let $d = p_o - p_u = \bar{p}_o - \bar{p}_u$ denote the difference in the unit forecast error penalties and substitute $p_o + p_u = 2p_u + d$. Note that, ceteris paribus, for any d the conditions for the three different cases of δ^{sep} are the same. Note further that $\frac{\partial \delta^{sep}}{\partial \gamma} \leq 0$ in all cases and that $\frac{\partial \delta^{sep}}{\partial \beta} \leq 0$ in the domain of overforecasting (first case) and $\frac{\partial \delta^{sep}}{\partial \beta} \geq 0$ in the domain of underforecasting (third case).

For ease of exposition of the following arguments, we drop the case conditions which are the same as above. We take the cross-partial derivative of $\frac{\partial \delta^{sep}}{\partial \gamma}$ with respect to p_u

$$\frac{\partial \delta^{sep}}{\partial \gamma \partial p_u} = \begin{cases} 2\frac{b(1-\alpha)-\beta}{\gamma^2(2p_u+d)^2}(G^{-1})'\left(\frac{b(1-\alpha)+\gamma p_u-\beta}{\gamma(p_u+p_o)}\right) \\ \quad +\frac{(2(b(1-\alpha)-\beta)-\gamma d)(b(1-\alpha)-\beta)}{\gamma^3(2p_u+d)^3}(G^{-1})''\left(\frac{b(1-\alpha)+\gamma p_u-\beta}{\gamma(p_u+p_o)}\right), \\ 0, \\ 2\frac{b(1-\alpha)+\beta}{\gamma^2(2p_u+d)^2}(G^{-1})'\left(\frac{b(1-\alpha)+\gamma p_u+\beta}{\gamma(p_u+p_o)}\right) \\ \quad +\frac{(2(b(1-\alpha)+\beta)-\gamma d)(b(1-\alpha)+\beta)}{\gamma^3(2p_u+d)^3}(G^{-1})''\left(\frac{b(1-\alpha)+\gamma p_u+\beta}{\gamma(p_u+p_o)}\right), \end{cases} \tag{A.11}$$

to express how the sensitivity of δ^{sep} with respect to γ changes for a given d but different values of p_u. Making use of the case conditions ($d < 2(b(1-\alpha)-\beta)/\gamma$ in the first case and $d > 2(b(1-\alpha)+\beta)/\gamma$ in the third case) and the properties of the distribution of the market uncertainty ($(G^{-1})'(x) \geq 0$ for all x, $(G^{-1})''(x) \leq 0$ for all $x \in (0;0.5)$ and $(G^{-1})''(x) \geq 0$ for all $x \in (0.5;1)$), we can show that $\frac{\partial \delta^{sep}}{\partial \gamma \partial p_u} \geq 0$ for all three cases. Hence, given that δ^{sep} decreases with increasing values of γ ($\frac{\partial \delta^{sep}}{\partial \gamma} \leq 0$), it decreases less strongly for bigger values of p_u and hence p_o.

Similarly, dropping the case conditions again, we take the cross-partial derivative of $\frac{\partial \delta^{sep}}{\partial \beta}$ with respect to p_u

$$\frac{\partial \delta^{sep}}{\partial \beta \partial p_u} = \begin{cases} +\frac{2}{\gamma(2p_u+d)^2}(G^{-1})'\left(\frac{b(1-\alpha)+\gamma p_u-\beta}{\gamma(p_u+p_o)}\right) + \frac{2(b(1-\alpha)-\beta)-\gamma d}{\gamma^2(2p_u+d)^3}(G^{-1})''\left(\frac{b(1-\alpha)+\gamma p_u-\beta}{\gamma(p_u+p_o)}\right), \\ 0, \\ -\frac{2}{\gamma(2p_u+d)^2}(G^{-1})'\left(\frac{b(1-\alpha)+\gamma p_u+\beta}{\gamma(p_u+p_o)}\right) - \frac{2(b(1-\alpha)+\beta)-\gamma d}{\gamma^2(2p_u+d)^3}(G^{-1})''\left(\frac{b(1-\alpha)+\gamma p_u+\beta}{\gamma(p_u+p_o)}\right), \end{cases} \tag{A.12}$$

to express how the sensitivity of δ^{sep} with respect to β changes for a given d but different values of p_u. Making use of the case conditions and the properties of $G(\cdot)$, we can show that $\frac{\partial \delta^{sep}}{\partial \beta \partial p_u} \geq 0$ in the first case (overforecasting) and $\frac{\partial \delta^{sep}}{\partial \beta \partial p_u} \leq 0$ in the third case (underforecasting). Hence, given that δ^{sep} decreases in β in the domain of overforecasting ($\frac{\partial \delta^{sep}}{\partial \beta} \leq 0$) and increases in β in the domain of underforecasting ($\frac{\partial \delta^{sep}}{\partial \beta} \geq 0$), it does so less strongly for bigger values of p_u and hence p_o. $\square$

A.1.3. Proof of Proposition 2

Given a trust-based belief $\mu_\xi(\phi \mid \hat{\phi}) = f_\xi(\phi \mid \hat{\phi})$, where f_ξ is the density function of $\xi\hat{\phi}+(1-\xi)\Phi$ and $\xi \in (0;1]$, it is straightforward to verify that the best response of Operations to a forecast $\hat{\phi}$ is $q_\xi(\hat{\phi}) = \xi\hat{\phi} + F_Z^{-1}(\alpha)$, where F_Z is the distribution function of $Z = (1-\xi)\Phi + E$. To see this, note that Operations faces a newsvendor problem with demand distribution $\xi\hat{\phi} + (1-\xi)\Phi + E$.

To derive the optimal response of Sales to an order strategy q_ξ, we differentiate $U_S(q_\xi(\hat{\phi}), \hat{\phi} \mid$

ϕ) with respect to $\hat{\phi}$

$$\frac{\partial U_S(q_\xi(\hat{\phi}), \hat{\phi} \mid \phi)}{\partial \hat{\phi}} = b\xi \left(1 - G\left(\xi\hat{\phi} + F_Z^{-1}(\alpha) - \phi\right)\right) - \gamma(p_o + p_u)G(\hat{\phi} - \phi) + \gamma p_u \begin{cases} -\beta & \text{for } \hat{\phi} > \phi, \\ +\beta & \text{for } \hat{\phi} < \phi, \end{cases} \tag{A.13}$$

and set $\frac{\partial U_S}{\partial \hat{\phi}} = 0$ to obtain the first-order condition

$$b\xi \left(1 - G\left(\xi\hat{\phi}_\xi + F_Z^{-1}(\alpha) - \phi\right)\right) = \gamma(p_o + p_u)G\left(\hat{\phi}_\xi - \phi\right) - \gamma p_u \begin{cases} +\beta \text{ for } \hat{\phi}_\xi > \phi, \\ -\beta \text{ for } \hat{\phi}_\xi < \phi, \end{cases} \tag{A.14}$$

where $\hat{\phi}_\xi$ is a forecast that solves the first-order condition.

We confirm concavity of U_S over $\mathbb{R} \setminus \{\phi\}$ by means of the second derivative

$$\frac{\partial^2 U_S(q_\xi(\hat{\phi}), \hat{\phi} \mid \phi)}{\partial \hat{\phi}^2} = -b\xi^2 g\left(\xi\hat{\phi} + F_Z^{-1}(\alpha) - \phi\right) - \gamma(p_o + p_u)g(\hat{\phi} - \phi) < 0 \text{ for all } \hat{\phi} \neq \phi.$$

Given that U_S is concave on $(-\infty, \phi)$ and on (ϕ, ∞) and we have $\lim_{\hat{\phi}\to\phi^-} \frac{\partial U_S}{\partial \hat{\phi}} \geq \lim_{\hat{\phi}\to\phi^+} \frac{\partial U_S}{\partial \hat{\phi}}$ for all ϕ, there is at most one solution to Equation (A.14) for a given ϕ. If that solution exists, it maximizes the utility of Sales.

We next derive the conditions under which a solution to Equation (A.14) exists. For $p_o > -\beta/\gamma$, we have $\lim_{\hat{\phi}\to-\infty} \frac{\partial U_S}{\partial \hat{\phi}} > 0$ and $\lim_{\hat{\phi}\to\infty} \frac{\partial U_S}{\partial \hat{\phi}} < 0$. It follows that there is either a solution to Equation (A.14) or we have $\lim_{\hat{\phi}\to\phi^-} \frac{\partial U_S}{\partial \hat{\phi}} > 0$ and $\lim_{\hat{\phi}\to\phi^+} \frac{\partial U_S}{\partial \hat{\phi}} < 0$, which implies that the truthful forecast $\hat{\phi}_\xi = \phi$ is the best response of Sales.

For $p_o \leq -\beta/\gamma$, we have $\lim_{\hat{\phi}\to\infty} \frac{\partial U_S}{\partial \hat{\phi}} \geq 0$, i.e., there exists no solution to Equation (A.14). Because we have $p_o \geq 0$, $\beta \geq 0$ and $\gamma \geq 1$ by definition, this case can only occur if we have $p_o = \beta = 0$ and hence $\lim_{\hat{\phi}\to\infty} \frac{\partial U_S}{\partial \hat{\phi}} = 0$. We conclude that there exists a unique equilibrium response $\hat{\phi}_\xi$ for all Sales types ϕ if at least $p_o > 0$ or $\beta > 0$.

For ease of interpretation, we rewrite the case conditions of Equation (A.14) in terms of the incentive parameters p_o and p_u by replacing $\hat{\phi}_\xi > \phi$ with the requirement

$$\lim_{\hat{\phi}\to\phi^+} \frac{\partial U_S}{\partial \hat{\phi}} > 0 \quad \Leftrightarrow \quad p_o - p_u < \frac{2}{\gamma} \left[b\xi \left(1 - G\left((\xi-1)\phi + F_Z^{-1}(\alpha)\right)\right) - \beta \right], \tag{A.15}$$

and by replacing $\hat{\phi}_\xi < \phi$ with the requirement

$$\lim_{\hat{\phi}\to\phi^-} \frac{\partial U_S}{\partial \hat{\phi}} < 0 \quad \Leftrightarrow \quad p_o - p_u > \frac{2}{\gamma} \left[b\xi \left(1 - G\left((\xi-1)\phi + F_Z^{-1}(\alpha)\right)\right) + \beta \right]. \tag{A.16}$$

□

A.1.4. Proof of Proposition 3

In analogy to the proof of Theorem 1 (Appendix A.1.1), it is straightforward to verify that the best response of Operations given a naïveté-based belief

$$\mu_\theta(\phi \mid \hat{\phi}) = \begin{cases} 1 \text{ for } \phi = \theta\hat{\phi} + (1-\theta)s^{-1}(\hat{\phi}), \\ 0 \text{ otherwise,} \end{cases}$$

is $q_\theta(\hat{\phi}) = \theta\hat{\phi} + (1-\theta)s^{-1}(\hat{\phi}) + G^{-1}(\alpha)$.

To derive the optimal response $s(\phi)$ of Sales to an order strategy q_θ, suppose a Sales type ϕ sent a signal $s(\bar{\phi})$ corresponding to some type $\bar{\phi}$. We first differentiate $U_S(q_\theta(s(\bar{\phi})), s(\bar{\phi}) \mid \phi)$ with respect to $\bar{\phi}$:

$$\begin{aligned} \frac{\partial U_S(q_\theta(s(\bar{\phi})), s(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}} =& b\left(1 - G\left(\theta s(\bar{\phi}) + (1-\theta)\bar{\phi} + G^{-1}(\alpha) - \phi\right)\right)\left(\theta s'(\bar{\phi}) + (1-\theta)\right) \\ &- \gamma\left((p_o + p_u)G(s(\bar{\phi}) - \phi) - p_u\right)s'(\bar{\phi}) \begin{cases} -\beta s'(\bar{\phi}) & \text{for } s(\bar{\phi}) > \phi, \\ +\beta s'(\bar{\phi}) & \text{for } s(\bar{\phi}) < \phi. \end{cases} \end{aligned} \tag{A.17}$$

Assuming a linear signaling strategy $s_\theta(\bar{\phi}) = \bar{\phi} + \delta_\theta$, Equation (A.17) simplifies to

$$\begin{aligned}\frac{\partial U_S(q_\theta(s_\theta(\bar{\phi})), s_\theta(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}} =& b\left(1 - G\left(\theta\delta_\theta + \bar{\phi} + G^{-1}(\alpha) - \phi\right)\right) \\ & - \gamma(p_o + p_u)G(\bar{\phi} + \delta_\theta - \phi) + \gamma p_u \begin{cases} -\beta & \text{for } \bar{\phi} + \delta_\theta > \phi, \\ +\beta & \text{for } \bar{\phi} + \delta_\theta < \phi. \end{cases}\end{aligned} \tag{A.18}$$

A sophisticated belief is based on the assumption that Operations can infer the true market condition ϕ from the signal $\hat{\phi}$. We hence require Sales to signal the true type and we thus evaluate Equation (A.18) at $\bar{\phi} = \phi$

$$\begin{aligned}\left.\frac{\partial U_S(q_\theta(s_\theta(\bar{\phi})), s_\theta(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}}\right|_{\bar{\phi}=\phi} =& b\left(1 - G\left(\theta\delta_\theta + G^{-1}(\alpha)\right)\right) \\ & - \gamma(p_o + p_u)G(\delta_\theta) + \gamma p_u \begin{cases} -\beta & \text{for } \delta_\theta > 0, \\ +\beta & \text{for } \delta_\theta < 0. \end{cases}\end{aligned} \tag{A.19}$$

We set Equation (A.19) equal to zero to derive the first-order condition for the best response of Sales

$$b\left[1 - G\left(\theta\delta_\theta + G^{-1}(\alpha)\right)\right] = \gamma\left[(p_o + p_u)G(\delta_\theta) - p_u\right] \begin{cases} +\beta \text{ for } \delta_\theta > 0, \\ -\beta \text{ for } \delta_\theta < 0. \end{cases} \tag{A.20}$$

We check concavity of U_S by means of the second derivative

$$\begin{aligned}\frac{\partial^2 U_S(q_\theta(s_\theta(\bar{\phi})), s_\theta(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}^2} =& - bg\left(\theta\delta_\theta + \bar{\phi} + G^{-1}(\alpha) - \phi\right) \\ & - \gamma(p_o + p_u)g(\bar{\phi} + \delta_\theta - \phi) \leq 0 \text{ for } \bar{\phi} + \delta_\theta \neq \phi.\end{aligned} \tag{A.21}$$

Since we have $\beta \geq 0$ by assumption, it is straightforward to verify that concavity of U_S extends to $\bar{\phi} + \delta_\theta = \phi$. Hence, if a solution to the first-order condition of Equation (A.20) exists, it is unique and it maximizes the utility of Sales.

We next derive the conditions under which a solution to Equation (A.20) exists. Note that we can treat Equation (A.19) as a sectionwise continuous function of δ_θ. For $p_o > -\beta/\gamma$, we have

$$\lim_{\delta_\theta \to -\infty} \frac{\partial U_S}{\partial \bar{\phi}}\bigg|_{\bar{\phi}=\phi} > 0 \text{ and } \lim_{\delta_\theta \to \infty} \frac{\partial U_S}{\partial \bar{\phi}}\bigg|_{\bar{\phi}=\phi} < 0.$$

It follows that there is either a solution to Equation (A.20) or we have

$$\lim_{\delta_\theta \to 0^-} \frac{\partial U_S}{\partial \bar{\phi}}\bigg|_{\bar{\phi}=\phi} > 0 \text{ and } \lim_{\delta_\theta \to 0^+} \frac{\partial U_S}{\partial \bar{\phi}}\bigg|_{\bar{\phi}=\phi} < 0,$$

and hence $\delta_\theta = 0$ is the best response of Sales. For $p_o \leq -\beta/\gamma$, we have

$$\lim_{\delta_\theta \to \infty} \frac{\partial U_S}{\partial \bar{\phi}}\bigg|_{\bar{\phi}=\phi} \geq 0,$$

i.e., there exists no solution to Equation (A.20). Because we have $p_o \geq 0$, $\beta \geq 0$ and $\gamma \geq 1$ by definition, this case can only occur if we have $p_o = \beta = 0$ and hence

$$\lim_{\delta_\theta \to \infty} \frac{\partial U_S}{\partial \bar{\phi}}\bigg|_{\bar{\phi}=\phi} = 0.$$

We conclude that there exists a unique equilibrium distortion value δ_θ for all Sales types ϕ if at least $p_o > 0$ or $\beta > 0$.

For ease of interpretation, we rewrite the case conditions of Equation (A.20) in terms of the incentive parameters p_o and p_u by replacing $\delta_\theta > 0$ with the requirement

$$\lim_{\delta_\theta \to 0^+} \frac{\partial U_S}{\partial \bar{\phi}}\bigg|_{\bar{\phi}=\phi} > 0 \quad \Leftrightarrow \quad p_o - p_u < \frac{2}{\gamma}\left(b\left(1-\alpha\right)-\beta\right), \tag{A.22}$$

and by replacing $\delta_\theta < 0$ with the requirement

$$\lim_{\delta_\theta \to 0^-} \frac{\partial U_S}{\partial \bar{\phi}}\bigg|_{\bar{\phi}=\phi} < 0 \quad \Leftrightarrow \quad p_o - p_u > \frac{2}{\gamma}\left(b\left(1-\alpha\right)+\beta\right). \tag{A.23}$$

□

A.1.5. Proof of Proposition 4

In analogy to the proof of Theorem 1, it is straightforward to verify that μ_R^{sep} satisfies Bayesian consistency of beliefs and that q_R^{sep} is a sequentially rational response of Operations given μ_R^{sep}. It remains to show that s_R^{sep} is a sequentially rational response of Sales given q_R^{sep}.

We first resolve the expected values and rewrite U_S^R as

$$\begin{aligned} U_S^R(q,\hat{\phi} \mid \phi) = C_S &+ b\left(q - \int_{-\infty}^{q-\phi} G(x)dx\right) - \gamma p_o \int_{-\infty}^{\hat{\phi}-\phi} G(x)dx \\ &- \gamma p_u \left(\int_{-\infty}^{\hat{\phi}-\phi} G(x)dx + \mathbb{E}_E(E) - (\hat{\phi}-\phi)\right) \\ &\begin{cases} +\tau p_u \left(\hat{\phi}-\phi-\int_0^{\hat{\phi}-\phi} G(x)dx\right) - \lambda\tau p_o \int_0^{\hat{\phi}-\phi} G(x)dx - \beta\left(\hat{\phi}-\phi\right) & \text{for } \hat{\phi} > \phi, \\ -\lambda\tau p_u \left(\phi-\hat{\phi}-\int_{\hat{\phi}-\phi}^0 G(x)dx\right) + \tau p_o \int_{\hat{\phi}-\phi}^0 G(x)dx + \beta\left(\hat{\phi}-\phi\right) & \text{for } \hat{\phi} < \phi. \end{cases} \end{aligned} \tag{A.24}$$

To derive the optimal response $s(\phi)$ of Sales to an order strategy q_R^{sep}, suppose a Sales type ϕ sent a signal $s(\bar{\phi})$ corresponding to some type $\bar{\phi}$. We first differentiate $U_S^R(q_R^{sep}(s(\bar{\phi})), s(\bar{\phi}) \mid \phi)$ with respect to $\bar{\phi}$:

$$\begin{aligned} \frac{\partial U_S^R(q_R^{sep}(s(\bar{\phi})), s(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}} &= b\left(1 - G\left(\bar{\phi} + G^{-1}(\alpha) - \phi\right)\right) - \gamma\left(\left(p_o+p_u\right) G\left(s(\bar{\phi})-\phi\right) - p_u\right) s'(\bar{\phi}) \\ &\begin{cases} +\tau p_u \left(1 - G\left(s(\bar{\phi})-\phi\right)\right) s'(\bar{\phi}) - \lambda\tau p_o G\left(s(\bar{\phi})-\phi\right) s'(\bar{\phi}) \\ \qquad -\beta s'(\bar{\phi}) \text{ for } s(\bar{\phi}) > \phi, \\ +\lambda\tau p_u \left(1 - G\left(s(\bar{\phi})-\phi\right)\right) s'(\bar{\phi}) - \tau p_o G\left(s(\bar{\phi})-\phi\right) s'(\bar{\phi}) \\ \qquad +\beta s'(\bar{\phi}) \text{ for } s(\bar{\phi}) < \phi. \end{cases} \end{aligned} \tag{A.25}$$

In a separating equilibrium, we require Sales to send a signal that corresponds to the true type ϕ. We hence evaluate Equation (A.25) at $\bar{\phi} = \phi$ and set it equal to zero to derive the first-order condition as a differential equation:

$$s'(\phi) = \begin{cases} \frac{b(1-\alpha)}{((1+\tau)p_u+(1+\lambda\tau)p_o)G(s(\phi)-\phi)-(1+\tau)p_u+\beta} \text{ for } s(\phi) > \phi, \\ \frac{b(1-\alpha)}{((1+\lambda\tau)p_u+(1+\tau)p_o)G(s(\phi)-\phi)-(1+\lambda\tau)p_u-\beta} \text{ for } s(\phi) < \phi. \end{cases} \tag{A.26}$$

Setting $s = s_R^{sep}$, with $(s_R^{sep})'(\phi) = 1$, Equation (A.26) holds for $\frac{b(1-\alpha)-\beta}{1+\lambda\tau} < p_o < 2\frac{b(1-\alpha)-\beta}{1+\lambda\tau} + \frac{1+\tau}{1+\lambda\tau}p_u$ and for $p_o > 2\frac{b(1-\alpha)+\beta}{1+\tau} + \frac{1+\lambda\tau}{1+\tau}p_u$. For $2\frac{b(1-\alpha)-\beta}{1+\lambda\tau} + \frac{1+\tau}{1+\lambda\tau}p_u \leq p_o \leq 2\frac{b(1-\alpha)+\beta}{1+\tau} + \frac{1+\lambda\tau}{1+\tau}p_u$, we have $s_R^{sep}(\phi) = \phi$ and Equation (A.26) is not defined, but we can verify for this case that the right hand side of Equation (A.25) evaluated at $\bar{\phi} = \phi$ is > 0 for $s(\phi) < \phi$ and < 0 for $s(\phi) > \phi$. Hence, the strategy s_R^{sep} fulfills the first-order condition.

For s_R^{sep}, the second derivative of Equation (A.24) is

$$\begin{aligned} \frac{\partial^2 U_S^R(q_R^{sep}(s_R^{sep}(\bar{\phi})), s_R^{sep}(\bar{\phi}) \mid \phi)}{\partial \bar{\phi}^2} = &- bg\left(\bar{\phi} + G^{-1}(\alpha) - \phi\right) - \gamma(p_o + p_u)g\left(\bar{\phi} + \delta_R^{sep} - \phi\right) \\ &\begin{cases} -(\tau p_u + \lambda\tau p_o)g\left(\bar{\phi} + \delta_R^{sep} - \phi\right) \text{ for } \bar{\phi} + \delta_R^{sep} > \phi, \\ -(\lambda\tau p_u + \tau p_o)g\left(\bar{\phi} + \delta_R^{sep} - \phi\right) \text{ for } \bar{\phi} + \delta_R^{sep} < \phi, \end{cases} \end{aligned} \tag{A.27}$$

which is < 0 for all $\bar{\phi} + \delta_R^{sep} \neq \phi$. That is, U_S^R is strictly concave in $\bar{\phi}$ for signals $s_R^{sep}(\bar{\phi})$ to the left and to the right of ϕ. Since we have $\beta \geq 0$ by assumption, concavity extends to $s_R^{sep}(\bar{\phi}) = \phi$. We can conclude that s_R^{sep} is a sequentially rational strategy for Sales and that $(s_R^{sep}, q_R^{sep}, \mu_R^{sep})$ is a separating equilibrium of the game with reference-dependent forecast error valuations.

□

A.1.6. Proof of Proposition 5

To derive the optimal response of Sales to an order strategy $q_c(\hat{\phi}) = \hat{\phi} - c$, we differentiate $U_S(q_c(\hat{\phi}), \hat{\phi} \mid \phi)$ with respect to $\hat{\phi}$

$$\frac{\partial U_S(q_c(\hat{\phi}), \hat{\phi} \mid \phi)}{\partial \hat{\phi}} = b\left(1 - G\left(\hat{\phi} - \phi - c\right)\right) - \gamma(p_o + p_u)G(\hat{\phi} - \phi) + \gamma p_u \begin{cases} -\beta & \text{for } \hat{\phi} > \phi, \\ +\beta & \text{for } \hat{\phi} < \phi, \end{cases} \tag{A.28}$$

and set $\frac{\partial U_S}{\partial \hat{\phi}} = 0$ to obtain the first-order condition

$$b\left(1 - G\left(\hat{\phi}_c - \phi - c\right)\right) = \gamma(p_o + p_u)G\left(\hat{\phi}_c - \phi\right) - \gamma p_u \begin{cases} +\beta \text{ for } \hat{\phi}_c > \phi, \\ -\beta \text{ for } \hat{\phi}_c < \phi, \end{cases} \tag{A.29}$$

where $\hat{\phi}_c$ is a forecast that solves the first-order condition.

To verify the existence and uniqueness of solutions to Equation (A.29), we first confirm concavity of U_S over $\mathbb{R} \setminus \{\phi\}$ by means of the second derivative

$$\frac{\partial^2 U_S(q_c(\hat{\phi}), \hat{\phi} \mid \phi)}{\partial \hat{\phi}^2} = -bg\left(\hat{\phi} - \phi - c\right) - \gamma(p_o + p_u)g(\hat{\phi} - \phi) < 0 \text{ for all } \hat{\phi} \neq \phi. \tag{A.30}$$

Given that U_S is concave on $(-\infty, \phi)$ and on (ϕ, ∞) and we have $\lim_{\hat{\phi} \to \phi^-} \frac{\partial U_S}{\partial \hat{\phi}} \geq \lim_{\hat{\phi} \to \phi^+} \frac{\partial U_S}{\partial \hat{\phi}}$ for all ϕ, there is at most one solution to Equation (A.29) for a given ϕ. If that solution exists, it maximizes the utility of Sales.

We next derive the conditions under which a solution to Equation (A.29) exists. For $p_o > -\beta/\gamma$, we have $\lim_{\hat{\phi} \to -\infty} \frac{\partial U_S}{\partial \hat{\phi}} > 0$ and $\lim_{\hat{\phi} \to \infty} \frac{\partial U_S}{\partial \hat{\phi}} < 0$. It follows that there is either a solution to Equation (A.29) or we have $\lim_{\hat{\phi} \to \phi^-} \frac{\partial U_S}{\partial \hat{\phi}} > 0$ and $\lim_{\hat{\phi} \to \phi^+} \frac{\partial U_S}{\partial \hat{\phi}} < 0$, and hence $\hat{\phi}_c = \phi$ is the optimal forecast.

For $p_o \leq -\beta/\gamma$, we have $\lim_{\hat{\phi} \to \infty} \frac{\partial U_S}{\partial \hat{\phi}} \geq 0$, i.e., there exists no solution to Equation (A.29). Because we have $p_o \geq 0$, $\beta \geq 0$ and $\gamma \geq 1$ by definition, this case can only occur if we have $p_o = \beta = 0$ and hence $\lim_{\hat{\phi} \to \infty} \frac{\partial U_S}{\partial \hat{\phi}} = 0$. We conclude that there exists a unique equilibrium response $\hat{\phi}_c$ for all Sales types ϕ if at least $p_o > 0$ or $\beta > 0$.

For ease of interpretation, we rewrite the case conditions of Equation (A.29) in terms of the incentive parameters p_o and p_u by replacing $\hat{\phi}_c > \phi$ with the requirement

$$\lim_{\hat{\phi} \to \phi^+} \frac{\partial U_S}{\partial \hat{\phi}} > 0 \quad \Leftrightarrow \quad p_o - p_u < \frac{2}{\gamma} \left[b(1 - G(-c)) - \beta \right], \tag{A.31}$$

and by replacing $\hat{\phi}_c < \phi$ with the requirement

$$\lim_{\hat{\phi} \to \phi^-} \frac{\partial U_S}{\partial \hat{\phi}} < 0 \quad \Leftrightarrow \quad p_o - p_u > \frac{2}{\gamma} \left[b(1 - G(-c)) + \beta \right]. \tag{A.32}$$

□

A.2. Phase Line Diagrams and Vector Fields

The following figures serve to visualize additional candidate equilibrium distortion strategies discussed in the proof of Theorem 1. The phase line diagrams describe the slope of u based on the properties of $w(u)$ (rightward pointing arrows: u is increasing, leftward pointing arrows: u is decreasing, filled dots: the slope of u is 0, empty dots: the slope of u is not defined). Expressions on top of the phase line denote the limits of $w(u)$ towards infinity and from both sides at points where $w(u)$ changes sign and/or is not defined. The vector fields illustrate the behavior of u depending on ϕ. They are based on an error distribution of $E \sim \mathcal{N}(0, 30)$ and a critical ratio of $\alpha = 0.5$. For each phase there are graphs of example initial conditions (IC) including an initial condition $u(0) = \delta^{sep}$ that corresponds to the Pareto-dominant differentiable separating equilibrium. Where applicable, dotted lines represent $u(\phi) = \delta^{pref}$.

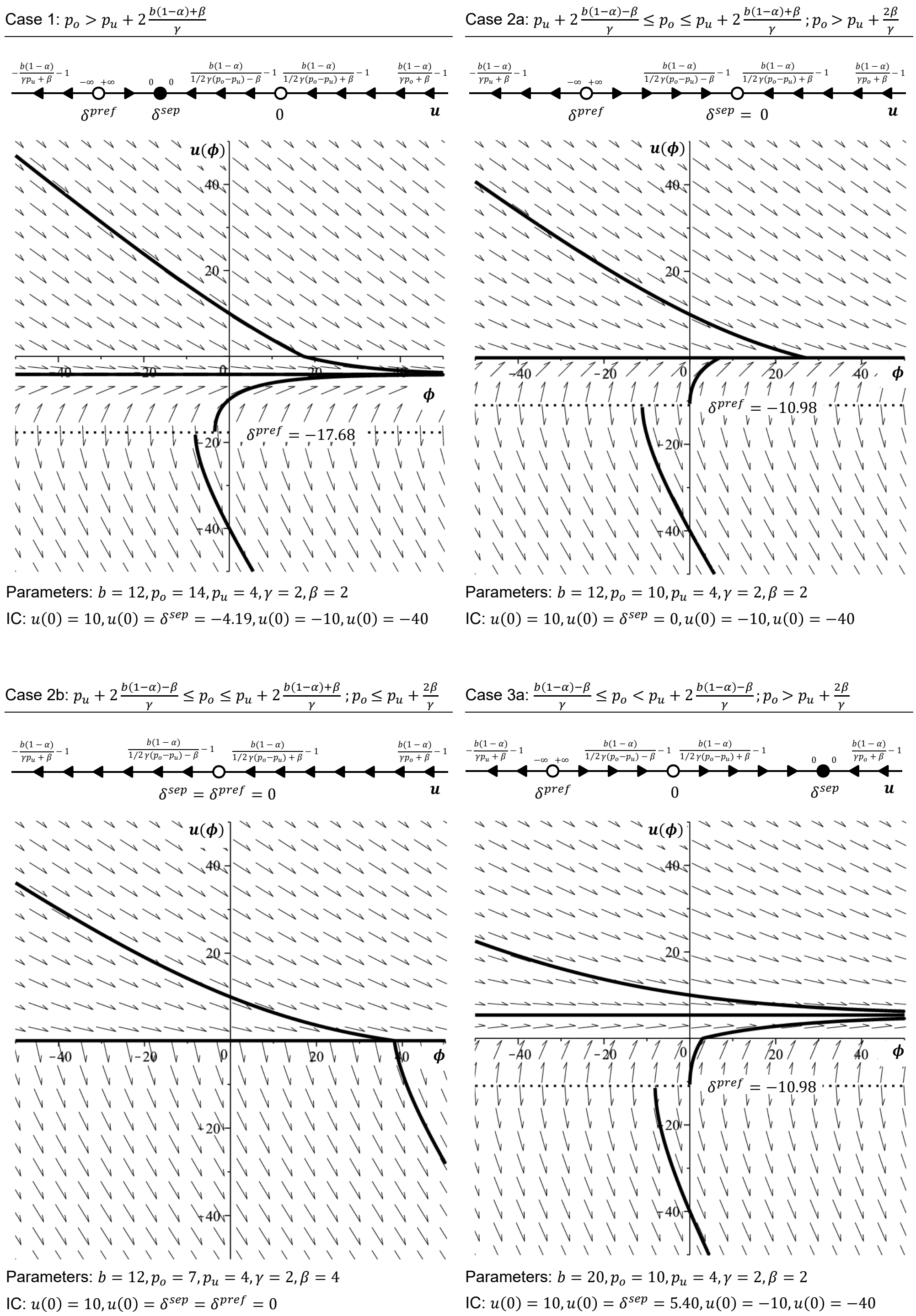


Figure A.1.: Vector field plots of candidate equilibrium distortion strategies

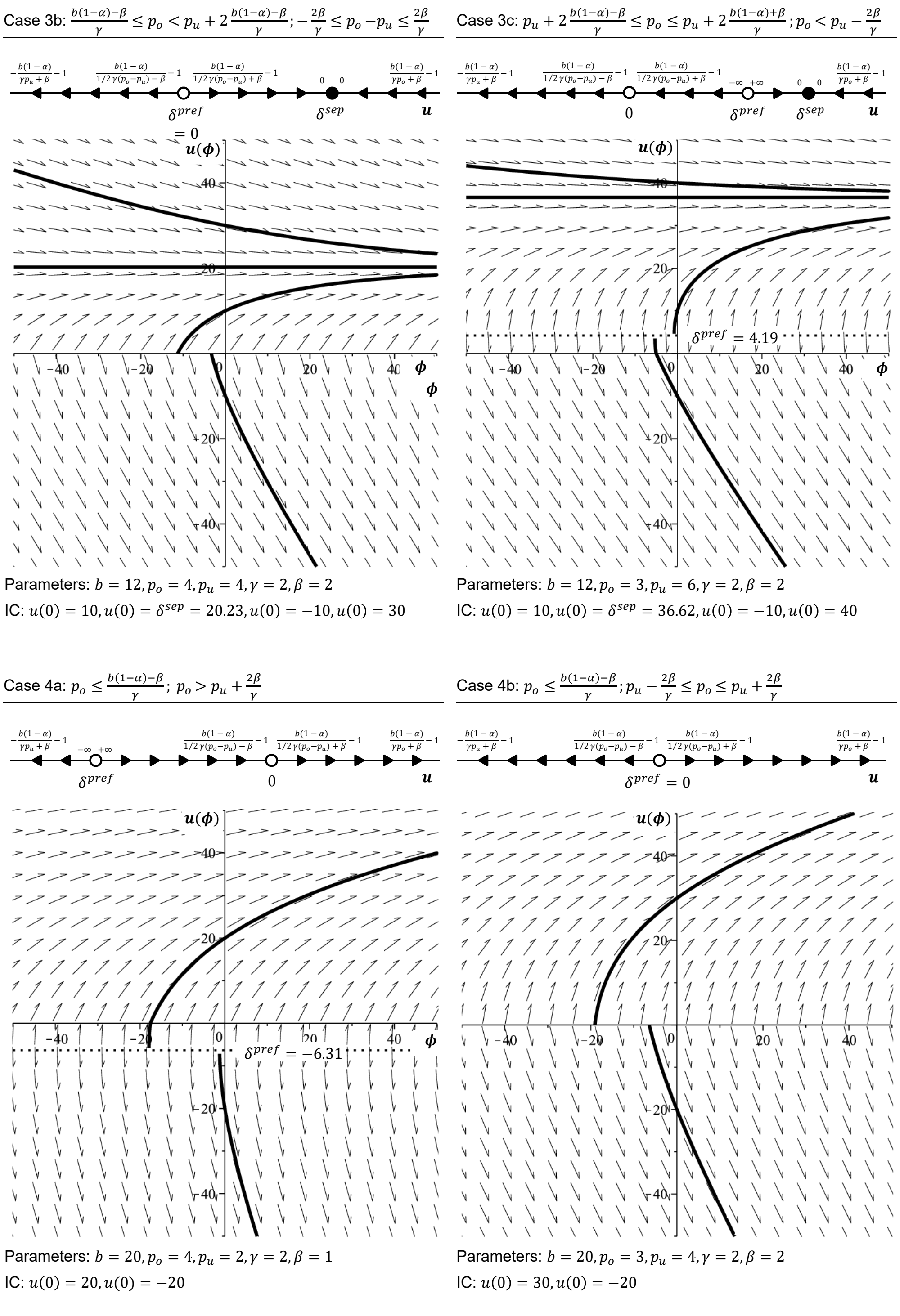


Figure A.1.: Vector field plots of candidate equilibrium distortion strategies (continued)

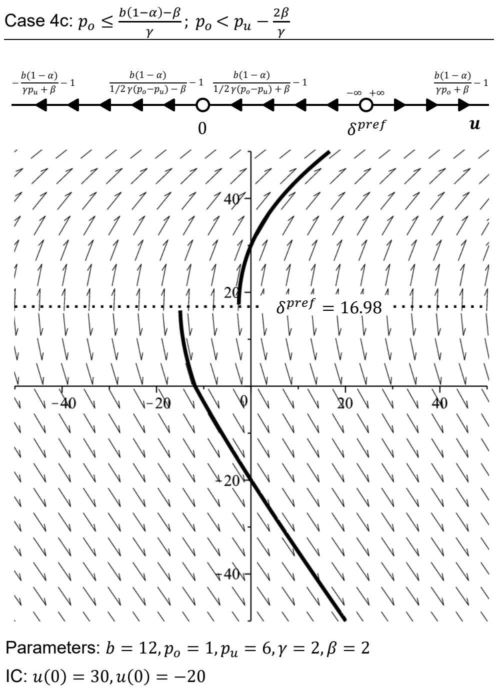


Figure A.1.: Vector field plots of candidate equilibrium distortion strategies (continued)

A.3. Details of the Main Experiment

The experiment was conducted in German. The corresponding instructions and questionnaires are available upon request from the author.

A.3.1. Instructions

At the beginning of the experiment, subject received the following instructions (exemplary for the second session (Treatments 5–8) of the main experiment), which were first read aloud by the instructor and then silently by each subject.

Overall situation.

In this experiment, you are working for a company that produces and sells a single product. Half of you will assume the role of a salesperson, the other half of you will assume the role of a production planner within this company.

- It is the *task of a salesperson* to generate a sales forecast for the product and to communicate this forecast to the production planner. The salesperson possesses information on demand for the product, that the production planner does not have.
- It is the *task of a production planner* to decide how much the company will produce based on the sales forecast.

For the entire duration of the experiment, you will assume the role either of a salesperson or of a production planner. You will be informed at the beginning of the experiment which role you have been assigned to.

Demand information.

The real demand for the product of the company is unknown at the beginning of a period. It depends on the overall market condition and additional random factors.

- The *market condition* is the expected value of demand and can take on a different value in each period. The market condition follows a normal distribution with a mean of 100 units and a standard deviation of 30 units. The realization of the market condition is known only to the salesperson.
- *Random factors* can cause the realization of demand to deviate from the market condition. The random factors follow a normal distribution with a mean of 0 units and a standard deviation of 30 units. The realization of random factors is known neither to the salesperson nor to the production planner.

A characteristic of normally distributed random variables is that on average 68 % of all realizations (i.e., of effectively drawn random numbers) are within an interval of plus/minus one standard deviation of the mean. On average 95 % of all realizations are within an interval of plus/minus two standard deviations of the mean. Figure A.2 shows an example with a market condition of 120 units. Then real demand will be within the interval of 90 ($= 120 - 1 \cdot 30$) to 150 ($= 120 + 1 \cdot 30$) with 68 % probability and within the interval of 60 ($= 120 - 2 \cdot 30$) to 180 ($= 120 + 2 \cdot 30$) with 95 % probability.

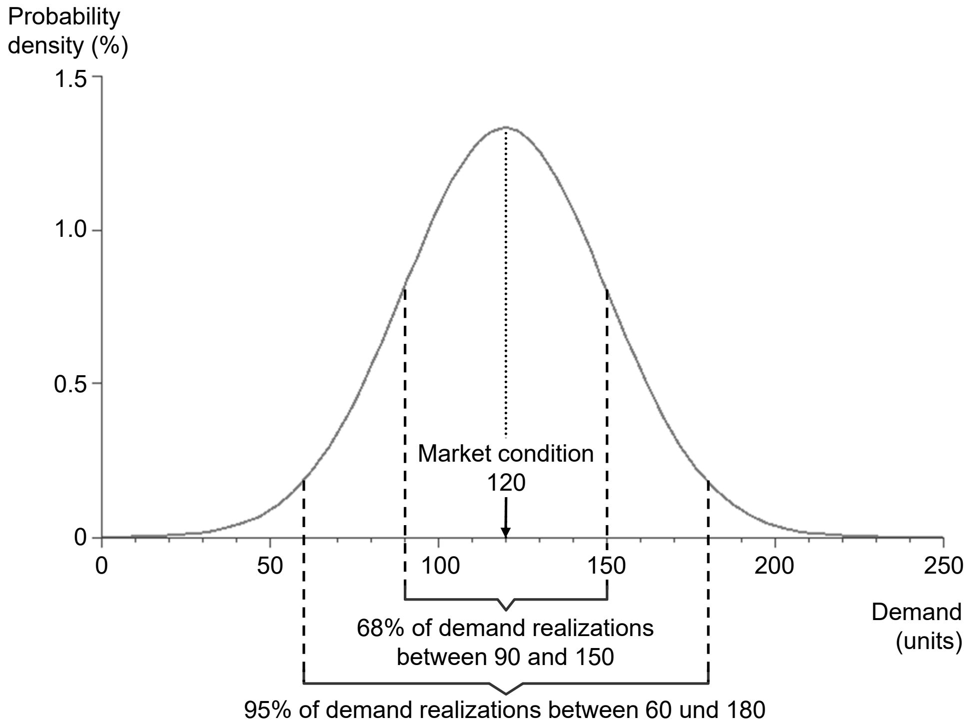


Figure A.2.: Example of demand distribution

At the end of each period (after the salesperson has communicated a forecast and the production planner has chosen a production quantity) real demand occurs. The production quantity chosen by the production planner determines how many units of the product can be sold. There are three possibilities:

- *Production quantity > demand:* If production exceeds demand, all demand can be filled. The remaining units (*overage quantity*) cannot be stored and therefore cannot be used in subsequent periods.
- *Production quantity < demand:* If production is less than demand, only part of the demand can be filled. The difference between demand and production quantity (*overage quantity*) cannot be filled in subsequent periods.
- *Production quantity = demand:* If production equals demand, all demand can be filled and there are neither overage nor underage quantities.

Information on your compensation.

In each period you have the opportunity to earn bonuses valued at virtual currency units (ECU = experimental currency unit). Over the course of the experiment you will play four different scenarios in random order (A, B, C, D).

Compensation of the production planner: The calculation of the production planner's bonus is the same across all four scenarios. The bonus of a period is composed of a fixed component, an overage component and an underage component.

- *Fixed component:* In each period, the production planner receives a fixed amount of 2,000 ECUs.
- *Overage component:* For each unit of overage quantity, 10 ECUs are deducted from the overall bonus. For example, for an overage quantity of 50 units, the overage component would be $50 \cdot (-10)$ ECUs = -500 ECUs.
- *Underage component:* For each unit of underage quantity, 10 ECUs are deducted from the overall bonus. For example, for an underage quantity of 50 units, the underage component would be $50 \cdot (-10)$ ECUs = -500 ECUs.

The magnitude of overage and underage quantities depends on the production quantity decision of the production planner and on the demand realization. The production planner does not know the demand but must estimate it based on the forecast of the salesperson. Table A.1 shows two examples of how to calculate the bonus of the production planner in case of an overage quantity (Example 1) and an underage quantity (Example 2).

Compensation of the salesperson: The calculation of the salesperson's bonus varies by scenario; the structural composition, however, is the same. The bonus of a period is composed of a fixed component, a forecast component and a sales component. In the following, the calculation will be explained with placeholders X and Y, which need to be replaced with values depending on the scenario.

- *Fixed component:* In each period, the salesperson receives a fixed amount of 1,000 ECUs.

	Example 1 Production quantity > demand	Example 2 Production quantity < demand
■ Forecast (of salesperson)	110	110
■ Production quantity (of production planner)	130	130
■ Demand realization	100	140
■ Number of units sold (minimum of production quantity and demand)	100	130
■ Overage quantity	30	--
■ Underage quantity	--	10
■ Fixed component	2,000	2,000
■ Overage component (minus 10 ECUs for each unit of overage quantity)	-300	--
■ Underage component (minus 10 ECUs for each unit of underage quantity)	--	-100
■ Overall bonus (Fixed component - overage component - underage component)	1,700	1,900

Table A.1.: Example bonus calculation of the production planner

- *Forecast component:* For each unit that the forecast deviates from the demand, some ECUs will be deducted from the salesperson's bonus. It will be distinguished between an overestimation of demand (the forecast is greater than the demand) and an underestimation of demand (the forecast is smaller than the demand):
  - *Overestimation of demand:* For each unit that the forecast is greater than the real demand, X ECUs will be deducted from the salesperson's bonus. For example, for an overestimation of 50 units, the forecast component would be $50 \cdot (-X)$ ECUs.
  - *Underestimation of demand:* For each unit that the forecast is smaller than the real demand, Y ECUs will be deducted from the salesperson's bonus. For example, for an underestimation of 50 units, the forecast component would be $50 \cdot (-Y)$ ECUs.

  The magnitude of the forecast component depends on the forecast decision of the salesperson and on the demand realization.

- *Sales component:* The salesperson receives 10 ECUs for each unit of the product sold. For example, for a sold volume of 50 units, the sales component would be $50 \cdot 10$ ECUs =

500 ECUs. The number of units sold depends on the production quantity decision of the production planner and the on the demand realization.

The placeholders X and Y of the forecast component must be replaced by scenario as shown in Table A.2.

	Overestimation of demand (placeholder X)	Underestimation of demand (placeholder Y)
▪ Scenario A	-6	-4
▪ Scenario B	-8	-2
▪ Scenario C	-10	--
▪ Scenario D	-12	-2

Table A.2.: Values of placeholders by scenario of the experiment

Both players (salesperson and production planner) will be informed, over the course of the experiment, which scenario is to be played. Table A.3 shows two examples of how to calculate the bonus of the salesperson in scenario A in case of an overestimation of demand (Example 1) and an underestimation of demand (Example 2).

In total you will play 32 periods (plus 4 test periods). At the end of the experiment, the computer will randomly choose 8 of these 32 periods. The sum of ECUs that you have earned in these 8 periods determines your overall payout. Additionally, you will have two more possibilities to increase your payout:

- You can earn an additional amount of 3,000 ECUs by correctly answering the questions of the *pre-experiment-questionnaire*. You will receive 1,000 ECUs for each block of 3 questions that you answer correctly at your first try.
- At the end of the experiment, you can increase your payout by performing two *additional tasks*.

When finished, you will receive your payout based on the sum of ECUs you earned. You will receive EUR 1 for every 1,000 ECUs.

	Example 1 Forecast > demand	Example 2 Forecast < demand
■ Forecast (of salesperson)	110	110
■ Production quantity (of production planner)	130	130
■ Demand realization	100	140
■ Overestimation of demand	10	--
■ Underestimation of demand	--	30
■ Number of units sold (minimum of production quantity and demand)	100	130
■ Fixed component	1,000	1,000
■ Forecast component: Overestimation of demand (Minus 6 ECUs for each unit of overestimation)	-60	--
■ Forecast component: Underestimation of demand (Minus 4 ECUs for each unit of underestimation)	--	-120
■ Sales component (Plus 10 ECUs for each unit sold)	1,000	1,300
■ Overall bonus (Fixed component - forecast component + sales component)	1,940	2,180

Table A.3.: Example bonus calculation of the salesperson in scenario A

Sequence of events.

At the beginning of each period, one salesperson and one production planner will be randomly matched. You will interact with a different participant of this experiment in each period and you will never get to know who you are playing with. The general sequence of actions in each period is the following (see Figure A.3):

1. *Generation of market condition:* At the beginning of each period the computer generates a random market condition (i.e., the expected value of demand for that period).

2. *Display of market condition:* The salesperson receives information about the market condition. This information will be displayed only on the salesperson's screen; the production planner will not see this information.

3. *Communication of forecast:* The salesperson sends a forecast to the production planner. The production planner sees the forecast on his screen.

4. *Decision of production quantity:* The production planner decides how much the company will produce. This decision can be, but does not have to be based on the forecast.

5. *Generation of demand:* The computer generates the real demand. Based on the demand and the production quantity the computer calculates how many units of the product can be sold.

6. *Calculation of bonuses:* Based on the forecast, the demand, the number of units sold, the overage and the underage quantity, the computer calculates the bonuses of both players.

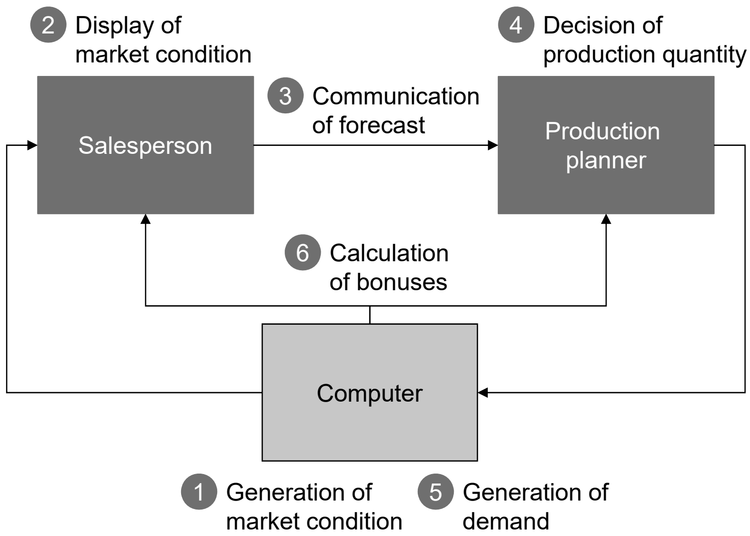


Figure A.3.: Sequence of actions in each period

A.3.2. Pre-Experiment Quiz

By way of example, the following set of questions relates to the second session (Treatments 5–8) of the main experiment. Correct answers are marked in the checkboxes.

1. Are you going to play all rounds of the experiment with the same partner?

   ☐ Yes, all rounds with the same partner

   ☒ No, each round with a different partner

   ☐ No, 10 rounds each with the same partner

2. What is the relationship between the forecast and the market condition?

   ☐ The forecast has to be equal to the market condition

   ☐ The forecast has to be higher than the market condition

   ☒ The salesperson can choose the forecast freely

3. What is the relationship between the forecast and the production quantity?

   ☐ The production quantity has to be equal to the forecast

   ☐ The production quantity has to be lower than the forecast

   ☒ The production planner can choose the production quantity freely

4. Who knows the market condition (=the expected value of demand) with certainty?

   ☒ The salesperson

   ☐ The production planner

   ☐ Both

5. Imagine the market condition is 85 units. In which interval will the real demand be with 95 % probability?

   ☐ Between 55 and 115 units

   ☒ Between 25 and 145 units

   ☐ Between 60 and 180 units

6. Evaluate the statement “The more the production planner produces, the higher the expected sales bonus of the salesperson.”

   ☒ True

   ☐ False

   ☐ Neither true nor false

7. Imagine the production planner produced 70 units. The demand is 90 units. What is the profit of the production planner?

   ☒ 1,800 ECUs

   ☐ 2,000 ECUs

   ☐ 1,600 ECUs

8. Imagine the salesperson chose a forecast of 120 units. The production planner produced 130 units. The demand is 90 units. What is the profit of the salesperson in scenario A?

   ☐ 1,780 ECUs

   ☐ 2,120 ECUs

   ☒ 1,720 ECUs

9. Imagine the salesperson chose a forecast of 60 units. The production planner produced 60 units. The demand is 90 units. What is the profit of the salesperson in scenario A?

   ☒ 1,480 ECUs

   ☐ 1,420 ECUs

   ☐ 1,780 ECUs

A.3.3. Screenshots

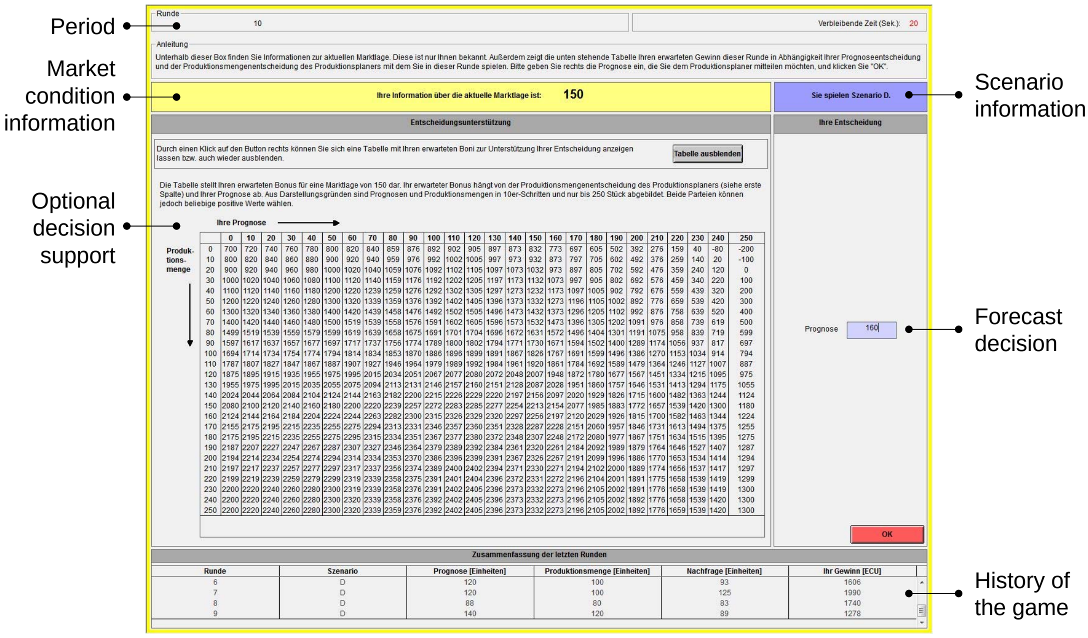


Figure A.4.: Example decision screen of Sales

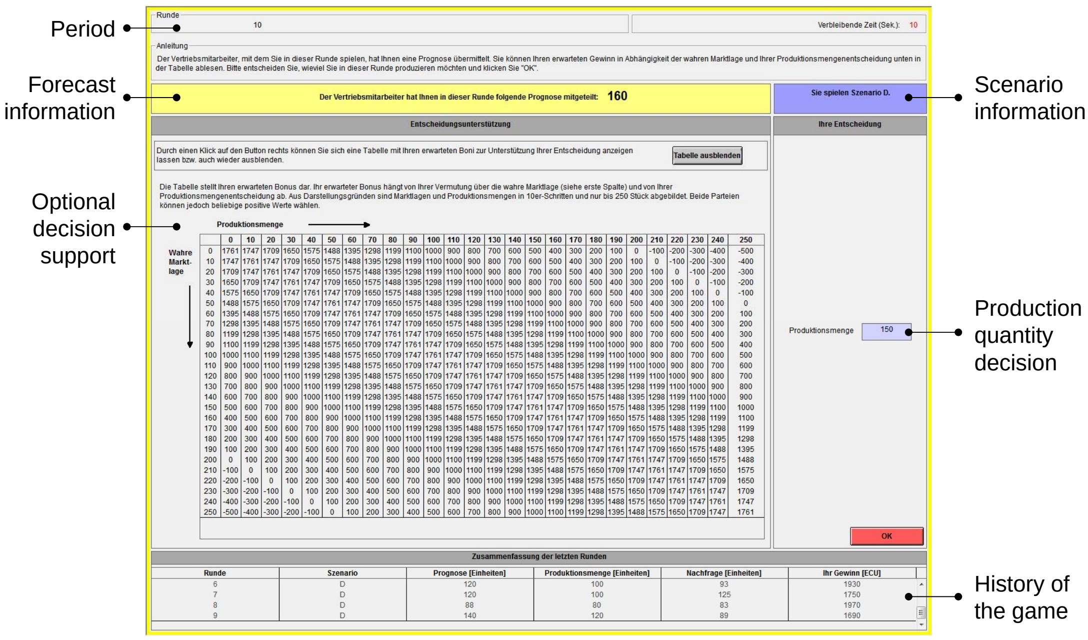


Figure A.5.: Example decision screen of Operations

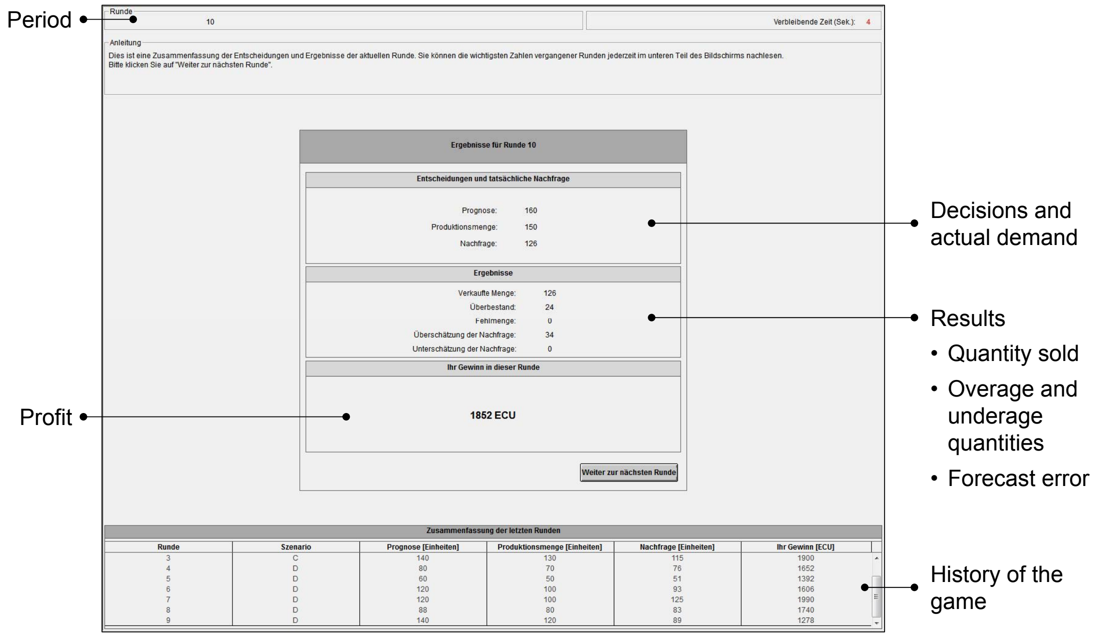


Figure A.6.: Example results screen of Sales

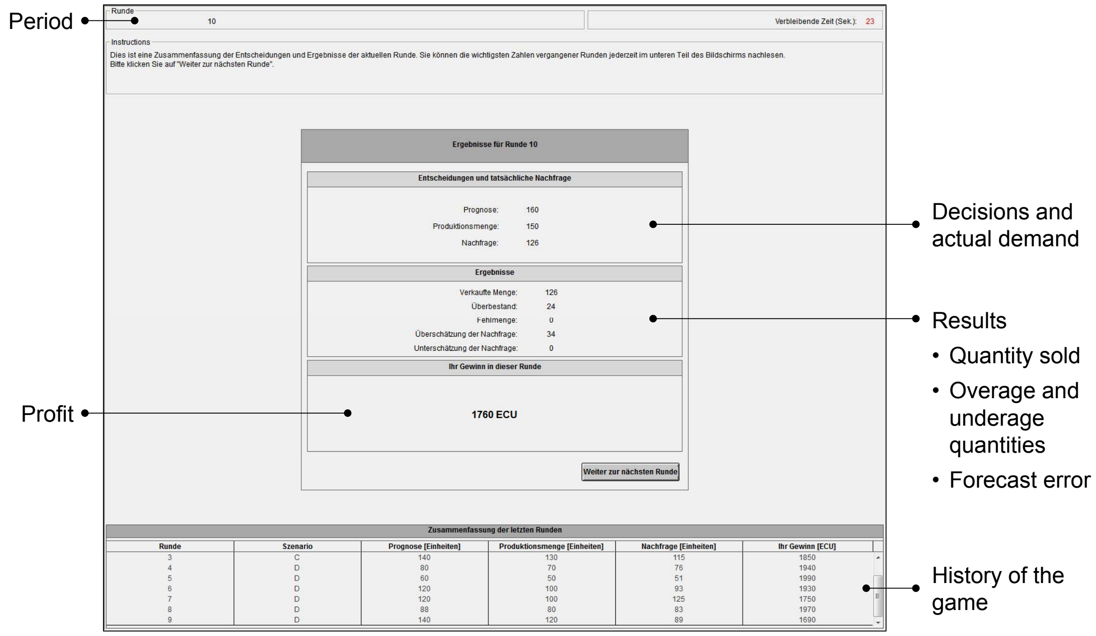


Figure A.7.: Example results screen of Operations

A.3.4. Elicitation of Risk Attitudes

After the regular duration of the experiment, subjects were asked to make ten lottery decisions. The following instructions (translated from German) were displayed on screen (see Figure A.8):

You now have the possibility to increase your payout for the experiment. Your profit in this part of the experiment depends on your own decisions only. It will be added to your profit from the forecast sharing task and will be paid out at the end of the experiment.

In the following, you have to make ten decisions. Each decision is a choice between "Option A" and "Option B". In the end, only one of the ten decisions will be used to calculate your additional profit. Before you proceed, please read the following instructions.

After having made your ten decisions, the computer will generate two random numbers. The first number determines, which one of the ten decisions is chosen for the profit calculation. The second number determines the profit for the option (A or B) that you chose in this particular decision task. Even though you will make ten decisions, only one will affect your profit. Every decision has the same probability to be chosen in the end and you do not know yet which one it will be.

Pleas take a look at the first decision task. In Option A, you will receive a profit of 1,000 ECUs with a probability of 10 % and a profit of 800 ECUs with a probability of 90 %. In Option B, you will receive a profit of 1,925 ECUs with a probability of 10 % and a profit of 50 ECUs with a probability of 90 %. The remaining decisions are similar, however, the chances of winning the higher payout are increasing. In decision number 10, the probability for the higher payout is even 100 %, i.e., you are choosing between a sure payout of 1,000 ECUs and a sure payout of 1,925 ECUs.

To summarize, you are making ten decisions of choosing between two Options A and B. You can choose Option A in some decisions and Option B in others. You can make your decisions in any order and you can change them as often as you like.

A.3.5. Post-Experiment Questionnaire

At the end of the experiment, subjects answered a questionnaire that tested their overall understanding of the game and gave them the possibility to explain their decisions. Answering

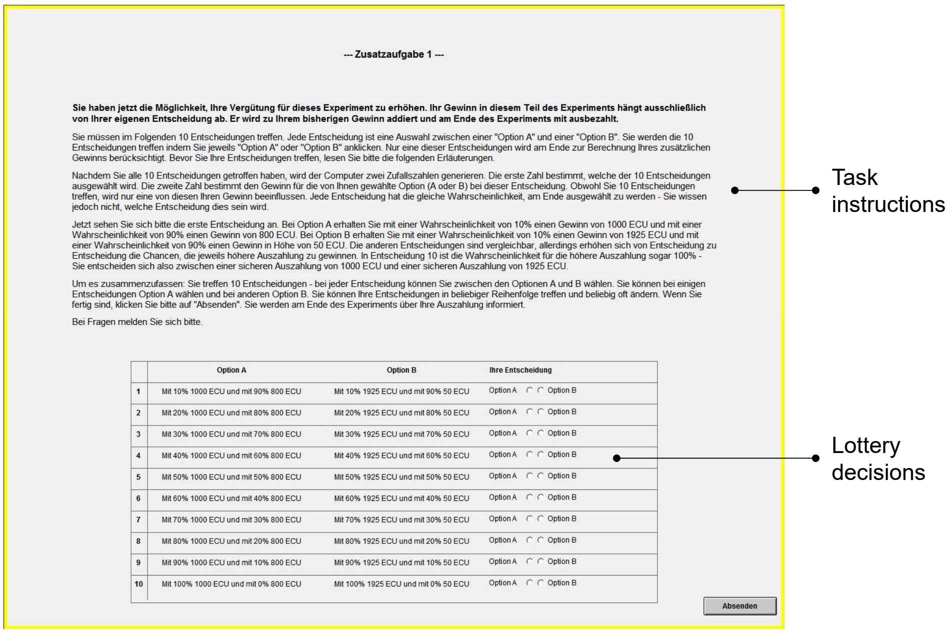


Figure A.8.: Screenshot of lottery decisions

the questionnaire was voluntary and the responses did not affect the payout. The questions differed according to the role the subject played throughout the experiment.

Questionnaire for Sales.

1. In each of the scenarios, did you report the true market condition to Operations?
   - Scenario A:
     - ☐ My forecast was usually higher than the market condition
     - ☐ My forecast usually corresponded to the market condition
     - ☐ My forecast was usually lower than the market condition
   - Scenario B:
     - ☐ My forecast was usually higher than the market condition
     - ☐ My forecast usually corresponded to the market condition
     - ☐ My forecast was usually lower than the market condition
   - Scenario C:

☐ My forecast was usually higher than the market condition

☐ My forecast usually corresponded to the market condition

☐ My forecast was usually lower than the market condition

- Scenario D:

  ☐ My forecast was usually higher than the market condition

  ☐ My forecast usually corresponded to the market condition

  ☐ My forecast was usually lower than the market condition

2. Please describe, in which situations you reported a forecast that was *higher* than the true market condition?

3. Please describe, in which situations you reported a forecast that was *lower* than the true market condition?

4. Which scenario was best for you personally?

   ☐ Scenario A

   ☐ Scenario B

   ☐ Scenario C

   ☐ Scenario D

5. Which scenario was best for the entire company?

   ☐ Scenario A

   ☐ Scenario B

   ☐ Scenario C

   ☐ Scenario D

Questionnaire for Operations.

1. Imagine, you knew the true market condition. Which order quantity would you choose if...

- ...the market condition was 75?
- ...the market condition was 100?
- ...the market condition was 125?

2. In each of the scenarios, do you think the salesperson reported the true market condition?
    - Scenario A:
        - ☐ The forecast was usually higher than the market condition
        - ☐ The forecast usually corresponded to the market condition
        - ☐ The forecast was usually lower than the market condition
    - Scenario B:
        - ☐ The forecast was usually higher than the market condition
        - ☐ The forecast usually corresponded to the market condition
        - ☐ The forecast was usually lower than the market condition
    - Scenario C:
        - ☐ The forecast was usually higher than the market condition
        - ☐ The forecast usually corresponded to the market condition
        - ☐ The forecast was usually lower than the market condition
    - Scenario D:
        - ☐ The forecast was usually higher than the market condition
        - ☐ The forecast usually corresponded to the market condition
        - ☐ The forecast was usually lower than the market condition
3. Please describe, in which situations you chose an order quantity that was *higher* than the forecast?
4. Please describe, in which situations you chose an order quantity that was *lower* than the forecast?

5. Which scenario was best for you personally?

   ☐ Scenario A

   ☐ Scenario B

   ☐ Scenario C

   ☐ Scenario D

6. Which scenario was best for the entire company?

   ☐ Scenario A

   ☐ Scenario B

   ☐ Scenario C

   ☐ Scenario D

A.3.6. Subject Pool Characteristics

Before collecting their payouts, we asked subjects to answer some statistical questions. Table A.4 summarizes the results.

	Sales	Operations	Total
Number of subjects	32	32	64
Demographics			
▪ Age (years)	25.9 (3.1)	24.2 (3.6)	25.0 (3.4)
▪ Gender			
– Female	43.8%	43.8%	43.8%
– Male	56.2%	56.2%	56.2%
Study background			
▪ Semester of studies	7.2 (4.1)	6.2 (4.9)	6.7 (4.5)
▪ Level of studies			
– Bachelor	46.9%	62.5%	54.7%
– Master	53.1%	37.5%	45.3%
▪ Course of studies			
– Business Administration	59.4%	65.6%	62.5%
– Economics	28.1%	25.0%	26.6%
– Information Systems	9.4%	6.3%	7.8%
– Other	3.1%	3.1%	3.1%
▪ Attended basic OM course			
– Yes	68.8%	68.8%	68.8%
– No	31.2%	31.2%	31.2%

Note: Standard errors are reported in parentheses.

Table A.4.: Subject pool characteristics of the main experiment

Bibliography

Abdellaoui, M., Bleichrodt, H., and Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science*, 53(10):1659–1674.

Albers, S. (1996). Optimization models for salesforce compensation. *European Journal of Operational Research*, 89(1):1–17.

Ashraf, N., Bohnet, I., and Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, 9(3):193–208.

Balasubramanian, S. and Bhardwaj, P. (2004). When not all conflict is bad: Manufacturing-marketing conflict and strategic incentive design. *Management Science*, 50(4):489–502.

Barratt, M., Choi, T. Y., and Li, M. (2011). Qualitative case studies in operations management: Trends, research outcomes, and future research implications. *Journal of Operations Management*, 29(4):329–342.

Becker-Peth, M., Katok, E., and Thonemann, U. W. (2013). Designing buyback contracts for irrational but predictable newsvendors. *Management Science*, 59(8):1800–1816.

Beer, R., Ahn, H.-S., and Leider, S. (2017). Can trustworthiness in a supply chain be signaled? *Management Science*. Published online in Articles in Advance: September 13, 2017. https://doi.org/10.1287/mnsc.2017.2817.

Bendoly, E., Croson, R., Goncalves, P., and Schultz, K. (2010). Bodies of knowledge for research in behavioral operations. *Production and Operations Management*, 19(4):434–452.

Bolton, G. E. and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity and competition. *American Economic Review*, 90(1):166–193.

Bostian, A. A., Holt, C. A., and Smith, A. M. (2008). Newsvendor "pull-to-center" effect: Adaptive learning in a laboratory experiment. *Manufacturing & Service Operations Management*, 10(4):590–608.

Bowles, S. (2009). When economic incentives backfire. *Harvard Business Review*, (March):22–24.

Cachon, G. P. (2003). Supply chain coordination with contracts. In de Kok, A. G. and Graves, S. C., editors, *Handbooks in Operations Research and Management Science: Supply Chain Management, Vol. 11*, pages 229–339. Elsevier, Amsterdam.

Cachon, G. P. and Larivière, M. A. (2001). Contracting to assure supply: How to share demand forecasts in a supply chain. *Management Science*, 47(5):629–646.

Cain, D. M., Loewenstein, G., and Moore, D. A. (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *The Journal of Legal Studies*, 34(1):1–25.

Camerer, C. and Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, 5(4):325–370.

Camerer, C. F. and Loewenstein, G. (2004). Behavioral economics: Past, present, future. In Camerer, C. F., Loewenstein, G., and Rabin, M., editors, *Advances in Behavioral Economics*, pages 3–51. Princeton University Press, Princeton, NJ.

Celikbas, M., Shanthikumar, J. G., and Swaminathan, J. M. (1999). Coordinating production quantities and demand forecasts through penalty schemes. *IIE Transactions*, 31(9):851–864.

Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6):1579–1601.

Charness, G., Gneezy, U., and Kuhn, M. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1):1–8.

Charness, G. and Levin, D. (2005). When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect. *American Economic Review*, 95(4):1300–1309.

Chen, F. (2003). Information sharing and supply chain coordination. In de Kok, A. G. and Graves, S. C., editors, *Handbooks in Operations Research and Management Science: Supply Chain Management, Vol. 11*, pages 341–421. Elsevier, Amsterdam.

Chen, F. (2005). Salesforce incentives, market information and production/inventory planning. *Management Science*, 51(1):60–75.

Chen, F., Lai, G., and Xiao, W. (2016). Provision of incentives for information acquisition: Forecast-based contracts vs. menus of linear contracts. *Management Science*, 62(7):1899–1914.

Cohen, M. A., Ho, T.-H., Ren, Z. J., and Terwiesch, C. (2003). Measuring imputed cost in the semiconductor equipment supply chain. *Management Science*, 49(12):1653–1670.

Cohen, S. A., Kulp, S., and Randall, T. (2007). Motivating supply chain behavior: The right incentives can make all the difference. *Supply Chain Management Review*, 11(4):18–24.

Cooper, D. J. and Kagel, J. H. (2016). Other regarding preferences: A selective survey of experimental results. In Kagel, J. H. and Roth, A. E., editors, *Handbook of Experimental Economics, Vol. 2*, pages 217–289. Princeton University Press, Princeton, NJ.

Crawford, V. P. (2003). Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, 93(1):133–149.

Croson, R., Schultz, K., Siemsen, E., and Yeo, M. L. (2013). Behavioral operations: The state of the field. *Journal of Operations Management*, 31(1-2):1–5.

Cui, R., Allon, G., Bassamboo, A., and Van Mieghem, J. A. (2015). Information sharing in supply chains: An empirical and theoretical valuation. *Management Science*, 61(11):2803–2824.

Cui, T. H., Raju, J. S., and Zhang, Z. J. (2007). Fairness and channel coordination. *Management Science*, 53(8):1303–1314.

Dal Bó, P. (2005). Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *The American Economic Review*, 95(5):1591–1604.

Dershem, B. (2007). Financial incentives improve the supply chain: Heineken's journey. *Journal of Business Forecasting*, 26(1):32–37.

Ebel, T., George, K., Shah, K., and Ungerman, D. (2013). Building new strengths in the healthcare supply chain. Accessed October 10, 2017, https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/strengthening-health-cares-supply-chain-a-five-step-plan.

Ebrahim-Khanjari, N., Hopp, W., and Iravani, S. M. R. (2012). Trust and information sharing in supply chains. *Production and Operations Management*, 21(3):444–464.

Eckel, C. C. and Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, 68(1):1–17.

Eichberger, J., Kelsey, D., and Schipper, B. C. (2008). Granny versus game theorist: Ambiguity in experimental games. *Theory and Decision*, 64(2-3):333–362.

Eliashberg, J. and Steinberg, R. (1993). Marketing-production joint decision-making. In Eliashberg, J. and Lilien, G. L., editors, *Handbooks in Operations Research and Management Science: Marketing, Vol. 5*, pages 827–880. Elsevier, Amsterdam.

Ellinger, A. E. (2000). Improving marketing/logistics cross-functional collaboration in the supply chain. *Industrial Marketing Management*, 29(1):85–96.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669.

Erat, S. and Gneezy, U. (2012). White lies. *Management Science*, 58(4):723–733.

Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2):293–315.

Fehr, E. and Falk, A. (2002). Psychological foundations of incentives. *European Economic Review*, 46(4-5):687–724.

Fehr, E. and Fischbacher, U. (2002). Why social preferences matter- The impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal*, 112:1–33.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868.

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

Fisher, M. (2007). Strengthening the empirical base of operations management. *Manufacturing & Service Operations Management*, 9(4):368–382.

Fudenberg, D. and Tirole, J. (1991). *Game theory.* MIT Press, Cambridge, MA.

Gaur, V., Giloni, A., and Seshadri, S. (2005). Information sharing in a supply chain under ARMA demand. *Management Science*, 51(6):961–969.

Gibbons, R. (2005). Incentives between firms (and within). *Management Science*, 51(1):2–17.

Gino, F. and Pisano, G. (2008). Toward a theory of behavioral operations. *Manufacturing & Service Operations Management*, 10(4):676–691.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1):384–394.

Gneezy, U., Rockenbach, B., and Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior and Organization*, 93:293–300.

Gneezy, U. and Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics*, 115(3):791–810.

Gonik, J. (1978). Tie salesmen's bonuses to their forecasts. *Harvard Business Review*, 56(3):116–123.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125.

Gümüş, M. (2014). With or without forecast sharing: Competition and credibility under information asymmetry. *Production and Operations Management*, 23(10):1732–1747.

Heath, C., Larrick, R. P., and Wu, G. (1999). Goals as reference points. *Cognitive Psychology*, 38(1):79–109.

Ho, T.-H., Lim, N., and Cui, T. H. (2010). Reference dependence in multilocation newsvendor models: A structural analysis. *Management Science*, 56(11):1891–1910.

Ho, T.-H., Su, X., and Wu, Y. (2014). Distributional and peer-induced fairness in supply chain contract design. *Production and Operations Management*, 23(2):161–175.

Ho, T.-H. and Zhang, J. (2008). Designing pricing contracts for boundedly rational customers: Does the framing of the fixed fee matter? *Management Science*, 54(4):686–700.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Hurkens, S. and Kartik, N. (2009). Would I lie to you? On social preferences and lying aversion. *Experimental Economics*, 12(2):180–192.

Hyndman, K., Kraiselburd, S., and Watson, N. (2013). Aligning capacity decisions in supply chains when demand forecasts are private information: Theory and experiment. *Manufacturing & Service Operations Management*, 15(1):102–117.

Inderfurth, K., Sadrieh, A., and Voigt, G. (2013). The impact of information sharing on supply chain performance under asymmetric information. *Production and Operations Management*, 22(2):410–425.

Johnson, R. B. and Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7):14–26.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, 98(6):1325–1348.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.

Karlsson, P.-M., Lal, S., and Rexhausen, D. (2017). Hidden sources of better supply-chain performance. *McKinsey Quarterly*, (August):1–3. Accessed October 5, 2017. https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/hidden-sources-of-better-supply-chain-performance.

Kartik, N., Ottaviani, M., and Squintani, F. (2007). Credulity, lies, and costly talk. *Journal of Economic Theory*, 134(1):93–116.

Katok, E. (2011). Using laboratory experiments to build better operations management models. *Foundations and Trends in Technology, Information and Operations Management*, 5(1):1–86.

Katok, E., Olsen, T., and Pavlov, V. (2014). Wholesale pricing under mild and privately known concerns for fairness. *Production and Operations Management*, 23(2):285–302.

Katok, E. and Pavlov, V. (2013). Fairness in supply chain contracts: A laboratory study. *Journal of Operations Management*, 31(3):129–137.

Katok, E. and Wu, D. Y. (2009). Contracting in supply chains: A laboratory investigation. *Management Science*, 55(12):1953–1968.

Kelsey, D. and le Roux, S. (2015). An experimental study on the effect of ambiguity in a coordination game. *Theory and Decision*, 79(4):667–688.

Kerr, S. (1975). On the folly of rewarding A, while hoping for B. *The Academy of Management Journal*, 18(4):769–783.

Kremer, M., Minner, S., and Van Wassenhove, L. N. (2014). On the preference to avoid ex post inventory errors. *Production and Operations Management*, 23(5):773–787.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Lawrence, M. and O'Connor, M. (2005). Judgmental forecasting in the presence of loss functions. *International Journal of Forecasting*, 21(1):3–14.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, 90(5):1346–1361.

Li, L. and Zhang, H. (2008). Confidentiality and information sharing in supply chain coordination. *Management Science*, 54(8):1467–1481.

Lindstrom, M. J. and Bates, D. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46(3):673–687.

Loch, C. H. and Wu, Y. (2008). Social preferences and supply chain performance: An experimental study. *Management Science*, 54(11):1835–1849.

López-Pérez, R. and Spiegelman, E. (2013). Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics*, 16(3):233–247.

Lundquist, T., Ellingsen, T., Gribbe, E., and Johannesson, M. (2009). The aversion to lying. *Journal of Economic Behavior & Organization*, 70(1):81–92.

Mailath, G. J. and von Thadden, E. L. (2013). Incentive compatibility and differentiability: New results and classic applications. *Journal of Economic Theory*, 148(5):1841–1861.

Malhotra, D. and Murnighan, J. K. (2002). The effects of contracts on interpersonal trust. *Administrative Science Quarterly*, 47(3):534–559.

Mantrala, L. K. and Raman, K. (1990). Analysis of a sales force incentive plan for accurate sales forecasting and performance. *International Journal of Research in Marketing*, 7(2):189–202.

McKenzie, J. (2011). Mean absolute percentage error and bias in economic forecasting. *Economics Letters*, 113(3):259–262.

Meredith, J. (1998). Building operations management theory through case and field research. *Journal of Operations Management*, 16(4):441–454.

Mishra, B. K., Raghunathan, S., and Yue, X. (2007). Information sharing in supply chains: Incentives for information distortion. *IIE Transactions*, 39(9):863–877.

Nowak, M. A. and Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437:1291–1298.

Ockenfels, A., Sliwka, D., and Werner, P. (2015). Bonus payments and reference point violations. *Management Science*, 61(7):1496–1513.

Oliva, R. and Watson, N. (2009). Managing functional biases in organizational forecasts: A case study of consensus forecasting in supply chain planning. *Production and Operations Management*, 18(2):138–151.

Özer, Ö., Subramanian, U., and Wang, Y. (2017). Information sharing, advice provision or delegation: What leads to higher trust and trustworthiness? *Management Science*. Published online in Articles in Advance: January 23, 2017. https://doi.org/10.1287/mnsc.2016.2617.

Özer, Ö. and Uncu, O. (2013). Competing on time: An integrated framework to optimize dynamic time-to-market and production decisions. *Production and Operations Management*, 22(3):473–488.

Özer, Ö. and Wei, W. (2006). Strategic commitments for an optimal capacity decision under asymmetric forecast information. *Management Science*, 52(8):1238–1257.

Özer, Ö. and Zheng, Y. (2017). Establishing trust and trustworthiness for supply chain information sharing. In Ha, A. Y. and Tang, C. S., editors, *Handbook of Information Exchange in Supply Chain Management, Vol. 5*, pages 287–312. Springer International Publishing, Cham, Switzerland.

Özer, Ö., Zheng, Y., and Chen, K.-Y. (2011). Trust in forecast information sharing. *Management Science*, 57(6):1111–1137.

Özer, Ö., Zheng, Y., and Ren, Y. (2014). Trust, trustworthiness, and information sharing in supply chains bridging China and the United States. *Management Science*, 60(10):2435–2460.

Pinheiro, J. and Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. Springer, New York.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2017). *nlme: Linear and nonlinear mixed effects models. R package version 3.1-131*. Accessed March 12, 2017, https://cran.r-project.org/package=nlme.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, 37(1):7–63.

Pulford, B. D. and Colman, A. M. (2007). Ambiguous games: Evidence for strategic ambiguity aversion. *The Quarterly Journal of Experimental Psychology*, 60(8):1083–1100.

PwC (2015). Reimagining operations - PwC's 2015 global operations survey. Accessed October 5, 2017. http://operationssurvey.pwc.com/PwC-2015-Global-Operations-Survey.pdf.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature*, 36(1):11–46.

Reese, S. (2001). The human aspects of collaborative forecasting. *Journal of Business Forecasting Methods & Systems*, 19(4):3–9.

Ren, Z. J., Cohen, M. A., Ho, T.-H., and Terwiesch, C. (2010). Information sharing in a long-term supply chain relationship: The role of customer review strategy. *Operations Research*, 58(1):81–93.

Scheele, L. M., Thonemann, U. W., and Slikker, M. (2017). Designing Incentive Systems for Truthful Forecast Information Sharing Within a Firm. *Management Science*. Published online in Articles in Advance: August 30, 2017. https://doi.org/10.1287/mnsc.2017.2805.

Schweitzer, M. E. and Cachon, G. P. (2000). Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science*, 46(3):404–420.

Shapiro, B. P. (1977). Can marketing and manufacturing coexist? *Harvard Business Review*, 55(5):104–114.

Shin, H. and Tunca, T. I. (2010). Do firms invest in forecasting efficiently? The effect of competition on demand forecast investments and supply chain coordination. *Operations Research*, 58(6):1592–1610.

Sobel, J. (2009). Signaling games. In Meyers, R. A., editor, *Encyclopedia of Complexity and Systems Science*, pages 8125–8139. Springer, New York.

Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Springer, Vienna, Austria.

Stamland, T. (1999). Partially informative signaling. *Journal of Economic Theory*, 89(1):148–161.

Terwiesch, C., Ren, Z. J., Ho, T.-H., and Cohen, M. A. (2005). An empirical analysis of forecast sharing in the semiconductor equipment supply chain. *Management Science*, 51(2):208–220.

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science*, 4(3):199–214.

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, 12(3):183–206.

Turner, R., Lasserre, C., and Beauchet, P. (2007). Innovation in field force bonuses: Enhancing motivation through a structured process-based approach. *Journal of Medical Marketing*, 7(2):126–135.

Vanberg, C. (2008). Why do people keep their promises? An experimental test of two explanations. *Econometrica*, 76(6):1467–1480.

Voigt, G. and Inderfurth, K. (2012). Supply chain coordination with information sharing in the presence of trust and trustworthiness. *IIE Transactions*, 44(8):637–654.

Wang, L. and Murnighan, J. K. (2017). How much does honesty cost? Small bonuses can motivate ethical behavior. *Management Science*, 63(9):2903–2914.

Wang, Y. (2006). Joint pricing-production decisions in supply chains of complementary products with uncertain demand. *Operations Research*, 54(6):1110–1127.

Yin, R. (2009). *Case study research: Design and methods*. Sage Publications, Thousand Oaks, CA, 4th edition.

Zhang, Y., Donohue, K., and Cui, T. H. (2016). Contract preferences and performance for the loss-averse supplier: Buyback vs. revenue sharing. *Management Science*, 62(6):1734–1754.