Omega 000 (2018) 1–11

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

Coordinating judgmental forecasting: Coping with intentional biases

Clint L.P. Pennings*, Jan van Dalen, Laurens Rook

Rotterdam School of Management, Erasmus University, T9-08, Burgemeester Oudlaan 50, Rotterdam, 3062 PA, The Netherlands

ARTICLE INFO

Article history:
Received 30 November 2017
Accepted 8 August 2018
Available online xxx

Keywords:
Judgmental
Forecasting
Biases
Departmental roles and incentives
Negotiation
Intentional biases
Weighting schemes

ABSTRACT

Human judgment, an almost inextricable ingredient in demand forecasting, introduces many unintentional and intentional biases to the forecasting and operations planning process. In the present research, we isolate intentional biases from this process and relate them to heterogeneous departmental roles and incentives. Through a laboratory experiment, which simulates forecasting operations planning in an interdepartmental decision-making context, we examine the effects of departmental roles, incentives and various weighting schemes on forecasting behavior and performance. We find that departmental roles, even without role-specific incentives, entail intentional biases of 8% of the forecast, and that role-specific incentives increase these biases to 14%. We further test the claim that accuracy-weighted schemes can remove biases in forecasting, and conclude that they halve, but don't fully remove them. Finally, a weighting scheme that explicitly corrects biased inputs shows great promise in reducing intentional and unintentional biases. In our experiment, this scheme reduces biases by 35%. This shows the importance of disentangling intentional and unintentional biases for more effective forecasting adjustments. Our insights have substantial ramifications for the design of the forecasting operations planning process in dynamic business environments determined by high levels of role- and incentive-based heterogeneity.

1. Introduction

Judgmental forecasting is a vital component in demand forecasting, with considerable implications for the functioning of a wide range of corporate supply chain-related procedures and activities, as well as a company's overall financial performance [28,29,77,78]. In organizational settings, judgmental forecasting is often used to capitalize on the valuable tacit or domain-specific knowledge, which is present within the corporation, but not always grounded in theoretical frameworks and/or statistical approaches [29,37,48,68]. People in companies tend to lack the knowledge and expertise in forecasting to validly support decision-making [37], even to the extent that the level of knowledge and forecast accuracy seem to have decreased over time [54]. In this respect, Davis and Mentzer [21] observe a gap between forecasting theory and practice in companies, and consider this a significant and urgent problem. Hence, researchers call for analyses that focus on how the forecasting process is managed and organized [2,20]. In response, we examine in the present research the origin and effect of biases due to judgmental forecasting in companies. Such biases can be unintentional (i.e., due to accidental errors in decision making), but they may also be intentional – that is, deliberately introduced in the process by the forecaster to suit a particular, perhaps strategically motivated, objective.

In the industry, the crucial but complex issue of forecasting operations planning is often conducted in a specific organizational unit dedicated to Sales & Operations Planning (S&OP) [74]. The S&OP process is an interdepartmental decision-making process, in which representatives from various departments, such as sales, operations, and finance, jointly generate a corporate forecast [44,48,61,62]. The demand planners in the S&OP unit typically deal with large volumes of data from various sources, offered in multiple formats. Not surprisingly, scholars in recent years have connected the S&OP process with data mining and big data analytics [15,53,73,83]. Notwithstanding, demand planners traditionally modify the information from a statistically generated, initial, forecast under the influence of judgmental information that comes up during a physical meeting. They eventually agree on a final forecast that is a mixture of statistics and human judgment [28,68–70]. Growing evidence suggests that manual judgmental adjustments to statistical forecasts may improve accuracy [26,28,39,45]. At the same time, however, human judgment also introduces all sorts of biases to forecasting operations planning. This is particularly true for the S&OP process, in which tensions between departmental roles and incentives may heavily impact the forecast.

Many role-dependent conflicts arise between demand planners during S&OP meetings, and especially notorious are the cross-

* Corresponding author.
E-mail addresses: cpennings@rsm.nl (C.L.P. Pennings), jdalen@rsm.nl (J.v. Dalen), l.rook@tudelft.nl (L. Rook).

https://doi.org/10.1016/j.omega.2018.08.007
0305-0483/© 2018 Elsevier Ltd. All rights reserved.

functional differences in approach and objectives between representatives of sales and operations departments [26,72,80,81]. The departmental roles of sales and operations within the organization commonly trigger opposing interests: the sales department prioritizes customer service and sales development; the operations department focuses on efficiency and costs [59,69,70]. As a consequence, departments do not necessarily share the same goal of minimizing forecast errors. Sales representatives may prefer to maintain high levels of inventory so as to be readily capable to respond to yet unforeseen sales peaks; operations personnel may feel more confident with lower inventory levels justified by historical data [58,62,70]. These opposing interests may lead to differentially contextualized inflated or deflated forecasts [42,64]. Departmental roles and incentives, therefore, clearly provide context for intentional as well as unintentional biases in demand forecasting.

However, the interplay of unintentional and intentional decision biases has been largely ignored in the literature on forecasting [61,62]. Admittedly, the general notion of decision bias has received much attention in this literature, as human judgment is rife with inherent biases [45]. Fildes et al. [28] summarize the extant knowledge about the ways in which excessive trust in one's own judgment (the illusion of control) and over-optimism may distort judgmental forecasting. Documented elsewhere are effects of the escalation of commitment [50], anchoring and adjustment [71], and other unintentional decision biases [33]. Nevertheless, matters get more complicated once we factor in intentional biases stemming from deliberate – role and incentive-based – human adjustment. Deliberate information processing tends to be slower and more precise than unintentional information processing; it also is strategic in nature [13,31]. Despite being different, intentional behavior will always contain a fraction of unintentionality [24,40]. Thus far, composite methodologies and weighting schemes have been proposed to cope with cross-functional demand forecasting within organizational setting [70]. Organizations such as Leitax, a manufacturer of consumer electronics [61,62], Norges Bank, and manufacturers of fast-moving consumer goods [8,66] seem to suggest that weighting schemes may remove (some of) the human bias due to departmental role and incentive. Still, the precise effects of departmental role, incentives, and the application of weighting schemes on the coordination of the cross-functional process of demand forecasting – as well as their impact on actual forecasting behavior – have not been examined.

In this paper, we study intentional biases in the context of different departmental roles and incentives in demand forecasting. Through a laboratory experiment, which simulates forecasting and production quantity decisions in an interdepartmental decision-making context, we disentangle intentional from unintentional biases and examine the effects of roles, incentives, and various weighting schemes on behavior and performance. Our work shows the importance of separately considering the two sources of biases for research. Also, our findings are practically relevant to the design of the demand forecasting process in terms of coordination mechanisms and incentives.

The remainder of this paper is organized as follows. Section 2 outlines the relevant literature on intentional biases in judgmental forecasting and on weighting schemes for formulating consensus forecasts, and states our hypotheses. Section 3 specifies our experiment and our methods to examine participants' behavior. Section 4 lists the results and their implications, while Section 5 concludes and gives suggestions for future research.

2. Theoretical background

Human judgment adds intentional and unintentional biases to the forecasting and operations planning process. Much knowledge nowadays exists on the often negative relationships between decision biases and forecasting accuracy, and various weighting schemes have been proposed as mechanisms to counter their detrimental effects. However, scholars have primarily devoted their research attention to the study of forecasting behavior flawed by unintentional forecasting biases. Systematic errors in forecasting caused by intentional biases – i.e., behaviors that are strategically and tactically motivated [13,31] – have been largely ignored. Weighting schemes supposedly also remove intentional biases, but their influence on forecasting behavior remains disregarded. In this section, we review the extant literature on forecasting and operations planning, and lay the claim that intentional biases in such settings should be isolated from unintentional biases. We formulate hypotheses and determine the objectives of our laboratory experiments to examine how the design of the forecasting process affects behavior and performance.

2.1. Intentional forecast biases

Forecast processes in organizations involve multiple stakeholders with distinct functional backgrounds, preferences, expectations, and behavioral repertoires. These factors do not necessarily contribute to the quality and accuracy of the forecasts, and possibly impair forecasts due to the negative effects of functional heterogeneity [12]. Also, there often is a lack of management of forecasting processes at companies, with a blurred distinction between forecasts, plans, and goals [56]. Scholars have recognized that contextual factors on the organizational level can cause the forecaster to maintain objectives other than accuracy when making demand forecasts: forecasts can be subject to managerial pressure [79], they can be intentionally flawed [48], or even be the “result of deliberate and rational decision making behavior on the part of the forecasters” [47, p.3]. Documented examples of intentional biases include forecasters who inflate forecasts to ensure that suppliers give them priority [79] or to increase the publicity of the forecast [5]. Wright and Rowe [88, p. 12] conclude that “[w]e must always remember that forecasts are rarely, in themselves, disinterested and innocent products of the group process in which they are produced and this reality should cause us to reconsider the way in which we evaluate forecasts.”

This particularly holds for the S&OP process – a collaborative, essentially cross-functional decision-making process, in which representatives from various departments in the organization, ranging from sales to operations, and finance, jointly generate a corporate forecast and a joint planning in a physical meeting [44,48,61,62]. As stated in the introduction, the differences in approach and operational execution between these functional roles are well-documented, and identified as a starting point of many organizational conflicts regarding demand forecasting [58,59,62,70,72]. The call to address these cross-functional conflicts has been made both in the marketing literature [80,81] and in the growing work on collaborative forecasting [26].

Oliva and Watson [61,62] illustrate the importance of role- and incentive-based heterogeneity for the forecasting process by describing the overhaul of the forecasting process at Leitax, a manufacturer of consumer electronics. Prior to the change, the forecasting process was fragmented over the various (sales, operations, finance) departments. Sales shared their generated – i.e., judgmental – forecasts informally with representatives from the operations and finance departments. The people at the operations department required forecasts for purchasing and production decisions. Those at finance needed forecasts for their financial planning and management. Not convinced of the adequacy of each other's forecasts, each functional department generated their own forecasts, resulting in major financial losses for the corporation. Centralizing and redesigning the S&OP process was successful, and had a large impact on the Leitax operations. The forecast accuracy increased by

30 percentage points from 58% to 88%, which entailed millions of savings in inventory.

Departmental roles and incentives have been identified as sources of unintentional and intentional forecast biases. Yaniv [89] fully ascribes biases to incentives and concludes that forecasting behavior differs substantially between departments with regard to financial incentives. By contrast, Kuo and Liang [42] highlight the importance of roles, and conclude that departmental roles affect forecasting behavior, even when forecasters receive exactly the same information and have no role-specific incentives or interests. Also participants in a study by Önkal et al. [64] displayed a strong commitment to their cross-functional roles. The authors found significant effects of organizational roles on the forecast, even without incentives. This clearly shows that role- and incentive-based heterogeneity may exert a substantial, intentional as well as unintentional, influence on forecasting operations planning.

2.2. Weighting schemes for combining forecasts

Demand forecasting is a genuinely difficult exercise. The decision maker evaluates and makes a forecast based on prior knowledge and information about demand and supply under high uncertainty about the future, about outliers, about the quantity of forecasts to be considered, and about trends or disruptive events [28,29]. Not surprisingly, many methodologies have been proposed for successful forecasting, ranging from general reminders on vigilance and conservatism in forecasting [4], and detailed guidelines in handbook format on how to keep forecasting simple [3], to structured overviews of various forecasting approaches. For instance, instead of having groups generate a single forecast, it has been suggested to combine the separate individual forecasts of group members into a final forecast using a weighting scheme [70]. Also, advanced statistical software packages have been developed that allow for the combination of standard approaches such as exponential smoothing with other, naive, forecast methods [28,30]. In recent years, business intelligence and analytics have been recommended for solving a wide range of forecasting problems characterized by large volumes and varieties of data [15,34,53,82], also in the setting of supply chain management [73,83]. However, the methodological rigor of big data applications to forecasting has been questioned. A major challenge to the use of big data analytics in forecasting seems to be the human factor – i.e., the biased or even erroneous processing and interpretation of data by the people involved [41,49]. It stands to reason that human decision biases also interfere with the use of predictive analytics in demand forecasting [83].

Especially, weighting schemes that combine forecasts obtained from different methods have been studied extensively. The dominant finding is that the combination of forecasts obtained from various methods can reduce the forecast error, by being more robust to particular assumptions and wrong inferences [7,22]. Indeed, empirical evidence exists that a combination of separate forecasts often substantially improves forecast accuracy [14,16,22]. Even a simple average of forecasts tends to outperform separate forecasts [27]. A simple average is generally more robust than a weighted average [65]. Moreover, adding additional forecasts as inputs has been shown to further improve accuracy [52]. These benefits of combining forecasts have received wide support [19,43,67,76,85]. They, interestingly, appear to hold also when inputs are provided by different judgmental forecasters rather than forecasting methods [6,18,51,57,63].

In addition to the simple average, various weighting schemes have been proposed to combine separate forecasts. The popular accuracy-weighted scheme, for instance, derives from the variance-covariance method. This method incorporates the accuracy of the individual forecasts, reflected by the variance of individual forecast errors, as well as the dependence between forecasts, reflected by the covariance between individual forecast errors [87]. The accuracy weights have been calculated by means of linear regression [36], principal component regression [14], or Bayesian shrinkage [1,23,55,84]. The accuracy-weighted scheme, seen at Norges Bank [8] and manufacturers [66], ignores the covariance between forecast errors to increase forecast accuracy, because of the sensitivity of the covariance to the sample cross-correlations. This results in highly unstable estimates of the weights [17,52,60,86].

The accuracy-weighted scheme was also used at Leitax, the case company studied by Oliva and Watson [61], 62], and which was introduced and described above. Importantly, the authors suggested that weighting schemes negate the effects of roles and incentives, improving forecasts by removing intentional biases [61,62]. It has also been observed elsewhere [26] that especially the literature on supply chain integration is heavily skewed towards case study approaches. As a consequence, no studies exist that explicitly put Oliva and Watson's suggestion to the test, and examine forecasting behavior under particular weighting schemes.

2.3. Research questions and hypotheses

Systematic errors in forecasting accuracy caused by intentional biases have been largely ignored in research, but it was suggested that weighting schemes can remove intentional biases [61,62]. We thus examine the design of the forecasting process, and how particular design choices affect intentional biases. Specifically, the research question is: How do roles and incentives affect the intentional biases of forecasters, and are these biases strengthened or mitigated by various popular weighting schemes?

Role- and incentive-based heterogeneity impact decision making, and engender both unintentional and intentional forecast biases. We posit that these biases can – and should – be separated from each other [13,24,31,40]. This can be done by distinguishing between a private forecast and a shared production quantity, determined sequentially by forecasters. The private forecast may be subject to unintentional biases, as it reflects expectations. The difference between the private forecasts and the proposed production quantities reflects the intentional bias in isolation from unintentional forecasting biases.

In recent years, it has been acknowledged that collaborative forecasting is strongly impacted by institutional forces – especially due to cross-functional differences in opinion and approach [26,28,81]. Admittedly, the S&OP process was also partly designed to formally address such cross-functional tensions. Eksoz et al. [26, p. 122] observe that "S&OP enables partners to internally synchronize their strategic, operational and financial plans to achieve consensus on a single plan." Nevertheless, previous research, with Leitax as a case in point, ascribes intentional biases solely to financial incentives [61,62]. Still, departmental roles, even when unconnected to rewards or penalties, can affect intentional biases, by implying goals. As prior research shows that heterogeneous departmental roles and incentives may cause unintentional biases [42,64,89], we hypothesize that both may also cause intentional biases and can substantially impact performance.

Hypothesis 1. Organizational roles, even without role-specific financial incentives, entail intentional biases.

Hypothesis 2. Financial role-specific incentives enlarge the intentional biases induced by organizational roles.

Interestingly, the Leitax case [61,62] demonstrates that design choices may positively influence S&OP. A key finding is that the re-organization of the S&OP process – i.e., such that separate forecasts of sales, product planning, strategy, and demand management together determine the final consensus forecast based on a weighting scheme derived from their past performance – apparently pays off. Prior research had already pointed at the benefits for forecasting of reliance on pooled historical and recent information and data [70]. Intriguingly, for Leitax specifically this led the authors to conclude that weighting schemes can negate the effects of role- and incentive-based heterogeneity by removing intentional biases [61,62]. However, no studies exist that explicitly put this suggestion to the test, and examine forecasting behavior under particular weighting schemes. In this research, we do so, and test whether the accuracy-weighted scheme reduces intentional biases relative to using the simple average.

Hypothesis 3. The accuracy-weighted scheme to combine separate forecasts results in lower intentional biases than the simple average of forecasts.

The different combination schemes discussed do not incorporate interaction between forecasters for a single decision. Yet, the S&OP process and subsequent forecast performance hinges on how members interact with each other [35,38]. A key finding in behavioral research on group decision making is that group members in negotiation meetings hardly share unique information; instead, the group decisions are made on pooled information that was already available to all group members individually. The unique information, referred to in this literature as 'the hidden profile', is hardly ever shared or discussed in the group decision making process [75]. Groups thus are implicitly biased against exchanging with others outside their in-group unique information (or judgments) that could lead to more accurate group decision making – an observation with implications for demand forecasting [38]. Nauta and Sanders [59] and Nauta et al. [58] mention operations and sales as examples of organizational departments that commonly have opposing interests, an observation that was also more recently made in the context of forecasting in supply chains [26,28]. It thus makes sense that information exchange in interactive negotiation meetings among operations and sales representatives will be particularly hampered by strategic and tactical, role- and incentive-based, considerations. By allowing revisions in response to other inputs and by having roles with opposing interests, a weighting scheme resembles a negotiation. This is likely to emphasize the competitive nature of the process and increase intentional biases.

Hypothesis 4. Incorporating interactions between members in the forecast meeting increases intentional biases.

3. Experimental design and data

A laboratory experiment was conducted to evaluate our hypotheses about forecaster behavior under different organizational roles, incentives, and weighting schemes. This section describes the experimental design, the nature of our sample, the role of the computer agent, roles and incentives, and the weighting schemes used.

3.1. Design

The experiment consisted of three phases, see Fig. 1: a training phase, in which participants were familiarized with the forecasting task, and two phases, Phase I and Phase II, that allowed for interaction among decision makers. In each phase, participants were provided with a time series of 18 periods of historic demand of a perishable single-period product, and asked to separately provide a private demand forecast and a shared production quantity for the next period. Similar to Kremer et al. [39], demand was simulated with a local level model, also known as a random walk with noise:

$$\begin{aligned} d_t &= l_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \\ l_t &= l_{t-1} + \nu_t, \qquad \nu_t \sim N(0, \sigma_\nu^2) \end{aligned} \tag{1}$$

In our case, σ_ν^2 and σ_ε^2 are set equal to 100, thus implying equal parts of signal and noise. Participants repeated the forecasting and quantity decision tasks for 18 periods. After each period, they received information about the actual demand and the final production decision, and updates of the available information —profit, lost sales, obsolescence, and forecast accuracy. The forecasts were neutral, representing the participants' expected demand, and were separate from the desired production quantities. The intentional bias was defined as the difference between the demand forecast and the proposed production quantity.

The experiment had a $2 \times 2 \times 2 \times 2$ mixed factorial design, varying with respect to: (*i*) departmental role (sales or operations); (*ii*) incentive scheme (absence or presence of a role-specific financial incentive); (*iii*) incentive of the computer agent (absence or presence of a role-specific financial incentive); and (*iv*) weighting scheme used (accuracy-weighted or interaction); see Table 1. Upon entering the experiment, participants were randomly allocated to one of the four conditions defined by departmental role (sales or operations), and incentive (not role-specific company incentive or role-specific department incentive). In addition, they were paired with a manager from the other department (operations or sales), represented by a computer agent. This other manager was randomly allocated to the presence or absence of a role-specific incentive. The involvement of a computer agent allowed us to simulate interdepartmental decision-making in a fully controlled environment. Importantly, the weighting scheme that defined the final production quantity differed between the phases of the experiment. Phase I took a simple average of the proposed production quantities of participant and agent, whereas Phase II either used an accuracy-weighted scheme (Phase IIa) or explicitly allowed for interaction between participant and agent to determine the final production quantity (Phase IIb).

3.2. Participants

The experiment was performed with three groups of participants. The first sample consisted of 357 students of a Business Administration program in the Netherlands (240 men and 117 women, mean age 21), all familiar with the topic. The second sample consisted of 72 practitioners (51 men and 21 women, mean age 34), all experienced forecasters or demand planners from various manufacturing companies. The third sample was similar to the first sample, and consisted of 97 students (46 men and 51 women, mean age 22). Behavioral experiments are commonly conducted with students to ensure that analyses are based on a large number of participants [e.g. [9,11,39,71]]. Importantly, experienced managers and students generally exhibit the same behavior while working on operations management problems – that is, the practitioner's level of experience does not set them apart from the student [10]. All samples performed Phase I of the experiment, which involved simple weighting of the production quantities. Additionally, the first sample experienced accuracy-based weighting of production quantities in Phase IIa, whereas the second and third sample experienced interaction with the other manager in Phase IIb, see Fig. 1. The third sample is only used for robustness – i.e., to ensure the results match with those of the second sample, the group of practitioners. Table 2 summarizes the allocation of participants from the samples to each role and incentive, for both samples.

Participants who did not correctly answer the control questions about the forecasting task, their role and incentive scheme at the end of the experiment, were excluded from further analyses. Further, participants who made typographical errors were left out. This was necessary, because even simple input errors influenced subsequent rounds of the experiment – something, which could not be corrected. For the first sample, out of the initial 467 participants (321 men and 146 women with an average age of 21),

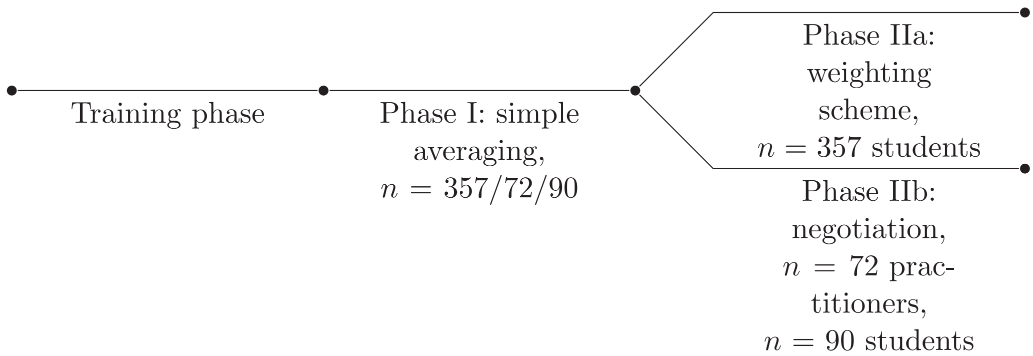

Fig. 1. Schematic presentation of the experiment.

Table 1
Summary of the mixed factorial design.

Department role		Financial incentive		Agent's incentive		Basis weighting scheme	
Sales	Operations	Absent	Present	Absent	Present	Accuracy	Interaction
×		×		×		×	
×		×		×			×
×		×			×	×	
×		×			×		×
×			×	×		×	
×			×	×			×
×			×		×	×	
×			×		×		×
	×	×		×		×	
	×	×		×			×
	×	×			×	×	
	×	×			×		×
	×		×	×		×	
	×		×	×			×
	×		×		×	×	
	×		×		×		×

Table 2
Experimental data over the four conditions for the two conditions of the weighting scheme. The four experimental conditions are based on the two roles of operations and sales, and the incentives of either department or company.

	Company incentive (Not role-specific)		Department incentive (Role-specific)	
	Operations	Sales	Operations	Sales
Accuracy-weighted				
Total: 357 students	92 (25.77%)	91 (25.49%)	85 (23.81%)	89 (24.93%)
Interaction				
Total: 72 practitioners	16 (22.22%)	19 (26.39%)	19 (26.39%)	18 (25.00%)
Total: 90 students	22 (24.44%)	23 (25.56%)	22 (24.44%)	23 (25.56%)

110 participants (24%) were dropped due to typographical errors, which corresponded to an input accuracy of over 99%. For the second sample no participants were dropped. For the additional sample of Phase IIb with students, six students were dropped due to not passing the manipulation check, leading to a sample of 90 students (43 men and 47 women, mean age 22).

The laboratory experiment was programmed in PHP and was accessible via a web browser, which ran on the computers at the behavioral laboratory of the university. Participants first received the background information about the company, a producer of fresh juice for which they presumably worked as a demand planner, and instructions containing the various manipulations. The detailed instructions provided to participants, and the specific questions asked such as manipulation checks, has been published [66, see esp. pages 58–64 and 79–83].

Below, we further specify the experiment, by detailing the forecast and production decisions of the computer agent, the roles and incentives of participants and agents, the weighting schemes applied, and the measures used in the analysis.

3.3. Forecasts and production decisions of the computer agent

The computer agent, who represents the manager of the other department involved in the planning, was designed to make demand forecasts and production decisions depending on its role and incentives in the experiment. Below, we subsequently describe the agent's forecast behavior, and the way it decides about the production quantity.

The agent's demand forecast. Agents adopt single exponential smoothing to forecast the next period's demand:

$$\hat{d}_{t+1|t} = \hat{d}_{t|t-1} + \alpha(d_t - \hat{d}_{t|t-1}) \quad (2)$$

where $\hat{d}_{t+1|t}$ denotes the forecast of demand for period $t+1$ at time t, d_t denotes the observed demand at time t, and α is a smoothing parameter. For a local level model (1), single exponential smoothing is the optimal forecast method [46]. The smoothing parameter α is set to minimize the mean squared forecast error, which gives [see, e.g., 25]:

$$\alpha^* = \frac{\sqrt{r(r+4)} - r}{2} \quad (3)$$

where $r = \sigma_\nu^2/\sigma_\varepsilon^2$ is the signal-to-noise ratio.

Please cite this article as: C.L.P. Pennings et al., Coordinating judgmental forecasting: Coping with intentional biases, Omega (2018), https://doi.org/10.1016/j.omega.2018.08.007

The agent's proposed production quantity. Following Schweitzer and Cachon [71], the agent's production quantity decision is based on the newsvendor model, in which a production quantity $q_{t+1|t}$ is decided at time t for sale during the next period $t+1$, knowing that the produced quantity is only available during time period $t+1$. Inventory costs are ignored, since a perishable single-period product is considered. After observing the outcome in period $t+1$, along with its consequences for lost sales and obsolete products, a new production quantity has to be set for period $t+2$. The agent is designed to maximize expected profits.

If we denote production quantity as q, demand as d and unit production costs as c, and consider that the number of units sold is equal to the minimum of the quantity produced and the demand, then profit $\pi(q, d)$ can be defined as:

$$\pi(q,d) = p \cdot \min(q,d) - c \cdot q \tag{4}$$

Moreover, assuming that the demand has distribution F, with density f, expected profit can be derived as:

$$E[\pi(q,d)] = [1-F(q)]\pi(q,q) + \int_0^q f(x)\pi(q,x)dx \tag{5}$$

The optimal production quantity, q^*, maximizes the expected profit (5) by balancing the costs of lost sales $(p-c)$ and the total cost (p) of being either overstocked (c) or understocked $(p-c)$, and follows from:

$$F(q^*) = \frac{p-c}{c+(p-c)} = \frac{p-c}{p} \tag{6}$$

which is referred to as the critical fractile [71].

The optimal order quantity is obtained as the inverse of this critical fractile (6). Considering that demand is generated by the local level model (1), Durbin and Koopman [25] show that the demand distribution F is normal with mean $\hat{d}_{t+1|t}$ and variance $\mathrm{Var}(\hat{d}_{t+1|t})$, $d_{t+1|t} \sim N(\hat{d}_{t+1|t}, \mathrm{Var}(\hat{d}_{t+1|t}))$, where the mean is based on (2) and the variance is derived as:

$$\mathrm{Var}(\hat{d}_{t+1|t}) = \sigma_\varepsilon^2 (r + \sqrt{r^2+4r})/2 + \sigma_\varepsilon^2 \tag{7}$$

with $r = \sigma_\nu^2/\sigma_\varepsilon^2$ the signal-to-noise ratio.

The optimal production quantity decision q^* can now be obtained by applying the inverse distribution function to the critical fractile (6):

$$q^*_{t+1|t} = F^{-1}\left(\frac{p-c}{p}\right) = \hat{d}_{t+1|t} + \sqrt{\mathrm{Var}(\hat{d}_{t+1|t})} \cdot \Phi^{-1}\left(\frac{p-c}{p}\right) \tag{8}$$

which gives the optimal order quantity, based on the optimal forecast and the trade-off of being either overstocked or understocked. Agents derive the optimal order quantity (8) using the optimal forecast and the steady state values of (7).

3.4. Roles and incentives

Participants are randomly assigned the role of either operations or sales managers, and are given a role-specific department or role-independent company incentive. The computer agent's role is complementary to the participant's role: it takes on the role of operations manager, when the participant has the role of sales manager; it adopts the role of sales manager, when the participant has the role of operations manager. The operations department focuses on production and inventory levels; the sales department focuses on product availability. Participants are penalized for outcomes straying from their department's objective, which is minimizing either obsolescence or lost sales, or from the company's objective, which is maximizing profit and minimizing ex-post inventory error.

Under incentives for the company's objective, the sales price p is 2 Euro and the cost of production c is 1 Euro. The optimal order quantity $q^*_{t+1|t}$ at time t, following (8), is equal to the expected demand $\hat{d}_{t+1|t}$, regardless of the departmental role, because of a symmetrical cost structure of lost sales and obsolescence. In this case, there is no incentive for an intentional bias.

Under incentives for the department's objective, the cost structure is asymmetric. The sales price p remains 2 Euro. Operations managers are penalized for obsolescence, which doubles the associated cost, while sales managers are similarly penalized for lost sales. This shifts the trade-off between the costs of lost sales and the total cost of being overstocked or understocked, which changes the critical fractile (6) for operations managers and sales managers respectively to:

$$F(q^*_{\text{operations}}) = \frac{p-c}{2c+(p-c)}$$
$$F(q^*_{\text{sales}}) = \frac{2(p-c)}{c+2(p-c)} \tag{9}$$

As a result, for $p=2$ and $c=1$ the optimal order quantities q^* (8) for operations managers and sales managers respectively become:

$$q^*_{\text{operations},t+1|t} = \hat{d}_{t+1|t} + \sqrt{\mathrm{Var}(\hat{d}_{t+1|t})}\,\Phi^{-1}\left(\frac{1}{3}\right)$$
$$q^*_{\text{sales},t+1|t} = \hat{d}_{t+1|t} + \sqrt{\mathrm{Var}(\hat{d}_{t+1|t})}\,\Phi^{-1}\left(\frac{2}{3}\right) \tag{10}$$

The incentives of these two roles are symmetric around the expected demand. If a simple average is taken of the two optimal order quantities, and if behavior is rational and based on unbiased forecasts, then the effects of the incentives will cancel out.

The change in the optimal order quantity illustrates the effect of the role-specific department incentive. In addition, (8) and (10) specify the desired production quantity of agents with a company objective (not role-specific) or departmental objective (role-specific), respectively.

3.5. Weighting schemes

Three different weighting schemes are used in different phases of the experiment in order to combine inputs from the participant and the agent into a final production quantity. In Phase I, the weighting scheme consists of a simple average of inputs, calculating the production quantity q_o, based on the proposed quantities of participant r and agent a, in each time period t, as:

$$q_{o,t} = (q_{r,t} + q_{a,t})/2 \tag{11}$$

In Phase IIa, the weighting scheme is accuracy-weighted, calculating the production quantity as a weighted average based on the past performance of the participant and agent, while ignoring covariance:

$$q_{o,t} = w_{r,t} q_{r,t} + (1-w_{r,t}) q_{a,t} \tag{12}$$

with:

$$w_{r,t} = \frac{e_{a,t}}{e_{r,t}+e_{a,t}}, \quad e_{r,t} = \frac{1}{t-1}\sum_{i=1}^{t-1}(q_{r,i}-d_i)^2 \text{ and}$$

$$e_{a,t} = \frac{1}{t-1}\sum_{i=1}^{t-1}(q_{a,i}-d_i)^2$$

Note that $q_{o,\,t}$ is a convex combination of the inputs based on the observed forecast accuracy up to time period t. This scheme is frequently applied in practice [8,17,86].

In Phase IIb, the weighting scheme facilitates interaction between participant and computer agent, by allowing them to revise their inputs after seeing each other's input in each period. By having roles with opposing interests, this weighting scheme resembles a negotiation. Due to practical constraints (i.e., the duration of the experiment), interaction is limited to four rounds. We pragmatically posit this is sufficient to determine whether the scheme affects participant behavior. In each negotiation round, the agent and the participant simultaneously propose production quantities. They can update these proposed quantities in the next negotiation round after seeing each other's proposals. Participants face no restrictions when revising the proposed production quantity: they can increase, decrease, or leave the proposed quantity unchanged. Agent behavior is restricted, as will be explained below. After the last round of interaction, the average of the last two inputs of the agent and the participant defines the production quantity.

The agent follows a simple algorithm during interactions, which includes random variation to avoid deterministic and predictable behavior. The agent neither behaves competitively, nor does it punish the participant: it either adjusts its input towards the participant's, or leaves it unchanged. Also, the agent never moves outside the range set by the 5th and 95th percentiles of the forecast distribution, to limit its reaction to possibly extreme inputs by participants. In the first negotiation round, the agent states its desired production quantity. In the second negotiation round, the agent adjusts its quantity in order to reduce the gap between its own and the participant's quantity. The agent's behavior in the third and fourth round depends on the preceding actions of participants. If the participant's quantity is not closer towards the agent's proposal, the agent will not adjust its proposal in return. But, if the participant decreases the gap, the agent will further adjust its proposal towards the participant.

If the agent adjusts its quantity, it will make its adjustment by one third of the distance between its own most recent proposal q_a and the latest proposal of the participant q_r, representing a substantial step towards the participant's quantity. The adjustment includes a random factor in order to avoid deterministic behavior, which can be quickly learned by participants:

$$\begin{aligned} x \quad & \sim \text{Beta}(2,2) \\ q_a \quad & := q_a + \tfrac{q_r - q_a}{3}(0.5 + x) \end{aligned} \tag{13}$$

This particular beta distribution is used, because it restricts the random factor x to $0 \le x \le 1$, is symmetric around the mean equal to 0.5, and its probability mass is highest at the mean after which it tapers off for higher or lower values of x.

3.6. Measures and analyses

The main measure of interest is the intentional bias, defined as the difference between participants' forecast $d_{i,\,t}$ and their production quantity $q_{i,\,t}$:

$$\delta_{i,t} = \begin{cases} d_{i,t} - q_{i,t} & \text{if } i \text{ has an operations role} \\ q_{i,t} - d_{i,t} & \text{if } i \text{ has a sales role} \end{cases} \tag{14}$$

We define the intentional bias separately for operations and sales roles to ensure that the intentional bias follows from the context: a positive bias for operations means that the forecast is deflated, whereas a positive bias for sales means that the forecast is inflated. For interactions, the final proposed quantity is used.

Preliminary insight into the effect of different incentives and roles is obtained by graphing the average of δ_t per incentive type over time. In addition, the intentional bias $\delta_{i,\,t}$ is modeled as a first-order autoregressive process AR(1) with random slopes α_i and coefficients ϕ_i to incorporate heterogeneity:

$$\begin{aligned} \delta_{i,t} &= \alpha_i + \phi_i \delta_{i,t-1} + \beta \nu_i + \eta_{it} \\ \eta_{it} &\sim N(0, \sigma_\eta^2) \\ \alpha_i &\sim N(\mu_\alpha, \sigma_\alpha^2) \\ \phi_i &\sim N(\mu_\phi, \sigma_\phi^2) \end{aligned} \tag{15}$$

The dummy variable ν_i indicates the absence ($\nu_i = 0$) or presence ($\nu_i = 1$) of the role-specific department incentive for participants. If $|\phi_i| < 1$, the autoregressive process is stationary, and the mean of δ_i is:

$$E[\delta_i] = \frac{\alpha_i + \beta \nu_i}{1 - \phi_i} \tag{16}$$

A positive mean for participants with the company incentive ($\nu_i = 0$) gives evidence for Hypothesis (1) that roles, even without role-specific financial incentives, entail intentional biases. An estimated β that is substantially and significantly higher than zero gives evidence for Hypothesis (2) that role-specific incentives result in larger intentional biases.

Behavior under the different weighting schemes is compared using the mean average intentional bias. In Phase I, the simple average is used; in Phase II, either the accuracy-weighted quantity or the negotiated quantity between the participant and the agent is used. The change in the intentional bias from Phase I to Phase II is evaluated using a Wilcoxon signed-rank test, to determine whether intentional biases are lower under the accuracy-weighted scheme, as posited by Hypothesis (3), and whether intentional biases are higher when there is interaction between participant and agent, as posited by Hypothesis (4).

If the intentional biases are not fully removed by the accuracy-weighted scheme, a possible alternative is offered by a scheme which de-biases the inputs. Suppose the intentional biases $\delta_{i,\,t}$ can be decomposed into a constant bias θ and a random component. If the biases are constant and do not cancel out, the inputs can be explicitly de-biased by estimating constant biases θ for each input. Given θ_r and θ_a, the estimated constant biases for the participant and the agent, the weights can be calculated as [cf. 65]:

$$\begin{aligned} w_{r,t} &= \frac{e_{a,t} + \theta_{a,t}(\theta_{a,t} - \theta_{r,t})}{e_{r,t} + e_{a,t} + (\theta_{r,t} - \theta_{a,t})^2} \\ w_{a,t} &= 1 - w_{r,t} \end{aligned} \tag{17}$$

In the first time period, the simple average of the proposed quantities is used. In subsequent time periods, the biases $\theta_{i,\,t}$, which is the sum of intentional and unintentional biases, up to time period t is calculated as:

$$\theta_{i,t} = \frac{1}{t - 1 - 18} \sum_{n=19}^{t-1} (q_{i,n} - d_n) \tag{18}$$

Performance under this de-biasing scheme is compared to the performance under the other weighting schemes, using the root-mean-square error (RMSE), a popular measure for forecasting accuracy. The MSE is attractive because it incorporates both the variance and the bias of the estimator. The root is taken so the measurement is on the same scale as the data. A Wilcoxon signed-rank test is used to determine whether this scheme outperforms weighting schemes without such a correction. Such a result is not straightforward, as performance can also deteriorate because of misspecification of the scheme and estimation errors. As this scheme is not as simple to interpret, its effect on behavior is not examined and is not part of the experiment. However, combining inputs post-hoc using this scheme may demonstrate its potential value.

4. Results

We first examine the average intentional bias of participants in the laboratory experiment for the two types of incentives. We then examine the results of our estimated statistical model (15) and statistical tests, and discuss the implications of these results for our hypotheses in turn. The results of the third sample matched with the results of the second sample, so that only the second sample is presented here.

Table 3
Mean intentional biases under the various weighting schemes. Mean and standard errors (in parentheses) of the observed intentional biases, determined for sub-samples defined by type of incentive and organizational role. Results pertain to the three weighting schemes: simple average (Phase I), accuracy-weighted (Phase IIa), and interaction (Phase IIb). Differences between incentives within weighting schemes and between the simple average and alternative weighting schemes are all significant at the 5% level.

	Company incentive (Not role-specific)		Department incentive (Role-specific)	
	Operations	Sales	Operations	Sales
Simple average (Phase I)	39.89 (5.73)	42.92 (3.19)	79.99 (3.65)	77.23 (3.69)
Accuracy-weighted (Phase IIa)	24.75 (4.52)	25.69 (2.56)	42.00 (3.45)	34.51 (3.10)
Interaction (Phase IIb)	27.44 (9.95)	28.39 (5.69)	17.94 (7.52)	29.28 (7.14)

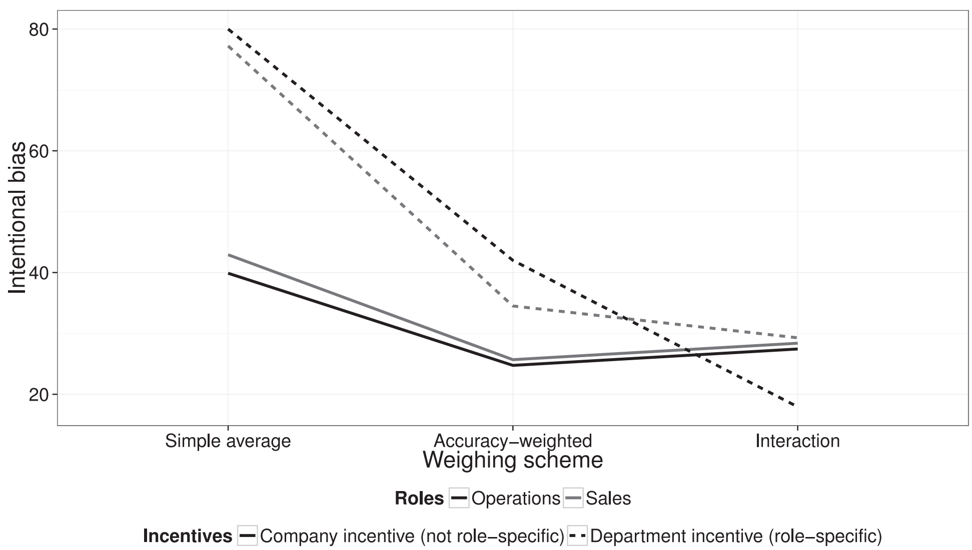

Fig. 2. Average intentional bias for roles, incentives and weighting schemes. This figure shows the descriptive intentional biases from Table 3.

4.1. Descriptive analysis of intentional biases by role and incentive

Table 3 summarizes the means and standard errors of the intentional biases per role and incentive for each of the weighting schemes; Fig. 2 illustrates these outcomes.

Under the simple average scheme (Phase I), the biases for the company incentive are substantial, and are seen to be larger for sales (42.921 (s.e.3.192)) than for operations (39.886 (s.e.5.731)). This implies that organizational roles, even without a role-specific incentive, induce intentional biases, supporting Hypothesis 1. The mean intentional biases substantially increase when department incentives apply. In fact, they increase relatively more for operations managers, which reveal a mean bias equal to 79.998 (s.e.3.654), which is larger than the 77.229 (s.e.3.685) observed for sales managers. This shows that financial, role-specific incentives favor intentional biases, supporting Hypothesis 2, and moreover, that the incentive effects interact with the organizational roles of managers.

A similar pattern is observed for the accuracy-weighted scheme (Phase IIa), except that the intentional biases is almost half of that under the simple average weighting scheme. The smallest change, from 39.886 to 24.751 (38%), occurs for operations managers under a company incentive, and the largest change, from 77.229 to 34.507 (55%), for sales managers under a department incentive. The accuracy-weighted scheme thus reduces, but does not remove, the intentional bias. The lowest bias, 24.751 for operations under a company incentive, is still substantial. So, accuracy-weighting of quantity decisions, though beneficial from the perspective of reducing intentional biases, neither removes the impact of role-specific incentives or eliminates the bias due to the organizational roles. This supports Hypothesis 3.

Including interaction between the participant and agent (Phase IIb), allowing them to revise their inputs in several negotiation rounds, lowers the intentional biases to a similar extent as the accuracy-weighted scheme does, albeit with more variation: the smallest drop, from 39.886 to 27.442 (31%), is for operations under a company incentive, and the largest drop, from 79.998 to 17.942 (78%), occurs for operations under a department incentive. This contrasts with Hypothesis 4, which posits that interactions emphasize the competitive nature of the task. Instead, this form of interaction appears to stimulate cooperation, reducing participants' intentional bias. Unlike for simple averaging and accuracy weighting schemes, the consideration of interaction affects the impact of incentives likewise for operation and sales managers – no moderation of incentives and roles is observed.

4.2. The development of intentional biases over time

Fig. 3 shows the development of the mean intentional bias over time, separately for the department and company incentives, assuming the simple-averaging weighting scheme (Phase I). Interestingly, the intentional bias appears to be positive and substantial under both incentives, and seems to increase over time. Further, the intentional bias is consistently and substantially larger for the department incentive than for the company incentive. This is in line with theorizing on the strategic nature of intentional information processing [13,31], and highlights that deliberation may lead to biases in favor of departmental stimuli.

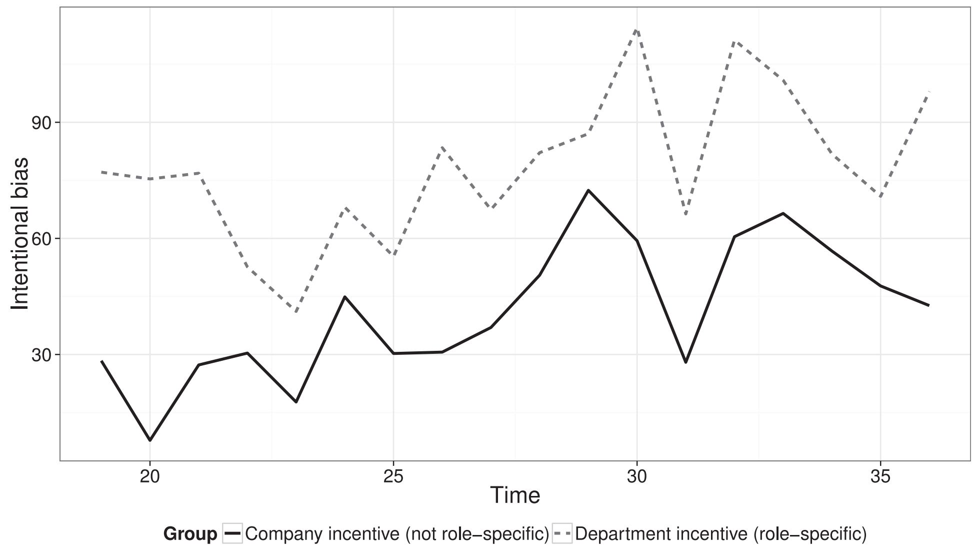

Fig. 3. Mean intentional biases. This figure shows the mean intentional bias δ_t aggregated over participants (students and practitioners combined) per time period, for the department and company incentive with the simple average as the weighting scheme.

Table 4
Estimates of the AR(1) model for intentional adjustments. Estimates of model (15) for the groups of students and practitioners separately. The effect of the department incentive β is significant. Standard errors are in parentheses.

	Students (n = 357)	Practitioners (n = 72)
μ_α	32.972 (0.122)	33.824 (0.756)
μ_ϕ	0.298 (0.079)	0.249 (0.655)
β	24.826 (0.003)	25.370 (0.721)
σ^2_α	8.661 (0.978)	12.685 (0.431)
σ^2_ϕ	0.091 (0.266)	0.143 (0.734)

Table 5
Accuracy. Accuracy under the various weighting schemes. Standard errors in parentheses.

	RMSE
Accuracy-weighted	197.163 (1.821)
Interaction	203.934 (3.547)
De-biasing scheme	128.097 (1.593)

The persistence of intentional biases is further illustrated by the parameter estimates for the autoregressive state space model (15) in Table 4. Intentional biases are strongly serially dependent ($\mu_\phi = 0.298(0.079)$ for the student sample, and $\mu_\phi = 0.249(0.655)$ for the practitioner sample): biases observed in one planning round tend to be predictive for the biases in the next round. Further, the estimated effect of role-specific department incentive β is substantially and significantly different from zero, for both students ($\beta = 24.826, s.e. = 0.003$) and practitioners ($\beta = 25.370, s.e. = 0.721$). This indicates that department incentives increase intentional biases, which is consistent with Hypothesis 2. Furthermore, the mean (16) of the intentional biases is equal to 46.968 for the company incentive, and equal to 82.333 for the department incentive. As the average forecast is equal to 572.856, the intentional bias for the company incentive and for the department incentive are respectively 8% and 14% relative to the private demand forecast. The intentional bias under the company incentive shows that roles, even without role-specific incentives, entail intentional biases. This is in line with theorizing [13,31], and offers support for Hypothesis 1.

4.3. The impact of de-biasing the weighting scheme

The three weighting schemes – i.e., simple averaging (11), accuracy-weighting (12), and interaction between participant and agent – each influence forecasting behavior and affect the size of the intentional biases. However, none of the schemes fully removes the intentional biases. Moreover, the private demand forecasts also contain unintentional biases, as theoretically assumed in the wider literature on goals and information processing [24,40]. All three weighting schemes, though commonly used in practice, are deficient in this regard. To explore the possible impact of taking these biases into account in the weighting scheme, the weights given to inputs are re-calculated using (17), to explicitly remove biases from the inputs. Table 5 shows the accuracy under the accuracy-weighted scheme, the interaction scheme, and the de-biasing scheme (17). The de-biasing scheme greatly reduces inaccuracy by 35% compared to the accuracy-weighted scheme. Compared to the interaction scheme, the de-biasing scheme reduces inaccuracy by 37%. Even such a basic adjustment, which estimates and removes a constant bias from the inputs of the participant, apparently yields considerable gains. Even though de-biasing was not implemented in the forecast experiment, and participants' responses to the use of such a scheme were not measured, this substantial improvement in performance holds promises for practice.

5. Conclusions

The present research examined the existence of intentional biases in forecast behavior in the context of different departmental roles and incentives. Specifically, we conducted an elaborate laboratory experiment to simulate forecasting and production quantity decisions in an interdepartmental decision-making context that resembled the Sales & Operations Planning (S&OP) process of real organizations [44,48,62]. Much like Önkal et al. [64], we offered our participants – a large group of students and practitioners – specific role descriptions that were intended to render salient a typical departmental (sales or operations) role, and evaluated the effects of those departmental roles on their forecasting decisions.

Please cite this article as: C.L.P. Pennings et al., Coordinating judgmental forecasting: Coping with intentional biases, Omega (2018), https://doi.org/10.1016/j.omega.2018.08.007

Also, we investigated the extent to which (presence or absence of) incentives, and (presence of accuracy-weighted or interaction-based) weighting scheme impacted forecasting decisions. We found that departmental roles, even in the absence of role-specific incentives, entailed intentional biases of 8% of the forecast, whereas role-induced decision making in the presence of role-specific incentives increased these biases to 14%. We, further, put to the test a claim that was first made in Oliva and Watson [62] that an accuracy-weighted scheme can remove unintentional biases. The results of our study allowed us to conclude that an accuracy-weighted scheme is capable of halving, but not fully removing, unintentional biases.

The contributions of our work are threefold. First, our study contributes to the accumulating evidence for the impact of organizational roles on demand forecasting. Consistent with extant research by Kuo and Liang [42], Önkal et al. [64], and Yaniv [89], our findings indicate that the departmental lens through which a decision maker looks at the forecast determines the outcome of the demand forecasting process. Second, and more importantly, our behavioral experiment is the first to isolate intentional forecast biases from unintentional ones in the setting of demand forecasting. Thus far, unlike behavioral researchers [13,24,31,40], scholars in the domain of supply chain management did not make this distinction. Our work clearly emphasizes the need to identify intentionality (i.e., strategic, goal-directed, role-inspired biases) in the setting of demand forecasting – as it is the intentional biases that are affected by weighting schemes. Third, and related, our study to large extent expands the case study of Oliva and Watson [61] and Oliva and Watson [62] to a laboratory setting. This allowed us to provide empirical support for their suggestion that weighting schemes have the potential to compensate for unintentional biases in S&OP decisions. Our study shows that this may be the case, even though we acknowledge the limits of the accuracy-weighted scheme, which does not entirely remove intentional biases. Interestingly, however, a simple de-biasing scheme that we tentatively explored, showed great promise in reducing intentional as well as unintentional biases by 35%. Future research should more carefully explore the merits of this alternative scheme, and address its potential for demand forecasting, also in comparison to other weighting schemes.

The current research was limited to the simple dichotomy in departmental roles of sales and operations. Even though this dichotomy has long been acknowledged among academics and practitioners as a ‘classic divide’ in supply chain management [26,44,58,59,70,72,80,81], more departmental roles, such as marketing and finance, should be taken into account in future research; see also [62]. It would be particularly fruitful to work towards a thorough inventory of the various role-specific, intentional biases in demand forecasting. Such an inventory could serve as the starting point in the development and testing of an “adaptive toolbox” [32] of targeted de-biasing schemes, designed to remove the decision biases associated with a specific departmental role. Such role-dependent de-biasing schemes would be beneficial in traditional demand forecasting processes, and would also add value to overcoming flawed decision making of employees in a demand forecasting landscape determined by big data and predictive analytics [15,53,73,83].

Likewise, our current insights have important consequences for the design of the forecasting process in terms of coordination mechanisms and incentives [74]. This is important for organizational practice, because forecasters’ behavior directly affects forecast performance, and can have large financial ramifications [29,77,78]. Our work on disentangling specific design choices and examining these in isolation paves the way for future work on forecast process design – specifically on the potential performance gain of weighting schemes. More immediately, however, it presses for a careful review of current policies, because choices in terms of roles, incentives, and weighting schemes meant to increase performance can have a detrimental effect.

References

[1] Anandalingam G, Chen L. Bayesian forecast combination and kalman filtering. Int J Syst Sci 1989;20(8):1499–507.
[2] Armstrong JS. The handbook of forecasting: the forecasting audit. In: Makridakis S, Wheelwright SC, editors. The Handbook of Forecasting. New York: John Wiley; 1987. p. 584–602.
[3] Armstrong JS. Principles of forecasting: a handbook for researchers and practitioners, 30. Springer Science & Business Media; 2001.
[4] Armstrong JS, Green KC, Graefe A. Golden rule of forecasting: be conservative. J Bus Res 2015;68(8):1717–31.
[5] Ashiya M. Strategic bias and professional affiliations of macroeconomic forecasters. J Forecast 2009;28(2):120–30.
[6] Ashton AH, Ashton RH. Aggregating subjective forecasts: some empirical results. Manage Sci 1985;31(12):1499–508.
[7] Bates J.M., Granger C.W.J.. The combination of forecasts. 1969. OR. 20, 4, 451.
[8] Bjørnland HC, Gerdrup K, Jore AS, Smith C, Thorsrud LA. Does forecast combination improve Norges bank inflation forecasts. Oxf Bull Econ Stat 2012;74(2):163–79.
[9] Bolton GE, Katok E. Learning by doing in the newsvendor problem: a laboratory investigation of the role of experience and feedback. Manuf Serv Oper Manag 2008;10(3):519–38.
[10] Bolton GE, Ockenfels A, Thonemann UW. Managers and students as newsvendors. Manage Sci 2012;58(12):2225–33.
[11] Bostian AA, Holt CA, Smith AM. Newsvendor ‘pull-to-center’ effect: adaptive learning in a laboratory experiment. Manuf Serv Oper Manag 2008;10(4):590–608.
[12] Brockhoff K. Group processes for forecasting. Eur J Oper Res 1983;13(2):115–27.
[13] Camerer CF. Behavioural studies of strategic thinking in games. Trends Cogn Sci (Regul Ed) 2003;7(5):225–31.
[14] Chan YL, Stock JH, Watson MW. A dynamic factor model framework for forecast combination. Span Econ Rev 1999;1(2):91–121.
[15] Chen H, Chiang RHL, Storey VC. Business intelligence and analytics: from big data to big impact. MIS Q 2012:1165–88.
[16] Clemen RT. Combining forecasts: a review and annotated bibliography. Int J Forecast 1989;5(4):559–83.
[17] Clemen RT, Winkler RL. Combining economic forecasts. J Bus Econ Stat 1986;4(1):39–46.
[18] Clemen RT, Winkler RL. Combining probability distributions from experts in risk analysis. Risk Anal 1999;19(2):187–203.
[19] Costantini M, Kunst RM. Combining forecasts based on multiple encompassing tests in a macroeconomic core system. J Forecast 2011;30(6):579–96.
[20] Danese P, Kalchschmidt M. The role of the forecasting process in improving forecast accuracy and operational performance. Int J Prod Econ 2011;131:204–14.
[21] Davis DF, Mentzer JT. Organizational factors in sales forecasting management. Int J Forecast 2007;23(3):475–95.
[22] Diebold FX, Pauly P. Structural change and the combination of forecasts. J Forecast 1987;6(1):21–40.
[23] Diebold FX, Pauly P. The use of prior information in forecast combination. Int J Forecast 1990;6(4):503–8.
[24] Dijksterhuis A, Aarts H. Goals, attention, and (un) consciousness. Annu Rev Psychol 2010;61:467–90.
[25] Durbin J, Koopman SJ. Time series analysis by state space methods. Oxford University Press; 2012.
[26] Eksoz C, Mansouri SA, Bourlakis M. Collaborative forecasting in the food supply chain: a conceptual framework. Int J Prod Econ 2014;158:120–35.
[27] Fang Y. Forecasting combination and encompassing tests. Int J Forecast 2003;19(1):87–94.
[28] Fildes R, Goodwin P, Lawrence M, Nikolopoulos K. Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. Int J Forecast 2009;25(1):3–23.
[29] Fildes R, Nikolopoulos K, Crone SF, Syntetos AA. Forecasting and operational research: a review. J Oper Res Soc 2008;59(9):1150–72.
[30] Fildes R, Petropoulos F. Is there a golden rule? J Bus Res 2015;68(8):1742–5.
[31] Gigerenzer G, Goldstein DG. Reasoning the fast and frugal way: models of bounded rationality. Psychol Rev 1996;103(4):650.
[32] Gigerenzer G, Selten R. Bounded rationality: the adaptive toolbox. MIT Press; 2002.
[33] Gino F, Pisano G. Toward a theory of behavioral operations. Manuf Serv Oper Manag 2008;10(4):676–91.
[34] Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009;457(7232):1012.
[35] Graefe A, Armstrong JS. Comparing face-to-face meetings, nominal groups, delphi and prediction markets on an estimation task. Int J Forecast 2011;27(1):183–95.
[36] Granger CWJ, Ramanathan R. Improved methods of combining forecasts. J Forecast 1984;3(2):197–204.

Please cite this article as: C.L.P. Pennings et al., Coordinating judgmental forecasting: Coping with intentional biases, Omega (2018), https://doi.org/10.1016/j.omega.2018.08.007

[37] Hughes MC. Forecasting practice: organisational issues. J Oper Res Soc 2001;52(2):143–9.
[38] Kerr NL, Tindale RS. Group-based forecasting? A social psychological analysis. Int J Forecast 2011;27(1):14–40.
[39] Kremer M, Moritz B, Siemsen E. Demand forecasting behavior: system neglect and change detection. Manage Sci 2011;57(10):1827–43.
[40] Kruglanski AW, Gigerenzer G. Intuitive and deliberate judgments are based on common principles. Psychol Rev 2011;118(1):97.
[41] Kugler L. What happens when big data blunders? Commun ACM 2016;59(6):15–16.
[42] Kuo Y, Liang K. Human judgments in new york state sales and use tax forecasting. J Forecast 2004;23(4):297–314.
[43] Kurz-Kim J. Combining forecasts using optimal combination weight and generalized autoregression. J Forecast 2008;27(5):419–32.
[44] Lapide L. An s&OP maturity model. J Bus Forecast. 2005;24(3):15.
[45] Lawrence M, Goodwin P, O'Connor M, Önkal D. Judgmental forecasting: a review of progress over the last 25 years. Int J Forecast 2006;22(3):493–518.
[46] Lawrence M, O'Connor M. Exploring judgemental forecasting. Int J Forecast 1992;8(1):15–26.
[47] Lawrence M, O'Connor M. Judgmental forecasting in the presence of loss functions. Int J Forecast 2005;21(1):3–14.
[48] Lawrence M, O'Connor M, Edmundson B. A field study of sales forecasting accuracy and processes. Eur J Oper Res 2000;122(1):151–60.
[49] Lazer D, Kennedy R, King G, Vespignani A. The parable of Google flu: traps in big data analysis. Science 2014;343(6176):1203–5.
[50] Lee WY, Goodwin P, Fildes R, Nikolopoulos K, Lawrence M. Providing support for the use of analogies in demand forecasting tasks. Int J Forecast 2007;23(3):377–90.
[51] Lipscomb J, Parmigiani G, Hasselblad V. Combining expert judgment by hierarchical modeling: an application to physician staffing. Manage Sci 1998;44(2):149–61.
[52] Makridakis S, Winkler RL. Averages of forecasts: some empirical results. Manage Sci 1983;29(9):987–96.
[53] McAfee A, Brynjolfsson E, Davenport TH, Patil DJ, Barton D. Big data: the management revolution. Harv Bus Rev 2012;90(10):60–8.
[54] McCarthy TM, Davis DF, Golicic SL, Mentzer JT. The evolution of sales forecasting management: a 20-year longitudinal study of forecasting practices. J Forecast 2006;25(5):303–24.
[55] Min C, Zellner A. Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. J Econom 1993;56(1):89–118.
[56] Moon MA, Mentzer JT, Smith CD. Conducting a sales forecasting audit. Int J Forecast 2003;19(1):5–25.
[57] Morris PA. Combining expert judgments: a Bayesian approach. Manage Sci 1977;23(7):679–93.
[58] Nauta A, De Dreu CKW, Vaart TVd. Social value orientation, organizational goal concerns and interdepartmental problem-solving behavior. J Organ Behav 2002;23(2):199–213.
[59] Nauta A, Sanders K. Causes and consequences of perceived goal differences between departments within manufacturing organizations. J Occup Organ Psychol 2001;74(3):321–42.
[60] Newbold P, Granger CWJ. Experience with forecasting univariate time series and the combination of forecasts. J R Stat Soc Ser A (Gen) 1974;137(2):131–65.
[61] Oliva R, Watson N. Managing functional biases in organizational forecasts: a case study of consensus forecasting in supply chain planning. Prod Oper Manag 2009;18(2):138–51.
[62] Oliva R, Watson N. Cross-functional alignment in supply chain planning: a case study of sales and operations planning. J Oper Manage 2011;29(5):434–48.
[63] Önkal D, Lawrence M, Sayim KZ. Influence of differentiated roles on group forecasting accuracy. Int J Forecast 2011;27(1):50–68.
[64] Önkal D, Sayim KZ, Lawrence M. Wisdom of group forecasts: does role-playing play a role? Omega (Westport) 2012;40(6):693–702.
[65] Palm FC, Zellner A. To combine or not to combine? issues of combining forecasts. J Forecast 1992;11(8):687–701.
[66] Protzner S. Mind the gap between demand and supply: a behavioral perspective on demand forecasting. ERIM; 2015. Ph.D. thesis.
[67] Rapach DE, Strauss JK. Forecasting US employment growth using forecast combining methods. J Forecast 2008;27(1):75–93.
[68] Sanders NR, Manrodt KB. The efficacy of using judgmental versus quantitative forecasting methods in practice. Omega (Westport) 2003;31(6):511–22.
[69] Sanders NR, Ritzman LP. On knowing when to switch from quantitative to judgemental forecasts. Int J Oper Prod Manag 1991;11(6):27–37.
[70] Sanders N.R., Ritzman L.P.. Integrating judgmental and quantitative forecasts: methodologies for pooling marketing and operations information. 2004. Int J Oper Prod Manag, 24, 5, 514–529.
[71] Schweitzer ME, Cachon GP. Decision bias in the newsvendor problem with a known demand distribution: experimental evidence. Manag Sci 2000;46(3):404–20.
[72] Shapiro BP. Can marketing and manufacturing co-exist. Harv Bus Rev 1977;55(5):104.
[73] Shapiro J. Advanced analytics for sales & operations planning. Anal Mag 2010:20–6.
[74] Singhal J, Singhal K. Holt, modigliani, muth, and simon's work and its role in the renaissance and evolution of operations management. J Oper Manage 2007;25(2):300–9.
[75] Stasser G. Information salience and the discovery of hidden profiles by decision-making groups: a 'thought experiment'. Organ Behav Hum Decis Process 1992;52(1):156–81.
[76] Swanson NR, Zeng T. Choosing among competing econometric forecasts: regression-based forecast combination using model selection. J Forecast 2001;20(6):425–40.
[77] Syntetos AA, Georgantzas NC, Boylan JE, Dangerfield BC. Judgement and supply chain dynamics. J Oper Res Soc 2011;62(6):1138–58.
[78] Syntetos AA, Nikolopoulos K, Boylan JE. Judging the judges through accuracy-implication metrics: the case of inventory forecasting. Int J Forecast 2010;26(1):134–43.
[79] Syntetos AA, Nikolopoulos K, Boylan JE, Fildes R, Goodwin P. The effects of integrating management judgement into intermittent demand forecasts. Int J Prod Econ 2009;118(1):72–81.
[80] Van Bruggen GH, Lilien GL, Kacker M. Informants in organizational marketing research: why use multiple informants and how to aggregate responses. Journal of Marketing Research 2002;39(4):469–78.
[81] Van Bruggen GH, Spann M, Lilien GL, Skiera B. Prediction markets as institutional forecasting support systems. Decis Support Syst 2010;49(4):404–16.
[82] Varian HR. Big data: new tricks for econometrics. J Econ Perspect 2014;28(2):3–28.
[83] Waller MA, Fawcett SE. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. J Bus Logist 2013;34(2):77–84.
[84] Walz DT, Walz DB. Combining forecasts: multiple regression versus a Bayesian approach. Decis Sci 1989;20(1):77–89.
[85] Wichard JD. Forecasting the NN5 time series with hybrid models. Int J Forecast 2011;27(3):700–7.
[86] Winkler RL, Clemen RT. Sensitivity of weights in combining forecasts. Oper Res 1992;40(3):609–14.
[87] Winkler RL, Makridakis S. The combination of forecasts. J R Stat Soc 1983:150–7.
[88] Wright G, Rowe G. Group-based judgmental forecasting: an integration of extant knowledge and the development of priorities for a new research agenda. Int J Forecast 2011;27(1):1–13.
[89] Yaniv I. Group diversity and decision quality: amplification and attenuation of the framing effect. Int J Forecast 2011;27(1):41–9.