http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 63, No. 11, November 2017, pp. 3609–3627
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

A Behavioral Model of Forecasting: Naive Statistics on Mental Samples

Jordan Tong,[a] Daniel Feiler[b]

[a] Wisconsin School of Business, University of Wisconsin–Madison, Madison, Wisconsin 53703; [b] Tuck School of Business, Dartmouth College, Hanover, New Hampshire 03755

Contact: jordan.tong@wisc.edu (JT); df@dartmouth.edu (DF)

Received: September 23, 2015
Revised: March 21, 2016
Accepted: April 4, 2016
Published Online in Articles in Advance: September 28, 2016

https://doi.org/10.1287/mnsc.2016.2537

Abstract. Most operations models assume individuals make decisions based on a perfect understanding of random variables or stochastic processes. In reality, however, individuals are subject to cognitive limitations and make systematic errors. We leverage established psychology on sample naivete to model individuals' forecasting errors and biases in a way that is portable to operations models. The model has one behavioral parameter and embeds perfect rationality as a special case. We use the model to mathematically characterize point and error forecast behavior, reflecting an individual's beliefs about the mean and variance of a random variable. We then derive 10 behavioral phenomena that are inconsistent with perfect rationality assumptions but supported by existing empirical evidence. Finally, we apply the model to two operations settings, inventory management and queuing, to illustrate the model's portability and discuss its numerous predictions. For inventory management, we characterize order decisions assuming behavioral demand forecasting. The model predicts that even under automated cost optimization, one should expect a pull-to-center effect. It also predicts that this effect can be mitigated by separating point forecasting from error forecasting. For base stock models, it predicts that safety stocks are too small (large) for short (long) lead times. We also express the steady-state behavior of a queue with balking, assuming rational joining decisions but behavioral wait-time forecasts. The model predicts that joining customers tend to be disappointed in their experienced waits. Also, for long (short) lines, it predicts customers have more (less) disperse wait-time beliefs and tend to overestimate (underestimate) the true wait-time variance.

History: Accepted by Serguei Netessine, operations management.

Keywords: behavioral operations • bounded rationality • forecasting • representativeness • optimizer's curse • overconfidence • law of small numbers • newsvendor • inventory • queuing • judgment and decision making

1. Introduction

Nearly all decisions are based on forecasts, whether more intuitive or deliberative. The use of forecasts as an input to decision making is prominent in classic operations management (OM) settings. In inventory management, a newsvendor's order decision depends on her forecasted demand for the selling season. In queuing, a customer's decision to join or balk depends on his forecasted waiting time. And in process management, a manager's improvement decision depends on her forecasted capacities of each stage. Although there is an increasing use of computerized systems to support forecasting in businesses, many managerial forecasts still involve human judgment. Furthermore, customers' decisions are even more likely to be based on human judgments as opposed to computerized forecasts. Therefore, accounting for behavioral elements of forecasting is important for capturing more realistic decision-making behavior in operations and management science.

Forecasting is critical in the face of significant uncertainty; in OM contexts, such uncertainty is typically modeled using random variables or stochastic processes. The two most fundamental forecasts needed to support decision making in the face of uncertainty are the point forecast and error forecast. In OM models, these forecasts typically correspond to the mean and variance of the random variable faced by the decision maker. Therefore, in this article we focus on capturing an individual's forecasting behavior by modeling their beliefs about the mean and variance of a random variable. To the extent that researchers in other disciplines in management science and economics implement random variables on which a decision maker optimizes, our model may also be useful to them; however, this article focuses on modeling behavior for operations management.

Most traditional OM models assume, either explicitly or implicitly, that individuals have a perfect understanding of random variables and stochastic processes.

Specifically, they typically assume individuals know and make decisions based on the correct mean and variance of any random variable they face. By contrast, considerable behavioral work in the field of judgment and decision making has found that individuals have cognitive limitations and make certain systematic errors in their judgments under uncertainty. These findings highlight the potential limitations of perfectly rational models for capturing realistic behavior. From an analytical modeling standpoint, however, it can be challenging to find a way to incorporate a long list of behavioral biases in a tractable and meaningful manner. Furthermore, classic models based on the premise of perfect rationality are well developed and have yielded important insights, so it is desirable to incorporate more realistic behavior in such a way that we need not completely abandon these formulations. Therefore, there is a need for analytical models that are powerful enough to capture a variety of behavioral phenomena while still being tractable and implementable in existing models. While there have been recent advances along these lines for profit-optimization decision tasks in operations contexts (e.g., Su 2008), there has been less advancement of behavioral models for forecasting, which must precede such decision making. The goal of this paper is to take a step toward filling this gap.

To be clear, our primary objective is *not* to study a specific operations setting in-depth to derive prescriptive managerial insights for that setting. Rather, our objective is to develop a model that can be readily inserted in various existing OM models that typically assume perfect rationality such that human forecasting behavior can be better accounted for. Furthermore, our objective is *not* to present new empirical evidence of behavioral biases. Rather, we seek to show how a simple model based on established psychological principles can capture a surprisingly rich representation of forecasting behavior consistent with existing empirical evidence.

The behavioral forecasting model we propose is grounded in a psychological perspective that researchers have started to refer to as the *naive intuitive statistician* (see Fiedler and Juslin 2006 for an overview). This perspective places emphasis on the role of the environment surrounding a decision maker and the extent to which imperfect samples of information in the environment explain behavioral anomalies, even with otherwise perfect cognitive computation. In line with this perspective, our model assumes that because of limitations in one's ability to gather, process, and recall information (Simon 1955), individuals tend to think of only a small random sample of possible outcomes instead of perfectly leveraging the true random variable. They then naively operate as though the statistical properties (the mean and variance) of this small sample are perfectly representative of the properties of the true random variable. Therefore, we assume that while the individual can correctly describe the basic statistical properties of the mental sample, they fail to correct for the problems inherent in relying on small samples in the first place (Tversky and Kahneman 1971). We build most directly from Juslin et al. (2007).

The model assumptions are precise enough to allow us to characterize the individual's point and error forecasts as fully specified random variables, and we show how to do so for a variety of commonly used distributions. For example, for a normal random variable, we show that the individual's point forecast is also a normal random variable while her error forecast is a gamma random variable. (Moreover, the two are independent.) By contrast, a typical perfectly rational model assumes the point and error forecasts are constants and equal to the true mean and variance. Note also that the point and error forecasts are fully specified based on only one new behavioral parameter—the mental sample size—and the model embeds perfect rationality as a special case (as the mental sample size goes to infinity). Therefore, the model serves as a generalization of the perfectly rational model, and one can compare the predictions of the model relative to the perfect rationality benchmark by comparing behavior under a finite mental sample size to behavior with an infinite one.

After defining our behavioral forecasting model, we derive 10 distinct phenomena that it captures (see Table 1), which are not captured by a perfectly rational model, and discuss how these results relate to existing empirical evidence. Doing so serves two purposes. First, it demonstrates that we can relate a large number of behavioral phenomena (some of which are well established) to a single model. In linking several behavioral results, the model increases our understanding of how these phenomena relate to each other.[1] Second, it shows that the model can be useful for future research by capturing many departures from perfect rationality with mathematical tractability.

The model performs well with respect to several dimensions that are desirable for formal behavioral theory (Rabin 2013)—portability, plausibility, parsimony, power, and precision. It is directly portable to classic OM model settings, is grounded in credible and plausible psychology, parsimoniously uses only one behavioral parameter, is powerful in that it captures a large number of behavioral phenomena, and predicts precise differences from the perfect rationality benchmark. The model is useful for analytical modelers because it can be implemented "off the shelf" to account for a rich set of forecasting behaviors with only a single parameter where one typically would simply assume perfect rationality. Modelers can derive new managerial insights when accounting for more realistic forecasting behavior, which can complement results derived under the perfect rationality paradigm.

Table 1. Phenomena Captured by the Behavioral Model of Forecasting

Behavioral phenomenon	Corresponding behavioral model of forecasting prediction
Wisdom of the crowd	The expected value of point forecasts is the optimal point forecast.
Forecaster dispersion predicts true uncertainty and average forecaster confidence	Point forecast variance is increasing in true uncertainty; expected error forecast is increasing in point forecaster dispersion.
Underweighting rare events	Point forecasts err more frequently toward the mode than away from the mode.
Optimizer's curse	When faced with multiple random variables, the individual's belief about the mean of the random variable associated with his largest point forecast is biased high.
Jensen's inequality neglect and the planning fallacy	Individuals tend to overestimate a convex function of the mean of a random variable, causing them to overestimate how much can be completed in a given time interval.
Overconfidence	The individual underestimates the error of her own point forecast.
Weak confidence-accuracy correspondence	For normal random variables, the individual's error forecast and her point forecast are independent.
Format dependence and egocentric assessment of others' forecasts	The individual's assessment of exogenously provided point forecasts is unbiased (for normal random variables).
Gambler's fallacy and the law of small numbers	Over short time intervals, individuals tend to underestimate the error of their point forecast.
Nonbelief in the law of large numbers	Over long time intervals, individuals tend to overestimate the error of their point forecast.

Furthermore, experimentalists and empiricists can use the model to generate testable predictions, or plausible explanations for observed anomalies, across a variety of settings. The model not only hypothesizes clear differences from perfect rationality but also predicts relationships between biases and individual differences according to the behavioral parameter.

We provide two specific examples to illustrate how the model can be applied to important OM settings. First, we show that it can be applied to the classic newsvendor and base stock models by capturing the manager's demand forecasting behavior. The model is tractable enough to derive the predicted distribution of order quantities, assuming order decisions are automated based on an individual's behavioral point and error forecast inputs. Second, we investigate the classic single-server queue with balking setting and show how the model can be applied to derive the steady-state distribution of customers, assuming joining decisions are rational given the individual's behavioral wait-time forecast inputs. For both examples, even without pushing the analyses beyond interpreting the previously derived 10 phenomena, we observe several implications from the behavioral forecasting model. For example, even if cost optimization is automated, the model predicts a strong pull-to-center effect in the newsvendor problem, although it also predicts that this effect can be potentially eliminated by separating the point forecast from the error forecast tasks. It also predicts that an overordering bias can be reduced by separating the product choice decision from the order quantity decision. For base stock models, it predicts too small of safety stocks for short lead times but too large of safety stocks for long lead times. In a queue with balking, the model predicts that customers who join the queue tend to be disappointed in their experienced wait times. For long lines, it predicts that customers have more disperse beliefs about wait times and tend to overestimate the true wait-time variance. For short lines, it predicts that customers have less disperse beliefs about wait times and tend to underestimate true wait-time variance. These predictions and others are discussed in Section 5.

There are important limitations of the model. Human judgment is complex and is certainly not exhaustively accounted for by naive statistics on mental samples. Even within the context of our model, the assumptions could be further relaxed and parameterized to better fit data. We will return to these limitations and discuss opportunities for future research in the conclusion.

2. Related Literature

While our model may be useful to other disciplines in which perceptions of uncertainty are important inputs to decision making, our article contributes most directly to the field of behavioral operations management (see Croson et al. 2013, Bendoly et al. 2010, Gino and Pisano 2008 for recent reviews). There are two categories of research in behavioral operations that relate directly to this paper: (i) applications of general behavioral economics models to operations settings and (ii) forecasting experiments that uncover anomalies in operations decision making.

The use of general behavioral models in operations has thus far been primarily focused on importing features of existing models from behavioral economics into operations settings. This is often a nontrivial translation across fields, and these models have yielded important operations insights. Arguably the most commonly imported behavioral modeling feature to operations management is a utility function of preferences.

For example, operations scholars have imported utility model features from prospect theory (Kahneman and Tversky 1979a) and mental accounting (Thaler 1985) to study topics such as inventory management (e.g., Schweitzer and Cachon 2000, Nagarajan and Shechter 2013, Chen et al. 2013), supply chain contracting (e.g., Zhang et al. 2016, Becker-Peth et al. 2013), and pricing (e.g., Popescu and Wu 2007). Similarly, utility models of social preferences have been adapted from economics to generate research in supply chain management (e.g., Loch and Wu 2008, Özer et al. 2011, Katok and Pavlov 2013). Furthermore, utility models of time preferences and hyperbolic discounting in economics (Laibson 1997) have recently been implemented into operations models to study queuing (Plambeck and Wang 2013). A second behavioral modeling feature used in operations management is random decision error to account for bounded rationality in optimization. For example, following the advancements of the quantal choice models (Luce 1959, McFadden 1981, Anderson et al. 1992) and the quantal response equilibrium (McKelvey and Palfrey 1995), we began to see these general models being applied to newsvendor models (Su 2008, Kremer et al. 2010), capacity allocation models (Chen et al. 2012), and service system models (Huang et al. 2013) to generate important insights.

The majority of this work has focused on deviations from perfect rationality in decision making given perfect knowledge of random variable inputs. By contrast, less work has examined deviations from perfect rationality in the formation of beliefs that serve as the inputs for decision making. This distinction is between two fundamental challenges faced by decision makers: the need to optimize (i.e., decision making) and the need to forecast (i.e., judgment). There are a few recent behavioral operations models that have focused on the forecasting task. Croson et al. (2008) apply a model of overconfidence in demand forecasting to newsvendor order decisions. Their model assumes that individuals act on a perceived demand distribution that is a mean-preserving but variance-reduced transformation of the true demand distribution. Additionally, Huang and colleagues have studied opaque selling, capacity management, and service pricing (Huang and Yu 2014, Huang and Liu 2015, Huang and Chen 2015) by applying models of "anecdotal reasoning" developed in economics (Osborne and Rubenstein 1998; Spiegler 2006a, b, 2011). These papers relax the assumptions of rational expectations in games and instead assume that customers rely on random anecdotes (e.g., from another customer's experience) in order to make decisions. Although the authors implement a similar sampling approach, the objectives and contributions of their work are quite different from ours. These papers examine specific operations and marketing contexts assuming customers use anecdotal reasoning. Here, we focus on developing a general model, showing how naive statistics on small mental samples can capture a large number of empirically supported behavioral phenomena, which can then be implemented in a variety of settings. We also study both error and point forecasting, whereas these papers focus on point forecasting. This enables one to examine operational decisions that depend on both the mean *and* variance of the random variable.

Our article also relates to research in behavioral operations management that experimentally examines behavioral demand forecasting as a key driver of operational decision making. For example, Kremer et al. (2011) studied biases in point forecasting behavior in time-series forecasting. It is worthwhile to note that the overreaction to signals they observe in stationary demand environments is consistent with naive statistics on a mental sample of recent outcomes. Moritz et al. (2014) study the effect of individual differences in cognitive reflection and decision speed also in time-series forecasting environments. Feiler et al. (2013) studied the effect of demand censoring on point forecasting behavior. Kremer et al. (2016) studied differences between top-down and bottom-up forecasting. Also, because the newsvendor decision-making task can be decomposed into a demand forecasting task and an order decision, some behavioral newsvendor research is related. Ren and Croson (2013) tested the hypothesis that the pull-to-center effect may be due to underestimation of the demand variance. Lee and Siemsen (2017) also provided experimental evidence that overconfidence plays a role in newsvendor decisions, focusing on whether decomposing the newsvendor task into its subtasks can improve the ultimate order. Much like these scholars, we desire to bring more descriptive accuracy into operations management. However, our article differs in that we do not empirically document behavioral biases but rather focus on formalizing psychological primitives and demonstrating how these simple assumptions can unify a variety of behavioral results.

3. A Model of Behavioral Forecasting

In this section, we first present a behavioral model of the point forecast and error forecast for a single random variable. Such a random variable corresponds, for example, to the random demand in a newsvendor model. Then, we extend the model to capture forecasting for a stationary stochastic process, which corresponds, for example, to the demand process in a base stock model or the service process in a queuing model.

3.1. Naive Statistics on Mental Samples

To model forecasting behavior, we first model the belief formation upon which forecasting judgments are made. To do so, we build from the psychology literature on the naive intuitive statistician (Fiedler and Juslin 2006). Let Z be a random variable in the real numbers with distribution function F_Z, mean μ, and variance σ^2. A perfectly rational model assumes that individuals know and perfectly base their decisions on μ and σ^2. By contrast, we begin by assuming that individuals must rely only on a *small sample* of discrete random and independent outcomes from Z:

$$\mathcal{S}(Z) = \{O_i\}_{i=1,\ldots,n}, \quad O_i \sim Z. \tag{1}$$

This approach has three important behavioral features. First, the individual *forecasts by sampling*: he or she thinks in terms of distinct possible outcomes. This assumption is consistent with empirical work in cognitive psychology, which has shown that individuals tend to think in terms of discrete counts and exemplars (e.g., Gigerenzer and Hoffrage 1995, Nosofsky and Palmeri 1997). It is also consistent with the sampling approaches in the anecdotal reasoning models in economics (e.g., Osborne and Rubenstein 1998, Spiegler 2006a).

Second, the number of outcomes considered by the individual is finite or "small" (i.e., usually less than seven). Again, such an assumption is consistent with psychological theory: "Time pressure, structural limitations of the cognitive system, or paucity of available data often force people to make do with but a sample, when they try to learn the characteristics of their environment" (Kareev 2006, pg. 33). Working memory is cognition dedicated to the active processing of thought, computation, and information. Research on working memory capacity suggests that a mental sample is likely less than seven and, for complex processes, can be expected to be between two and five (Kareev 2000), but we do not take a strong position on the exact mental sample size we expect. We refer to the parameter n as the *mental sample size*. It can generally be interpreted as the degree to which cognition is bounded, with smaller sample sizes capturing more bounded cognition. As n approaches infinity, the distribution of the individual's discrete mental sample approaches the true distribution.

Third, we assume the mental outcomes considered are random draws from the true distribution Z. Thus, the probability that a value will be considered by the individual for any single mental draw is proportional to its true likelihood of occurring. Although this assumption is strong, it allows us to tractably capture a reasonable and important feature: outcomes that are more likely to occur are also more likely to be considered by the individual. It will also enable us to isolate the consequences of naive statistics on small samples without confounding them with biases as a result of drawing from incorrect distributions or nonrandom sampling.[2] One situation in which this assumption holds closely is when the individual has access to historical data of many random realizations of Z and then only recalls a subset of them (e.g., the most recent n random outcomes). When historical realizations are not available, such as is the case for judgments of novel situations, the random sampling assumption may be conceptualized as a simulation process in which the individual leverages available predictive information to simulate possible future outcomes.

Given only a small sample of random outcomes, one could apply normative statistical methods to infer the properties of the underlying distribution. By contrast, we assume that individuals apply *naive statistics* and operate as though properties of the mental sample are equal to the properties of the true random variable. Specifically, the individual naively believes the true mean μ (a constant) is exactly

$$\mu_b = \frac{1}{n}\sum_{i=1}^{n} O_i, \tag{2}$$

which is a random variable. Similarly, the individual naively believes the true variance σ^2 (a constant) is exactly

$$\sigma_b^2 = \frac{1}{n}\sum_{i=1}^{n}(\mu_b - O_i)^2, \tag{3}$$

which is a random variable. Here, the subscript b denotes the behavioral belief.

Applying naive statistics in the above manner can be thought of as a belief in the representativeness of small samples (Kahneman and Tversky 1972). The individual in our model assumes the mean and variance of her mental sample are equal to the mean and variance of the true distribution. Tversky and Kahneman (1971) found that "people view a sample randomly drawn from a population as highly representative, that is, similar to the population in all essential characteristics" (p. 105). Additional empirical examinations of the psychology of sampling have led to similar conclusions. In studying subjective confidence intervals, Juslin et al. (2007) concluded that "people tend to assume that sample properties can be directly used to estimate the corresponding population properties" (p. 678). In a study of memories of variability, Kareev (2006) concluded that "[p]eople tend to rely on sample data and do not correct for the biased values likely to be observed in small samples" (p. 34). Similarly, having reviewed a large body of psychological literature, Fiedler (2000) concluded that individuals suffer from a "lack of metacognitive devices that would be necessary . . . to correct sample statistics accordingly." He also argued that "given that even scientists [who are] specialized in sampling issues fall prey to the metacognitive weakness, everyday judgments should be even more vulnerable" (p. 660).

Naive statistics on small samples lead to two noteworthy deviations from perfectly rational beliefs. First, (μ_b, σ_b^2) are random variables because they are based on the randomly drawn mental outcomes, but the individual believes both to be equal to the true mean and variance μ and σ^2, which are constants. That is, he lacks the metacognition to account for the fact that his sample mean and variance are not necessarily equal to the true mean and variance. Put in sampling terms, he does not account for the fact that his sample mean and sample variance have sampling error. Second, note that σ_b^2 is the sample variance uncorrected for sample size. An unbiased estimator of the variance would be $(1/(n-1)) \cdot \sum_{i=1}^{n}(\mu_b - O_i)^2$. Consistent with this notion, psychologists have found evidence that individuals often fail to account for the fact that small samples tend to underestimate the variance of the population. On the basis of the results from five experiments, Kareev et al. (2002) concluded that "the variance of the actually observed sample was a better predictor of people's behavior than sample variance corrected for sample size" (p. 296). These two deviations from normative statistics will lead to several predicted forecasting biases, as we will see in subsequent analyses.

3.2. Point and Error Forecast Behavior

Now that we have modeled the individual's beliefs about random variables, we can use those beliefs to characterize forecast behavior. We begin with the most common type of forecast, the *point forecast*. Throughout this paper we assume that the point forecast seeks to minimize the expected mean squared error (MSE). This criterion incentivizes one to guess the mean of the random variable. Of course, one could use other possible criteria, such as the mean absolute deviation. However, operations decisions typically use the mean of a random variable as a key input (see Sections 5.2 and 5.1), so it is natural to focus on MSE.

Under perfect rationality, the optimal point forecast is simply the true mean,

$$\begin{aligned} x_* &= \arg\min_x E[(x-Z)^2] \\ &= \mu, \end{aligned}$$

which is a constant. By contrast, the individual believes the true mean is equal to the mean of her mental sample. Therefore, the behavioral point forecast is simply

$$\begin{aligned} X_b &= \arg\min_x \frac{1}{n}\sum_{i=1}^{n}(x-O_i)^2 \\ &= \mu_b, \end{aligned}$$

which is a random variable.

An attractive feature of formalizing behavioral forecasting with this sampling approach is that we can fully characterize the distribution of X_b by leveraging statistical theory. Below, we provide several examples.

Example 1. If Z is normally distributed with mean μ and variance σ^2, then $X_b \sim \text{Normal}(\mu, \sigma^2/n)$.

Example 2. If Z is uniformly distributed on the interval $[\alpha, \beta]$, then $X_b \sim \text{Bates}(n, \alpha, \beta)$.

Example 3. If Z is exponentially distributed with mean τ, then $X_b \sim \text{Erlang}(n, \tau/n)$.

Example 4. If Z is Bernoulli distributed with success probability p, then $nX_b \sim \text{Binomial}(n, p)$.

Example 5. If Z is Poisson distributed with parameter λ, then $nX_b \sim \text{Poisson}(n\lambda)$.

Next, we derive the behavioral *error forecast*, a measure of one's confidence in a point forecast. It is the key qualifier to the point forecast that is necessary for nearly any problem under uncertainty. Recall that the perfectly rational point forecast x_* is equal to the true mean μ. Because the mean squared error of the true mean is equal to the true variance, the perfectly rational error forecast for x_* is simply equal to the true variance,

$$\begin{aligned} \epsilon_*^2(x_*) &= E[(x_* - Z)^2] \\ &= \sigma^2 + (x_* - \mu)^2 \\ &= \sigma^2, \end{aligned}$$

which is a constant. By contrast, the individual naively believes the true mean is $X_b = \mu_b$, so she expects the error about X_b to be equal to what she believes is the variance:

$$\begin{aligned} \epsilon_b^2(X_b) &= \sigma_b^2 + (X_b - \mu_b)^2 \\ &= \sigma_b^2, \end{aligned}$$

which is a random variable.

The behavioral error forecast tends to be more cumbersome to characterize analytically than the point forecast. Still, it follows well-known distributions in some cases, such as the example below.

Example 6. If Z is normally distributed, then $\epsilon_b^2(X_b) \sim \text{Gamma}((n-1)/2, 2\sigma^2/n)$.

Observe that, in general, the behavioral error forecast is defined as a function of a point forecast and therefore may be correlated with the behavioral point forecast (e.g., see Section 4.7). Also, the error forecast can apply to *any* point forecast, not only one's own. For instance, one can also evaluate other error forecasts such as the behavioral error forecast of the optimal point forecast $(\epsilon_b^2(x_*))$ or the optimal error forecast of the behavioral point forecast $(\epsilon_*^2(X_b))$. We will leverage these quantities as points of comparison in Section 4.

3.3. Extension to Stationary Stochastic Processes

In many operations models, uncertainty is captured as a stochastic process rather than a single random variable. Our model can be extended to forecasting of stationary stochastic processes using the following approach: the random draws in the mental sample come from some natural time period, and individuals naively assume that the mean and variance of the *rate* in the mental sample are equal to the mean and variance of the true rate over any time period.

Formally, let $\mathbf{Z} = \{Z_t \mid t \in \mathbb{N}\}$ be a stationary stochastic process with independent increments, where t is the time period to be forecasted. For every Z_t, we denote its distribution function F_{Z_t}, mean $\mu(Z_t)$, and variance $\sigma^2(Z_t)$. To generalize the sampling process, we assume there is a certain time interval or *reference period length* (denoted l) of the stochastic process from which outcomes are typically recorded, experienced, and/or observed by the individual. That is, the mental samples are drawn from Z_l, and we have $\mathcal{S}(\mathbf{Z}) = \{O_i\}_{i=1,\dots,n}$, $O_i \sim Z_l$.

We apply naive statistics by assuming that the individual naively believes the mean and variance of the *rates* in the mental sample are representative of the mean and variance of the true rate over any time period. Let $\lambda_b = (1/n)\sum_{i=1}^{n} O_i/l$ be the individual's perception about the mean rate, and let $\nu_b^2 = (1/n)\sum_{i=1}^{n}(\lambda_b - O_i/l)^2$ be the individual's perception about the variance of the rate. Thus, the individual believes that $E[Z_t/t] = (1/t)\mu(Z_t)$ is λ_b for all t. Similarly, she believes that $\mathrm{Var}[Z_t/t] = (1/t^2)\mathrm{Var}[Z_t]$ is ν_b^2 for all t. By solving both of these equations for Z_t, we have that the individual believes the mean of Z_t is exactly

$$\mu_b(Z_t) = t\lambda_b \tag{4}$$

and the variance of Z_t is exactly

$$\sigma_b^2(Z_t) = t^2\nu_b^2 \tag{5}$$

for any t. Note that when $t = l$, the model reduces to the single random variable case: $\mu_b(Z_l) = l(1/n)\sum_{i=1}^{n} O_i/l = (1/n)\sum_{i=1}^{n} O_i$ and $\sigma_b^2(Z_l) = l^2(1/n)\cdot\sum_{i=1}^{n}(\lambda_b - O_i/l)^2 = (1/n)\sum_{i=1}^{n}(\mu_b(Z_l) - O_i)^2$. Otherwise, the individual extrapolates or interpolates from her mental sample using her perception of the rate.

It may be helpful to consider the following example. A company's demand every *day* is an independent and identically distributed (i.i.d.) random variable. Employees always record and report sales outcomes by the *week*. Therefore, $\mathbf{Z}$ is the daily demand process, and the reference period length is $l = 7$ days. To form beliefs about demand in a *year* (Z_{365}), the individual thinks of n outcomes of weekly demand (Z_7). She then extrapolates to a year by assuming the mean and variance of the demand rate in a year is the same as the mean and variance of the demand rate in her mental sample. If, instead, she were assessing demand in a week, she would not need to make such an extrapolation because her mental outcomes would already be in the appropriate time lengths.

Relative to the single random variable case, the stochastic process case presents an additional deviation from normative statistics: irrespective of the magnitude of t, the individual naively assumes that the variance of the rate in her sample is representative. Kahneman and Tversky (1972) presented experimental evidence consistent with this notion. Participants were told that the probability of success for a single random draw was 50%. They then considered either 10, 100, or 1,000 draws with replacement and were asked to report a likelihood distribution for different possible proportions of successes (10%, 20%, etc.). Participants generated nearly identical distributions for the proportion of successes, irrespective of the number of draws. Kahneman and Tversky (1972) referred to this as the "universal sampling distribution," and it is consistent with the insensitivity to t in the model. This insensitivity to the time horizon leads to important biases, as we will see in Sections 4.9 and 4.10.

Now that we have expressions for $\mu_b(Z_t)$ and $\sigma_b^2(Z_t)$, the extension to point forecast behavior follows the same structure as in the single random variable case. The perfectly rational point forecast for the random variable Z_t is

$$x_*(Z_t) = \mu(Z_t). \tag{6}$$

By contrast, the behavioral point forecast for Z_t is

$$\begin{aligned} X_b(Z_t) &= \mu_b(Z_t) \\ &= t\lambda_b. \end{aligned} \tag{7}$$

The normative error forecast for the normative point forecast for Z_t is

$$\epsilon_*^2(x_*(Z_t)) = \sigma^2(Z_t). \tag{8}$$

By contrast, the individual believes the error of her own behavioral point forecast for Z_t is

$$\begin{aligned} \epsilon_b^2(X_b(Z_t)) &= \sigma_b^2(Z_t) \\ &= t^2\nu_b^2, \end{aligned} \tag{9}$$

where the second equality follows from (5).

4. Behavioral Phenomena Captured by the Model

Recall that we define the perfect rationality model in this context as one that assumes that decision makers have a perfect understanding of the properties of the random variables or stochastic processes that they face. We now show that the behavioral model captures several distinct behavioral phenomena that are not captured by such a model of perfect rationality.

4.1. Wisdom of the Crowd

The behavioral model relaxes the perfectly rational benchmark by capturing random forecasting behavior. Thus, it captures the reality that even given the same information, people may not always report identical point forecasts equal to the optimal value.

A well-documented effect involving human forecasting is the "wisdom of the crowd" (Surowiecki 2005). This phenomenon states that the average of many people's point forecasts tends to be very accurate, more accurate than most individuals' point forecasts (e.g., Armstrong 2001, Clemen 1989, Einhorn et al. 1977, Larrick and Soll 2006). Even though, given the same information, people do not always report identical point forecasts, the average of their point forecasts tends to be quite accurate.

Our model captures this well-documented phenomenon in that although point forecasts are random, the *expected* point forecast is optimal.

Proposition 1. *The individual's point forecast is equal to the optimal forecast in expectation* $E[X_b] = x_*$.

Of course, even for the optimal point forecast, the forecast error is not zero because Z is still random. However, in general, the expected point forecast will outperform an individual with finite n in the long run. What drives the wisdom of the crowd effect here is the sampling mechanism: people make their forecasts based on random pieces of relevant information such that those people who rely on evidence for high values tend to be canceled out by others who randomly rely on evidence for low values. Therefore, averaging imperfect point forecasts both reduces variability and generally improves accuracy.

4.2. Forecaster Dispersion Predicts True Uncertainty and Average Confidence

Forecaster dispersion is the extent to which there is variance or dispersion in point forecasts across individuals. Another key feature of the model's sampling formulation is that the degree of point forecast dispersion is *endogenous*: it depends on the parameters of the model. By contrast, the perfect rational model does not predict any relationships with point forecast dispersion because it does not predict any dispersion to begin with.

Empirically, a strong relationship between forecaster dispersion and true uncertainty has been documented in demand forecasting in practice (Fisher and Raman 1996, 2010; Gaur et al. 2007). In a popular operations management case study (Hammond et al. 1994), the retailer Sport Obermeyer made demand point forecasts by averaging the point forecasts of seven internal experts who all had access to the same information. They then obtained a good predictor of their overall point forecast error by multiplying the experts' point forecast dispersion by 1.75. In this manner, they converted point forecast dispersion into a proxy of true uncertainty.

A relationship between forecaster dispersion and average forecaster confidence has also been documented empirically (Zarnowitz and Lambros 1987). More dispersion across forecasts tends to correspond with less confidence among those forecasters. The importance of this relationship stems from the fact that there are many situations in which only point forecast data are accessible, but one would also like to estimate what forecasters think about uncertainty.

Consistent with these empirical observations, our model captures the following relationships between forecaster dispersion, true uncertainty, and average forecaster confidence.

Proposition 2. *The individual's point forecast has expected dispersion* $\mathrm{Var}[X_b] = \sigma^2/n$. *The expected error forecast is proportional to the expected dispersion,* $E[\epsilon_b^2(X_b)] = (n-1)\cdot \mathrm{Var}[X_b]$.

Thus, the larger the uncertainty in the environment, the more dispersion we should expect in point forecasts, and vice versa. The second part of the proposition notes that because both error forecasts and point forecast dispersion are functions of true uncertainty, point forecast dispersion also predicts the average confidence level of the population. Although these results are straightforward, they are inconsistent with the perfectly rational model that captures no dispersion. They are also intuitively appealing: we would expect some dispersion and expect more dispersion across point forecasts when there is more uncertainty. For example, if Z is determined by the sum of two fair six-sided dice, we would not necessarily expect all individuals to make the perfectly rational point forecast of seven. And we would expect more point forecast dispersion (and forecasters to be less confident about their point forecast accuracy in general) if the dice were 47-sided as opposed to 6-sided.

Note that the model also predicts that the slope of these relationships depends on the mental sample size n. A given level of forecaster dispersion indicates larger true uncertainty (and less average confidence) when n is large compared with when n is small. In this way, the model provides an interpretation of the constant used at Sport Obermeyer, mentioned above. For example, if forecasters are under more time pressure (captured by a smaller n), it suggests that Sport Obermeyer should expect the constant to decrease below 1.75. On the other hand, as forecasters become more sophisticated (a larger n), even small levels of disagreement among the forecasters can indicate large true demand uncertainties, and Sport Obermeyer should multiply by a larger constant.

4.3. Underweighting Rare Events

The model predicts endogeneity not only in the degree of forecaster dispersion but also in the *shape* of dispersion. Because of the central limit theorem, X_b approaches a normal distribution as n grows large. However, in our model, n is small such that X_b may deviate significantly from normality. In particular, individuals are unlikely to consider rare events when thinking of their mental sample of possible outcomes. Thus, most point forecasts trend toward the outcomes that are most likely. Consequently, if Z is skewed, then the distribution of point forecasts will also be skewed in the same direction. We state it formally as follows.

Proposition 3. *Inequality* $\mathrm{Mode}(X_b) < x_*$ *holds if and only if* $\mathrm{Mode}(Z) < x_*$. *If* Z *belongs to the Pearson family, then* $\mathrm{Mode}(X_b) < \mathrm{Median}(X_b) < x_*$ *if* Z *is positively skewed, but* $\mathrm{Mode}(X_b) > \mathrm{Median}(X_b) > x_*$ *if* Z *is negatively skewed.*

In Proposition 2 we showed that the expected point forecast is equal to the true mean. How can these two propositions be reconciled? Although individuals are more likely to err away from rare events, those that err *toward* rare events tend to do so much more severely. Consider $Z \sim \mathrm{Bernoulli}(0.01)$; there is a 1% probability of a disaster, and the optimal point forecast is $x_* = 0.01$. By contrast, under our behavioral model with $n = 5$, the point forecasts have the distribution $5X_b \sim \mathrm{Binomial}(5, 0.01)$. Consequently, the model predicts that approximately 95% of individuals will act as though there is zero probability of a disaster. However, about 5% of individuals will act as though there is about 0.2 probability. And about 0.1% of individuals will act as though there is a 0.4 probability of a disaster. In short, the model captures a phenomenon where the majority of individuals underestimate the likelihood of rare events, but a minority of individuals greatly overestimate it.

This pattern of behavior is consistent with empirical work in psychology on decisions from experience. When sampling alternatives before making a risky choice, individuals tend to make strong inferences from their small samples of experience such that the majority underweights rare events, but a minority overweights them (Hertwig et al. 2004, Rakow et al. 2008, Hadar and Fox 2009). Hadar and Fox (2009, p. 324) related their results to the following example:

> Before taking a long trip a driver may seem to "underweight" and/or "underestimate" the possibility of a tire blowout by failing to check tire wear and inflation because the possibility of this outcome never occurs to him. However, if the driver has experienced (personally or vicariously) a blowout or is reminded about this possibility by a companion then he may "overweight" and/or "overestimate" this outcome, going to great lengths to avoid a low-probability catastrophe (blowout).

Returning to the wisdom of the crowd effect in Section 4.1, this pattern of point forecast dispersion also implies that one should not necessarily eliminate outliers when averaging point forecasts (Larrick and Soll 2006, Soll and Mannes 2011) because the large errors of the minority may help offset the small errors of the majority. If Z is positively (negatively) skewed, removing the point forecasts of the outliers will tend to bias the resulting average downward (upward). In this way, the model captures a distribution of perceptions of rare events in a manner consistent with empirical findings in psychology.

4.4. Optimizer's Curse

The belief that the mean of one's mental sample is equal to the true mean of the random variable is especially problematic when optimizing over several random variables. Imagine that a manager must choose one product from a set of alternatives and then decide how many units of that product to produce. He makes a point forecast for each alternative's demand, selects the product associated with his highest point forecast, and produces an amount equal to that point forecast.

Interestingly, this manager has likely just produced too many units. If each point forecast has random error, then the largest point forecast tends to be *larger* than its true mean value because of what is known as the optimizer's curse (Smith and Winkler 2006). In the process of choosing the maximum point forecast, one is more likely to choose a point forecast that had positive error than one that had negative error. To account for this problem, the maximal forecast should be adjusted downward and more so when point forecasts have more random error, and when the true means are close together (Smith and Winkler 2006, Harrison and March 1984). However, Tong et al. (2016) demonstrate experimentally that individuals generally fail to make such an adjustment, resulting in an overestimation bias (see also Kahneman and Tversky 1973 and Thaler 1988 for similar work on regression-to-the-mean effects and the winner's curse).

Our behavioral model captures the optimizer's curse as follows. Because of naive statistics, individuals incorrectly operate as though the mean of one's mental sample is identical to the mean of the random variable, even though it has random error. If an individual assumes that his point forecasts for several random variables are equal to their respective means, then there is no reason for him to believe that the forecast for the random variable with the highest predicted outcome is systematically biased. We state this formally as follows.

Proposition 4. *Let* $Z_1, Z_2, \ldots, Z_k$ *be* k *independent random variables with true means* $\mu_1, \mu_2, \ldots, \mu_k$. *Let* i^* *denote the index associated with the individual's maximal point forecast* $X_{b,i^*} = \max\{X_{b,1}, X_{b,2}, \ldots, X_{b,k}\}$. *Then, the individual's belief about* μ_{i^*} *is biased high.*

It is important to observe that the above biased belief about the chosen alternative is only for the *same* individual that made the choice. From Proposition 1, an individual is unbiased if he does not first choose among a set. This combination of results suggests that the optimizer's curse may be reduced if the selection task and forecasting task are given to two separate and independent people.

The optimizer's curse has important consequences in numerous operations settings beyond product designs and production decisions. For example, imagine that a manager conducts process improvement by making capacity forecasts under various designs and then choosing the one that is forecasted to yield the greatest improvement. The optimizer's curse suggests that she will overestimate the expected benefit of the chosen design, potentially yielding an inflated willingness to pay for it or affecting future budgetary and logistical planning. Similarly, imagine that a customer decides which product to purchase based on his quality forecasts for multiple products. The optimizer's curse suggests that he will be disappointed in the quality of the purchased product, on average. We discuss further applications in Section 5.

4.5. Jensen's Inequality Neglect and the Planning Fallacy

As previously noted, an important mechanism in the behavioral model is that the forecaster naively believes that her point forecast (which has random error) is exactly equal to the true mean. In addition to neglecting the optimizer's curse, this naivety makes the individual subject to what we call *Jensen's inequality neglect*. Because the individual treats X_b as the true mean, he applies functions to it as one would to the true mean. Mathematically speaking, he thinks that $g(X_b) = g(\mu)$ for any function g. This belief is problematic because X_b is a random variable while μ is a constant. Moreover, by Jensen's inequality, we know that $E[g(X_b)] \geq g(\mu)$. The individual does not take this inequality into account. Therefore, we have the following.

Proposition 5. *Let $g(\cdot)$ be a convex function. The individual's belief about $g(\mu)$ is biased high.*

To our knowledge, no empirical work in psychology or management has directly examined Jensen's inequality neglect. However, there is some evidence that even trained statisticians fail to correct for Jensen's inequality when conducting estimations (Silva and Tenreyro 2006), so it is reasonable to hypothesize that customers and managers will also fail to do so.

An important example of a convex function that may impact common decisions is the reciprocal.

Corollary 1. *The individual's belief about $1/\mu$ is biased high.*

Suppose a manager is asked how many projects he can complete in six years. If he bases his estimate on how long he thinks it takes to complete each project on average, the above corollary suggests that he tends to *overestimate* how many projects he can complete. The intuition is that misestimation gets magnified more when the manager underestimates how long it takes to complete a project than when he overestimates it. For example, suppose that in actuality a project takes one year on average so that, in expectation, he can complete six projects in six years. If he overestimates the average time needed to complete a project by six months, then he will underestimate his six-year productivity by two projects (6 years/(1.5 projects/year) = 4 projects). However, if he *underestimates* the average time needed to complete a project by six months, then he will overestimate his six-year productivity by six projects (6 years/(0.5 projects/year) = 12 projects). Therefore, he tends to overestimate his six-year productivity overall.

Corollary 1 can be interpreted as a type of *planning fallacy*. Traditionally, the planning fallacy refers to the empirical observation that individuals tend to underestimate how long projects will take to complete (e.g., Buehler et al. 1994). Kahneman and Tversky (1979b) suggested that the planning fallacy occurs because people focus on how the components of a project can be successfully coordinated and completed, underappreciating the combined impact of the many ways in which a plan can go awry. The planning fallacy may also emerge as a consequence of failing to unpack all of the individual steps that are required to complete a complex project (Kruger and Evans 2004). Our model generates an alternative source of the planning fallacy: neglecting errors in the forecasted cycle time—the time needed to complete one unit of work—leads to overestimation of the work completion rate, even if cycle time estimates are unbiased.

4.6. Overconfidence

Perhaps the most well-known empirical result related to the error forecast is overconfidence in the accuracy of one's point forecast, sometimes more specifically referred to as overprecision (Moore and Healy 2008). Empirical evidence supporting such overconfidence has typically employed the task of having individuals provide two numbers such that they are 90% sure the answer will lie between them. Surveying the overconfidence literature, Jain et al. (2013, p. 1970) found that self-reported 90% confidence intervals "are likely to capture much less than 90% of the actual realizations, often only 40% to 70% of the realizations." There is a clear tendency of individuals to provide overly narrow confidence intervals.

Of course, the perfectly rational model predicts no overconfidence; under perfect rationality, the point forecast equals the true mean and the error forecast is equal to the true variance. By contrast, our behavioral model generates such overconfidence. To see this, note that, on average, the behavioral error forecast can be expressed as

$$E[\epsilon_b^2(X_b)] = (n-1)\sigma^2/n.$$

It is smaller than the true variance by a factor $(n-1)/n$. Next, let $\epsilon_*^2(X_b)$ denote the *true* expected error of the behavioral point forecast. It can be evaluated as

$$\begin{aligned} E[\epsilon_*^2(X_b)] &= E[(X_b - Z)^2] \\ &= \sigma^2/n + \sigma^2 \\ &= \frac{n+1}{n}\sigma^2 \end{aligned}$$

and is larger than the true variance by a factor $(n+1)/n$. Combining these two observations yields the following proposition.

Proposition 6. *The individual underestimates the error of his point forecast for the random variable Z by a factor of $(n+1)/(n-1)$. That is, $E[\epsilon_b^2(X_b)] = ((n-1)/(n+1))E[\epsilon_*^2(X_b)]$.*

The magnitude of the above underestimation can be significant. If $n = 3$, the individual thinks his point forecast error will be only 1/2 of its true expected value. Put another way, he underestimates the root mean squared error by a factor $\sqrt{0.5}$, which suggests that an individual's 90% confidence interval will capture the true outcome only about 75% of the time.

The model captures two separate drivers of overconfidence. First, the individual fails to account for his own random point forecast error (recall that $\mathrm{Var}(X_b) = \sigma^2/n > 0$) and instead naively assumes his own point forecast is equal to the true mean.[3] Second, he tends to underestimate the variance of Z because small samples have lower variances than their populations (recall that $\sigma_b^2 = (n-1)\sigma^2/n < \sigma^2$). The manner in which our model captures overoptimism in the accuracy of one's point forecast is similar to the work of Juslin et al. (2007), which demonstrated that sampling and statistical naivete can lead to overly narrow confidence intervals even in the absence of any more pernicious bias, such as confirmatory information search.

4.7. Weak Confidence–Accuracy Correspondence

We have shown that the model captures overconfidence: individuals tend to underestimate their own point forecast error. However, are more confident individuals more accurate? From a managerial perspective, one would hope that individuals displaying more confidence are in fact more accurate. However, empirical work in psychology has often found a surprisingly weak correspondence between confidence and point forecast accuracy (Henry 1993, Sniezek and Henry 1989, Tsai et al. 2008).

Interestingly, the behavioral model also captures a surprisingly weak correlation between an individual's point forecast and error forecast. In fact, if Z is normally distributed, the individual's point and error forecasts are independent. Consequently, the individual's error forecast and her true expected error are also independent. We state this formally below.

Proposition 7. *If Z is normally distributed, then $\epsilon_b^2(X_b)$ and X_b are independent. Moreover, $\epsilon_b^2(X_b)$ and $\epsilon_*^2(X_b)$ are independent.*

If Z is normally distributed, the model states that a confident person is no more likely to be accurate than an unconfident person. At first glance, this result may appear surprising. After all, the error forecast is defined as a function of the point forecast. Also, one might expect that an individual who has a terribly erroneous point forecast should also have very low confidence. However, the result is a consequence of a statistical fact of normal distributions: the sample mean and sample variance of a normal distribution are independent. By connecting this statistical fact with forecasting behavior, the model provides a plausible mechanism for the empirically observed weak correspondence between confidence and accuracy.

There may, of course, exist other factors that make point and error forecast accuracy positively correlated. For example, a subset of individuals may face a random variable with a smaller true uncertainty (i.e., smaller σ^2), which would generate a positive correlation between confidence and accuracy. Furthermore, when Z is nonnormally distributed, the point and error forecast are not independent; if Z is positively (negatively) skewed, then X_b and $\epsilon_b^2(X_b)$ are positively (negatively) correlated.[4] Proposition 7 does not preclude these possibilities, but it does show why under certain conditions confidence and accuracy may not correspond to the degree one might expect.

Finally, we note that the independence of the behavioral point and error forecasts for normally distributed Z is very useful for analytical tractability. Indeed, we will leverage it later in this section and in Section 5.

4.8. Format Dependence and Egocentric Assessment of Others' Forecasts

Although the behavioral model predicts that individuals will be overconfident in their *own* point forecast, surprisingly, it predicts that overconfidence may be eliminated when assessing an *exogenous* point forecast.

Proposition 8. *Let Z be normally distributed and y be an exogenous point forecast. The individual's error forecast for y is unbiased; $E[\epsilon_b^2(y)] = \epsilon_*^2(y)$.*

The intuition is as follows. Although the individual underestimates σ^2, such underestimation is perfectly canceled out by the added randomness in his belief about μ, which he also ignores. By making the point forecast exogenous, we have flipped one of the drivers of overconfidence (incorrect beliefs about μ) into a driver of underconfidence. This perfect cancellation relies on the result from Proposition 7 that $X_b = \mu_b$ and $\epsilon_b^2(X_b) = \sigma_b^2$ are independent when Z is normally distributed. If Z is not normally distributed, then the two effects may not perfectly cancel out, although overconfidence should still generally be smaller for exogenous versus endogenous point forecasts.

Overconfidence in the accuracy of one's own forecast but good confidence calibration for an exogenous point forecast may seem like a strange pattern of results. However, there exists empirical support for exactly this pattern. As mentioned previously, when employing a confidence interval generation task (e.g., "Set an upper and lower bound such that there is an 90% chance that the correct answer falls in that range"), experiments have consistently revealed that individuals provide intervals that are too narrow (see Section 4.6). However, when employing a probability estimation task for exogenously provided intervals (e.g., "What is the probability that the correct answer is between 1,700 and 1,800?"), experiments have found overconfidence is greatly reduced (Gigerenzer et al. 1991, Hansson et al. 2008, Klayman et al. 1999, Haran et al. 2010). This phenomenon is referred to by psychologists as *format dependence* because the likelihood of observing overconfidence depends on the format of the question (Juslin et al. 1999). Our model provides a sampling-based explanation consistent with that proposed by Juslin et al. (2007) but in the context of point and error forecasting.

Finally, by comparing this result with the overconfidence in one's own point forecast (see Proposition 6), note that the model captures a systematically higher confidence in one's own forecasts than others' forecasts, even when unwarranted. Such egocentric evaluation of others' point forecasts is consistent with evidence of egocentric advice discounting (e.g., see Yaniv and Kleinberger 2000), in which individuals do not adjust their own forecasts sufficiently in response to observing another person's independent forecast.

4.9. The Gambler's Fallacy and the Law of Small Numbers

The next two phenomena concern individuals' beliefs about stochastic processes and therefore require our extension to stationary stochastic processes.

A well-known behavioral forecasting bias for stochastic processes is the gambler's fallacy, which is the tendency to believe that if an event occurred less frequently than its theoretical probability in the past, it will occur with higher probability in the future, even if the process is truly random and memoryless. It can also be described as a belief in local balancing within random sequences. For example, people tend to believe that a fair coin flip will be significantly more likely to be tails after observing several heads in a row than after observing several tails in a row (Rappoport and Budescu 1992, 1997). The result is consistent with the "law of small numbers," which, generally speaking, says that individuals exaggerate how likely it is that a small number of random outcomes will have the same characteristics as the true distribution from which they were drawn (Tversky and Kahneman 1971, Rabin 2002). If one believes that the true rate should be achieved in any short time interval, one will expect deviations in one direction to soon be canceled out by deviations in the other (Rabin and Vayanos 2010).

The behavioral model captures a bias consistent with the gambler's fallacy and a belief in the law of small numbers by way of an underestimation of point forecast error in the short run. Recall that Z_t is the random variable of interest. When t is small (relative to l), the behavioral model predicts that the individual is overconfident in the accuracy of her point forecast. Formally, we have the following.

Proposition 9. *Let* $\mathbf{Z}$ *be a stationary stochastic process with independent increments. The individual underestimates her point forecast error for* $t < ln/(n-2)$. *This bias is decreasing in* t.

The main driving force behind why the individual underestimates her point forecast error for small t is that she believes that highs and lows in the stochastic process should average out with as much force in small time t as it does in the reference period length l. For example, imagine that a manager typically observes monthly demand ($l = 30$ days) but must forecast demand for the next 10 days. The model predicts she expects the demand rate in 10 days to only vary as much as the demand rate varies in a month, even though the variance of the demand rate in 10 days is likely to be much larger.

Notice that when t is equal to l, the individual is still overconfident. The overconfidence in this case reduces to the overconfidence in the single random variable case, discussed in Section 4.6.

4.10. The Nonbelief in the Law of Large Numbers

The previous result concerns beliefs about stochastic processes in the short run. What pattern emerges regarding beliefs about stochastic processes in the long run?

Past work has found evidence of a "nonbelief in the law of large numbers" (Benjamin et al. 2016): people have a tendency to falsely believe that the characteristics of even very large random samples may still deviate from the population's true characteristics. For example, for 1,000 fair coin flips, what is the chance that the number of heads will fall within the range of 450 to 550? In Kahneman and Tversky (1972), participants assigned a probability of 0.21 when answering this question, even though the true probability is greater than 0.99.

The behavioral model captures a bias consistent with a nonbelief in the law of large numbers by way of an overestimation of point forecast error in the long run. When t is large (relative to l), the individual is underconfident in the accuracy of her point forecast. Specifically, we have the following.

Proposition 10. *Let* $\mathbf{Z}$ *be a stationary stochastic process with independent increments. The individual overestimates her point forecast error for* $t > ln/(n-2)$. *This bias is increasing in* t.

The mechanism driving this result is similar to that in Proposition 9: the individual does not believe that highs and lows will average out any more than they do in the reference time length, even for very large time intervals. Of course, by the actual law of large numbers, we know that the average observed rate will indeed equal the true expected rate as $t \to \infty$. Therefore, for large enough values of t, this false belief in the representativeness of the rate in her mental sample is enough to overcome the overconfidence result in Section 4.6, resulting in underconfidence. Therefore, when forecasting for a sufficiently long time horizon, our model predicts overestimation of the point forecast error.

5. Illustrative Examples

We now provide two examples of how one can apply the behavioral model of forecasting to OM settings—the newsvendor model and a single-server queue with balking model. We will show that the behavioral forecasting model can be tractably imported to derive the distribution of inventory order decisions (assuming behavioral demand forecasting but rational cost minimization) and the steady-state distribution of customers in the queue (assuming behavioral wait time forecasting but rational utility maximization). We then discuss the direct implications of the previously derived 10 behavioral phenomena for these settings. To be clear, our purpose is not to derive optimal policies nor performance implications, but rather to show that the model opens the door to studies of this kind. Similarly, our purpose is not to provide evidence for the implications of the model in each setting, but rather to show that the model can generate interesting and testable predictions when applied to OM settings.

5.1. Demand Forecasting in the Newsvendor Model

5.1.1. Relaxing Perfect Rationality in Forecasting. The newsvendor model is the fundamental building block of stochastic inventory management. In it, a manager must determine an inventory order size q in advance of a random demand D. We will assume that D is normally distributed with distribution function F_D, mean μ, and variance σ^2. For every unit the manager's q falls short of the realized demand, he incurs an underage cost c_u. For every unit he exceeds demand, he incurs an overage cost c_o. Thus, his inventory cost is

$$C(q) = c_o[q-D]^+ + c_u[D-q]^+, \tag{10}$$

where $[x]^+ = \max\{x, 0\}$. The optimal inventory level q^* minimizes the expected inventory cost. It is straightforward to show that q^* satisfies

$$F_D(q^*) = \frac{c_u}{c_u + c_o}.$$

The right-hand side is called the "critical fractile." The left-hand side is called the "in-stock probability." Therefore, the optimal order quantity achieves a probability of being in stock equal to the critical fractile.

Let $z_* = \Phi^{-1}(c_u/(c_u + c_o))$, where Φ^{-1} is the inverse of the standard normal distribution function. It is well known that for normally distributed D, the optimal inventory level can be written as

$$\begin{aligned} q^* &= \mu + \sigma z_* \\ &= x_* + \epsilon_*(x_*) z_*, \end{aligned}$$

(e.g., see Zipkin 2000). The first term of the optimal inventory level is often referred to as the "cycle stock" because one expects it to be used based on expected demand. The second term is often referred to as the "safety stock" because while one does not expect to use it, demand is uncertain, and it hedges against the asymmetric costs of underage and overage costs. Standard models assume that the individual's decision is based on a perfectly rational forecast—the individual uses the true demand mean and variance in the formula above to determine the order quantity.

By contrast, we can incorporate our model of behavioral forecasting to relax the perfect rationality assumption in the newsvendor model. Instead of basing the order on the true demand mean and variance, the individual bases it on her point forecast and error forecasts. Therefore, the resulting inventory decision based on the behavioral forecast is

$$Q_b = X_b + \epsilon_b(X_b) z_*, \tag{11}$$

which, in contrast with q^*, is a random variable. Because demand is normally distributed, from Examples 1 and 6, we know that $X_b \sim \text{Normal}(\mu, \sigma^2/n)$ and $\epsilon_b^2(X_b) \sim \text{Gamma}((n-1)/2, 2\sigma^2/n)$. Moreover, from

Proposition 7, we have that X_b and $\epsilon_b^2(X_b)$ are independent random variables. Therefore, Q_b is a fully defined random variable that we can express using the convolution of the probability density functions of X_b and $\epsilon(X_b)$.

5.1.2. Direct Implications of the Behavioral Phenomena. Given that the newsvendor order decision serves as a building block to many operations models, one could leverage the behavioral forecasting model to conduct an in-depth study on a variety of topics using the same methods as with the perfectly rational model. However, even without conducting additional analyses (solving for optimal policies, determining contract performances, etc.), we can generate implications for inventory management by interpreting the 10 previously derived behavioral phenomena.

From overconfidence (see Section 4.6), the model predicts that order quantities will be biased toward the demand mean even under automated ordering decisions. This pattern is consistent with the well-documented pull-to-center effect (Schweitzer and Cachon 2000, Bolton and Katok 2008) and laboratory evidence suggesting that the effect is at least in part due to demand overconfidence (Ren and Croson 2013) and the order-to-demand framing of the newsvendor problem (Kremer et al. 2010).

Furthermore, recall that there were two components driving the overconfidence bias according to our model: an underestimation of the true variance and an assumption that one's point forecast is centered on the true underlying mean. Therefore, Proposition 6 suggests that in order to achieve an in-stock probability level equal to the critical fractile, one must increase individuals' safety stocks by a factor $\sqrt{(n+1)/(n-1)}$, even though the average safety stock in Q_b is only less than the safety stock in q^* by a factor $\sqrt{n/(n-1)}$. In other words, increasing safety stocks to account for behavioral underestimation of variance observed in the laboratory is not enough—one must increase safety stocks even further to account for behavioral naivety with point forecast error.

Newsvendor orders in the laboratory exhibit random error (Su 2008), which sometimes is even more costly than the pull-to-center effect (Rudi and Drake 2014). The model relaxes perfectly rational orders by capturing random error generated by behavioral forecasting. Specifically, the model predicts that we should expect to see larger dispersion in order quantities for (1) products with high true demand variance, (2) products with very high or very low profit margins (consistent with Chen et al. 2013), and (3) people with larger pull-to-center effects (consistent with Moritz et al. 2013). To see this, note that endogenous dispersion (see Section 4.2) implies that the dispersion of the cycle stock is increasing in the variance of demand. The dispersion of the safety stock is also increasing in the demand variance (see Example 6). Because these dispersions are independent (Proposition 4.7), we have point (1). Point (2) follows from the fact that $\epsilon_b(X_b)$ is multiplied by z^* in the formula for Q_b. And point (3) follows from the fact that both overconfidence and dispersion are decreasing in n.

However, based on format dependence (see Section 4.8), if one person makes a point forecast and a different person takes that point forecast as exogenous to make an error forecast, the resulting error forecast should be unbiased. Therefore, applying Proposition 8, the model predicts that if we decomposed the order decision such that the cycle stock was determined by one person's point forecast and the safety stock was determined by another person's error forecast of the first person's point forecast, the resulting order quantity should achieve an in-stock probability closer to the critical fractile.

Similarly, the model also predicts that it should be advantageous to separate product selection decisions from inventory decisions. The optimizer's curse (see Section 4.4) suggests that if a manager is placing an order for a product because it is associated with his highest forecasted demand among several products, then his cycle stock decision will be biased high. However, by separating the product selection decision and the inventory order decision, firms may be able to reduce this bias.

Finally, Propositions 9 and 10 regarding the law of small numbers and the nonbelief in the law of large numbers (see Sections 4.9 and 4.10) suggest that point forecast error is significantly underestimated for short time intervals but significantly overestimated for long time intervals. Substituting these results into the expression for Q_b, we see that the implications of these findings for inventory decision making in base stock models (or other inventory control models where lead-time demand plays an important role) are that individuals will have too little safety stock for short lead times but too much safety stock for long lead times.

5.2. Wait-Time Forecasting in an Observable Queue

5.2.1. Relaxing Perfect Rationality in Forecasting. Consider the classic model setting of the single-server queue with balking, such as the one considered in Naor (1969). Customers arrive to a server according to a Poisson process with rate λ customers per minute. Service times are i.i.d. and exponential with expected time τ minutes. A customer receives benefit r from completed service, but waiting (both in line and while receiving service) is costly at rate c. Customers are served on a first-come, first-served basis. Customers are homogeneous with $r \geq c\tau$.

Upon arriving to the system and observing the number of customers in line, the customer forecasts the wait time based on the queue length. From this forecast, she decides whether or not to join the queue. Let W_{k+1} be the true waiting plus service time for a customer who arrives when there are k people in the system (both in line and in service). A customer would like to join the line if $cW_{k+1} \leq r$ (i.e., if the reward for joining is greater than the cost of waiting). Of course, at the time he must decide whether or not to join, W_{k+1} is uncertain, and he must make a decision based on his point forecast.

Naor (1969) models all customers as making perfectly rational forecasts and maximizing expected net rewards. Thus, the join or balk decision is determined by the true expected wait time $E[W_{k+1}]$. Upon observing k customers in line ahead of him, a customer joins the queue if and only if

$$c(k+1)\tau \leq r. \tag{12}$$

The resulting joining process is Poisson with rate λ if $k+1 \leq r/(\tau c)$ and zero otherwise. The steady-state queuing behavior is equivalent to a finite-capacity queue model with a capacity of $\bar{k} = \lfloor r/(c\tau) \rfloor$. That is, it is an $M/M/1/\bar{k}$ system. The steady-state probabilities are well known (see, for example, Kulkarni 2009):

$$P_k = \frac{1-\rho}{1-\rho^{\bar{k}+1}}\rho^k, \quad k = 0, 1, \ldots, \bar{k},$$

where $\rho = \lambda\tau$.

Our behavioral model enables us to relax Naor's model (Naor 1969) to capture behavioral aspects of forecasting by capturing join or balk decisions that are based on individuals' behavioral point forecasts, which are not always equal to the true expected wait time. Let $X_b(W_{k+1})$ be the behavioral point forecast for W_{k+1}. Then, upon observing k individuals in the queue, a customer joins the queue if and only if

$$cX_b(W_{k+1}) \leq r.$$

Because each service time is exponential with mean τ, we know that $W_{k+1} \sim \text{Erlang}(k+1, \tau)$. By contrast, by applying the behavioral forecasting model in Section 3.3, it is straightforward to show that $X_b(W_{k+1}) \sim (k+1)\tau_b$, where $\tau_b \sim \text{Erlang}(nl, \tau/(nl))$ with the distribution function denoted by $F_{\tau_b}(\cdot)$. Here, τ_b denotes the individual's belief about the average time per customer, and l denotes the reference number of customers. The individual will join the queue if

$$c(k+1)\tau_b \leq r, \tag{13}$$

which occurs with probability $F_b(r/(c(k+1)))$. Thus, the joining process when there are k customers in line is Poisson with rate

$$\lambda_k = \lambda F_{\tau_b}\left(\frac{r}{c(k+1)}\right),$$

and the resulting steady-state probabilities can be shown to be

$$P_k = A_k P_0, \quad k \geq 1,$$
$$P_0 = \frac{1}{\sum_{i=0}^{\infty} A_i},$$

where

$$A_k = \frac{\lambda}{\tau} F_{\tau_b}\left(\frac{r}{c(k+1)}\right) A_{k-1},$$
$$A_0 = 1.$$

Thus, the behavioral model of forecasting can tractably be imported to express the steady-state behavior of the queue.

5.2.2. Direct Implications of the Behavioral Phenomena. Again, one could apply this model to queuing systems and conduct in-depth analyses to try to address various questions. For example, the model can be inserted in Cui and Veeraraghavan (2016), which takes arbitrary service rate belief distributions to study congestion and revenues. However, even without doing so, we can generate implications for queuing theory by interpreting the behavioral phenomena derived earlier.

In contrast with the perfectly rational model, the present model captures dispersion in arriving customers' wait-time expectations. In particular, from Section 4.2 and (7), it predicts that wait-time forecasts are more disperse when the lines are longer.

Furthermore, because service times are exponential, the model predicts that most people underestimate their wait time and a few people will greatly overestimate their wait time (see Subsection 4.3). Thus, when the true expected cost of waiting equals the reward (i.e., when (12) holds at equality), individuals are more likely to join than balk in the behavioral model (i.e., the probability that (13) holds is greater than 0.5; see Proposition 3).

The implication from the optimizer's curse (see Section 4.4) is that, on average, customers will be disappointed in their wait-time experiences. Individuals choose the maximum of the net reward from joining and zero (from not joining). Therefore, the true expected wait time among those individuals who choose to join tends to be longer than their respective point forecasts (see Proposition 4). This gap between expectations and experience is important because a wealth of behavioral research shows that the outcome relative to the expectation is a fundamental determinant of (dis)satisfaction.

One prediction from Jensen's inequality neglect (see Section 4.5) is that customers overestimate how many positions they should expect to move forward in a given time interval. This result follows from Corollary 1: calculating how much progress one expects to make in a given time interval requires taking the reciprocal of the forecasted wait time per customer, and the reciprocal is a convex function. Therefore, an individual may choose to renege because of slower-than-expected progress within the queue.

Finally, the predicted law of small numbers and the nonbelief in the law of large numbers (see Sections 4.9 and 4.10) imply that individuals underestimate the uncertainty of the wait times for short lines but overestimate the uncertainty of wait times for long lines. This pattern suggests that, all else equal, customers may appear to prefer shorter lines even if there is no difference in actual wait time.

6. Conclusion

Accounting for behavioral elements of forecasting is important for more accurately modeling decision-making behavior in operations and management science. In this paper we presented a behavioral model of forecasting based on the psychological process of mental sampling and naive statistics. It captures many forecasting-related behavioral phenomena and can be directly imported into formal models that typically assume individuals have perfect understanding of random variables and stochastic processes. Our model can serve as a building block such that a rich set of forecasting behavior can be accounted for with a single parameter within more complex models. It may also be useful for empiricists because it is grounded in credible psychology and may help explain observed anomalies or be a source for new testable hypotheses. We have illustrated some of these opportunities for inventory management and queuing applications.

Researchers who are interested in achieving greater descriptive accuracy can further develop this model. For instance, rather than assuming inferences from small samples are purely naive, one could parameterize the extent to which an individual applies statistical corrections to infer the properties of the true uncertainty. Additionally, our model assumes that the mean and variance of a sample is computed without error, but one could add random error in these calculations. Furthermore, one could relax the assumption that the possible outcomes considered by the individual are random draws from the true uncertainty distribution (see Section 3.1). Such complexities reduce the parsimony of the model; however, achieving greater descriptive accuracy through additional parameterization may build a stronger linkage to the psychology literature on judgment and decision making and generate new insights and hypotheses.

We make no strong claims about how to increase mental samples or reduce statistical naivete. However, at a high level, the model suggests that a focus on the consideration of more possible outcomes can improve forecasting. Some strategies observed in management practice are seemingly consistent with an attempt to increase one's mental sample size. For example, BlackRock, the world's largest asset management firm, encourages its employees to consider 20 possible outcome scenarios before forming an opinion about how a given alternative affects their risk management strategy. Additionally, the model suggests that if a debiasing strategy works for one forecasting bias, there is reason to believe that it will also work for the others, because they can be the product of the same underlying process. Finally, the model may be useful from a system design standpoint. Rather than attempting to improve the judgment of individuals, system designers can use the model to intelligently structure the environment in anticipation of behavioral deviations from optimal forecasting.

The illustrative OM examples we provided raise other important and unanswered research questions that can be fully developed. Researchers in other disciplines in management science and economics may also be able to use our model to relax assumptions of perfect rationality in belief formation within other contexts. Our hope is that this work serves as a useful building block for future analytical and empirical research both in operations and more broadly.

Acknowledgments

The authors thank the following people for their valuable conversations and feedback: Greg DeCroix, Justin Sydnor, Steve Leider, Laurens Debo, Xuanming Su, Andy King, Connie Helfat, and Nagesh Gavirneni. Additionally, the authors are grateful for the helpful comments and questions from the seminar participants at Duke, the University of Texas at Dallas, and Tuck, and the conference participants at the 2015 Behavioral Operations Conference. Finally, the authors thank the department editor and the anonymous associate editor and reviewers for their constructive feedback.

Appendix

Proof of Propositions 1 and 2. The proof follows directly from the definitions of X_b, x_*, and $\epsilon_b^2(X_b)$. □

Proof of Proposition 3. The first part follows from the random sampling assumption. The second part follows from the fact that $Mode(Z) < Median(Z) < \mu$ holds if Z belongs to the Pearson family (see Groeneveld and Meeden 1977). □

Proof of Proposition 4. The proof follows very closely that of Proposition 1 in Smith and Winkler (2006). Observe that the true mean values $\mu_1, \mu_2, \ldots, \mu_k$ are fixed, but the point forecasts $X_{b,1}, X_{b,2}, \ldots, X_{b,k}$, which are what the individual believes to be the true mean values, are uncertain. Let j^* denote the random variable with the maximum true mean $\mu_{j^*} = \max \mu_1, \mu_2, \ldots, \mu_k$. Thus, j^* is fixed, but i^* is a random variable that depends on the outcomes of the point forecasts. From the definitions of i^* and j^*, we have

$$\mu_{i^*} - X_{b,i^*} \leq \mu_{j^*} - X_{b,i^*} \leq \mu_{j^*} - X_{b,j^*}.$$

Taking expectation over the uncertainty regarding the point forecasts, we have $E[\mu_{i^*} - X_{b,i^*}] \le E[\mu_{j^*} - X_{b,j^*}]$. The right-hand side equals zero because j^* is independent of the point forecasts and $E[X_{b,j}] = \mu_j$ for all j. Thus, $E[\mu_{i^*} - X_{b,i^*}] \le 0$. The left-hand side is not necessarily zero because i^* is a random variable that depends on the point forecasts of each Z. If there is a chance that i^* does not equal j^*, then the inequality is strict. Because the individual believes that X_{b,i^*} is the true mean, we have the result. □

Proof of Proposition 5. By Jensen's inequality, we have $g(E(X_b)) \le E[g(X_b)]$. Substituting $E[X_b] = \mu$, we obtain $g(\mu) < E[g(X_b)]$. The right-hand side is the expected value of the individual's belief about $g(\mu)$. □

Proof of Proposition 6. The proof follows from the expressions for $\epsilon_b^2(X_b)$ and $\epsilon_*^2(X_b)$. □

Proof of Proposition 7. The proof follows the same proof that the sample mean and the sample variance are independent for a normal population (e.g., Casella and Berger 2002, p. 218), with some minor adjustments to correct for the fact that the individual's belief about the variance is the sample variance uncorrected for the sample size. □

Proof of Proposition 8. The actual error of y is given by $E[(y-Z)^2] = E[y^2 - 2yZ + Z^2] = y^2 - 2y\mu + E[Z^2] = y^2 - 2y\mu + \sigma^2 + \mu^2 = \sigma^2 + (y-\mu)^2$. The individual expects the error to be

$$\begin{aligned}
E[\epsilon_b^2(y)] &= E[\sigma_b^2 + (y - \mu_b)^2] \\
&= \frac{n-1}{n}\sigma^2 + y^2 - 2y\mu + E[\mu_b^2] \\
&= \frac{n-1}{n}\sigma^2 + y^2 - 2y\mu + \mathrm{Var}[\mu_b] + \mu^2 \\
&= \frac{n-1}{n}\sigma^2 + y^2 - 2y\mu + \frac{\sigma^2}{n} + \mu^2 \\
&= \sigma^2 + (y-\mu)^2, \quad \square
\end{aligned}$$

Proof of Proposition 9. The individual expects his forecast error to be

$$\epsilon_b^2(X_b(Z_t)) = \frac{t^2}{l^2}\sigma_b^2(Z_l),$$

which has expected value $(t^2/l)((n-1)/n)\sigma^2(Z^1)$. His actual error is

$$\begin{aligned}
E[\epsilon_*^2(X_b(Z_t))] &= \frac{t^2}{l^2 n}\sigma^2(Z_l) + \sigma^2(Z_t) \\
&= \frac{t^2}{ln}\sigma^2(Z^1) + t\sigma^2(Z^1) \\
&= \frac{t+ln}{ln} t\sigma^2(Z^1).
\end{aligned}$$

The individual underestimates her own point forecast error if $E[\epsilon_b^2(X_b(Z_t))] < E[\epsilon_*^2(X_b(Z_t))]$. Substituting the values above, this inequality is equivalent to

$$\begin{aligned}
\frac{t^2}{l}\frac{n-1}{n}\sigma^2(Z^1) &< \frac{t+ln}{ln} t\sigma^2(Z^1), \\
t(n-1) &< t + ln, \\
t &< l\frac{n}{n-2}.
\end{aligned}$$

To see that the bias is decreasing in t, we take the derivative of the quotient:

$$\begin{aligned}
\frac{\partial}{\partial t}\left(\frac{E[\epsilon_b^2(X_b(Z_t))]}{E[\epsilon_*^2(X_b(Z_t))]}\right) &= \frac{\partial}{\partial t}\left(\frac{(t^2/l)((n-1)/n)\sigma^2(Z^1)}{((t+ln)/(ln))t\sigma^2(Z^1)}\right) \\
&= \frac{\partial}{\partial t}\left(\frac{t(n-1)}{t+ln}\right) \\
&= \frac{(n-1)(t+ln) - t(n-1)}{(t+ln)^2} \\
&= \frac{ln(n-1)}{(t+ln)^2} > 0,
\end{aligned}$$

so the individual underestimates error further (the numerator is smaller) for smaller values of t. That is, one should expect to find overconfidence when the time horizon being forecasted is similar to, or shorter than, the reference time unit, l. Note that for $t = l$, the result reduces to the single random variable case, and the individual underestimates the error by a factor $(n-1)/(n+1)$ (see Proposition 6). □

Proof of Proposition 10. Similar to the proof of Proposition 9. □

Endnotes

[1] Consistent with this aspect, leading psychologists have called for psychological theories that "serve an integrative function by explaining multiple phenomena, providing an organizing principle for a field criticized for being long on effects and short on unifying explanations" (Weber and Johnson 2009, p. 56).

[2] For example, the "availability heuristic" (Tversky and Kahneman 1973) suggests that mental samples are not random because some outcomes are systematically harder to imagine. Similarly, there is also a stream of literature that investigates how nonrandom samples emerge in the environment and from individuals' decision making (e.g., March 1996, Denrell 2005, Feiler et al. 2013, Feiler and Kleinbaum 2015).

[3] While random error is not the most commonly cited explanation for overconfidence, some psychologists have stressed its potential importance (e.g., Soll 1996, Soll and Klayman 2004).

[4] Recall our comment that when leveraging the wisdom of the crowd, one should not necessarily discard point forecasts that seem like outliers. Here, we find a reason why it may be difficult to follow this advice: when Z is skewed, individuals with extreme point forecasts also tend to report lower confidence, making their advice tempting to ignore.

References

Anderson SP, De Palma A, Thisse JF (1992) *Discrete Choice Theory of Product Differentiation* (MIT Press, Cambridge, MA).

Armstrong JS (2001) Combining forecasts. Armstrong JS, ed. *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Kluwer, New York), 417–439.

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

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

Benjamin DJ, Rabin M, Raymond C (2016) A model of nonbelief in the law of large numbers. *J. European Econom. Assoc.* 14(2): 515-–544.

Bolton GE, Katok E (2008) Learning by doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback. *Manufacturing Service Oper. Management* 10(3):519–538.

Buehler R, Griffin D, Ross M (1994) Exploring the "planning fallacy": Why people underestimate their task completion times. *J. Personality Soc. Psych.* 67(3):366–381.

Casella G, Berger RL (2002) *Statistical Inference* (Duxbury, Pacific Grove, CA).

Chen L, Kök AG, Tong JD (2013) The effect of payment schemes on inventory decisions: The role of mental accounting. *Management Sci.* 59(2):436–451.

Chen Y, Su X, Zhao X (2012) Modeling bounded rationality in capacity allocation games with the quantal response equilibrium. *Management Sci.* 58(10):1952–1962.

Clemen RT (1989) Combining forecasts: A review and annotated bibliography. *Internat. J. Forecasting* 5:559–609.

Croson D, Croson R, Ren Y (2008) How to manage an overconfident newsvendor. Working paper, Southern Methodist University, Dallas.

Croson R, Schultz K, Siemsen E, Yeo ML (2013) Behavioral operations: The state of the field. *J. Oper. Management* 31(1):1–5.

Cui S, Veeraraghavan S (2016) Blind queues: The impact of consumer beliefs on revenues and congestion. *Management Sci.* 62(12): 3656–3672.

Denrell J (2005) Why most people disapprove of me: Experience sampling in impression formation. *Psych. Rev.* 112(4):951–978.

Einhorn HJ, Hogarth RM, Klempner E (1977) Quality of group judgment. *Psych. Bull.* 84(1):158–172.

Feiler DC, Kleinbaum AM (2015) Popularity, similarity and the network extraversion bias. *Psych. Sci.* 26(5):593–603.

Feiler DC, Tong JD, Larrick RP (2013) Biased judgment in censored environments. *Management Sci.* 59(3):573–591.

Fiedler K (2000) Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psych. Rev.* 107(4):659–676.

Fiedler K, Juslin P (2006) Taking the interface between mind and environment seriously. Fiedler K, Juslin P, eds. *Information Sampling and Adaptive Cognition* (Cambridge University Press, New York), 3–32.

Fisher M, Raman A (1996) Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.* 44(1):87–99.

Fisher M, Raman A (2010) *The New Science of Retailing: How Analytics Are Transforming the Supply Chain and Improving Performance* (Harvard Business Press, Cambridge, MA).

Gaur V, Kesavan S, Raman A, Fisher ML (2007) Estimating demand uncertainty using judgmental forecasts. *Manufacturing Service Oper. Management* 9(4):480–491.

Gigerenzer G, Hoffrage U (1995) How to improve Bayesian reasoning without instruction: Frequency formats. *Psych. Rev.* 102(4): 684–704.

Gigerenzer G, Hoffrage U, Kleinbölting H (1991) Probabilistic mental models: A Brunswikian theory of confidence. *Psych. Rev.* 98(4):506–528.

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

Groeneveld RA, Meeden G (1977) The mode, median, and mean inequality. *Amer. Statistician* 31(3):120–121.

Hadar L, Fox CR (2009) Information asymmetry in decision from description versus decision from experience. *Judgment Decision Making* 4(4):317–325.

Hammond JH, Obermeyer WR, Raman A (1994) Making supply meet demand in an uncertain world. *Harvard Bus. Rev.* 72(3):83–93.

Hansson P, Juslin P, Winman A (2008) The role of short-term memory capacity and task experience for overconfidence in judgment under uncertainty. *J. Experiment. Psych. Learn., Memory, Cognition* 34(5):1027–1042.

Haran U, Moore DA, Morewedge CK (2010) A simple remedy for overprecision in judgment. *Judgment Decision Making* 5(7): 467–476.

Harrison J, March J (1984) Decision making and postdecision surprises. *Admin. Sci. Quart.* 29(1):26–42.

Henry R (1993) Group judgment accuracy: Reliability and validity of postdiscussion confidence judgments. *Organ. Behav. Human Decision Processes* 56(1):11–27.

Hertwig R, Barron G, Weber EU, Erev I (2004) Decisions from experience and the effect of rare events in risky choice. *Psych. Sci.* 15(8):534–539.

Huang T, Chen YJ (2015) Service systems with experience-based anecdotal reasoning customers. *Production Oper. Management* 24(5):778–790.

Huang T, Liu Q (2015) Strategic capacity management when customers have boundedly rational expectations. *Production Oper. Management* 24(12):1852–1869.

Huang T, Yu Y (2014) Sell probabilistic goods? A behavioral explanation for opaque selling. *Marketing Sci.* 33(5):743–759.

Huang T, Allon G, Bassamboo A (2013) Bounded rationality in service systems. *Manufacturing Service Oper. Management* 15(2): 263–279.

Jain K, Mukherjee K, Bearden JN, Gaba A (2013) Unpacking the future: A nudge toward wider subjective confidence intervals. *Management Sci.* 59(9):1970–1987.

Juslin P, Wennerholm P, Olsson H (1999) Format dependence in subjective probability calibration. *J. Experiment. Psych.: Learn., Memory, Cognition* 25(4):1038–1052.

Juslin P, Winman A, Hansson P (2007) The naive intuitive statistician: A naive sampling model of intuitive confidence intervals. *Psych. Rev.* 114(3):678–703.

Kahneman D, Tversky A (1972) Subjective probability: A judgment of representativeness. *Cognitive Psych.* 3(3):430–454.

Kahneman D, Tversky A (1973) On the psychology of prediction. *Psych. Rev.* 80(4):237–251.

Kahneman D, Tversky A (1979a) Intuitive prediction: Biases and corrective procedures. *TIMS Stud. Management Sci.* 12(1): 313–327.

Kahneman D, Tversky A (1979b) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.

Kareev Y (2000) Seven (indeed, plus or minus two) and the detection of correlation. *Psych. Rev.* 107(2):397–402.

Kareev Y (2006) Good sampling, distorted views: The perception of variability. Fiedler K, Juslin P, eds. *Information Sampling and Adaptive Cognition* (Cambridge University Press, Cambridge, UK), 33–52.

Kareev Y, Arnon S, Horwitz-Zeliger R (2002) On the misperception of variability. *J. Experiment. Psych.: General* 131(2):287–297.

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

Klayman J, Soll JB, González-Vallejo C, Barlas S (1999) Overconfidence: It depends on how, what, and whom you ask. *Organ. Behav. Human Decision Processes* 79(3):216–247.

Kremer M, Minner S, Van Wassenhove LN (2010) Do random errors explain newsvendor behavior? *Manufacturing Service Oper. Management* 12(4):673–681.

Kremer M, Moritz B, Siemsen E (2011) Demand forecasting behavior: System neglect and change detection. *Management Sci.* 57(10):1827–1843.

Kremer M, Siemsen E, Thomas DJ (2016) The sum and its parts: Judgmental hierarchical forecasting. *Management Sci.* 62(9): 2745–2764.

Kruger J, Evans M (2004) If you don't want to be late, enumerate: Unpacking reduces the planning fallacy. *J. Experiment. Soc. Psych.* 40(5):586–598.

Kulkarni VG (2009) *Modeling and Analysis of Stochastic Systems*, 2nd ed. (CRC Press, Boca Raton, FL).

Laibson D (1997) Golden eggs and hyperbolic discounting. *Quart. J. Econom.* 112(2):443–478.

Larrick RP, Soll JB (2006) Intuitions about combining opinions: misappreciation of the averaging principle. *Management Sci.* 52(1):111–127.

Lee YS, Siemsen E (2017) Task decomposition and newsvendor decision making. *Management Sci.* 63(10):3226–3245.

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

Luce RD (1959) *Individual Choice Behavior: A Theoretical Analysis* (John Wiley & Sons, New York).

March J (1996) Learning to be risk-averse. *Psych. Rev.* 103(2):309–319.

McFadden (1981) Econometric models of probabilistic choice. Manski C, McFadden D, eds. *Structural Analysis of Discrete Data* (MIT Press, Cambridge, MA), 198–272.

McKelvey RD, Palfrey TR (1995) Quantal response equilibria for normal form games. *Games Econom. Behav.* 10(1):6–38.

Moore DA, Healy PJ (2008) The trouble with overconfidence. *Psych. Rev.* 115(2):502–517.

Moritz BB, Hill AV, Donohue KL (2013) Individual differences in the newsvendor problem: Behavior and cognitive reflection. *J. Oper. Management* 31(1):72–85.

Moritz B, Siemsen E, Kremer M (2014) Judgmental forecasting: Cognitive reflection and decision speed. *Production Oper. Management* 23(7):1146–1160.

Nagarajan M, Shechter S (2013) Prospect theory and the newsvendor problem. *Management Sci.* 60(4):1057–1062.

Naor P (1969) The regulation of queue size by levying tolls. *Econometrica* 37(1):15–24.

Nosofsky RM, Palmeri TJ (1997) An exemplar-based random walk model of speede classification. *Psych. Rev.* 104(2):266–300.

Osborne M, Rubinstein A (1998) Games with procedurally rational players. *Amer. Econom. Rev.* 88(4):834–847.

Özer Ö, Zheng Y, Chen KY (2011) Trust in forecast information sharing. *Management Sci.* 57(6):1111–1137.

Plambeck EL, Wang Q (2013) Implications of hyperbolic discounting for optimal pricing and scheduling of unpleasant services that generate future benefits. *Management Sci.* 59(8):1927–1946.

Popescu I, Wu Y (2007) Dynamic pricing strategies with reference effects. *Oper. Res.* 55(3):413–429.

Rabin M (2002) Inference by believers in the law of small numbers. *Quart. J. Econom.* 117(3):775–816.

Rabin M (2013) Incorporating limited rationality into economics. *J. Econom. Literature* 51(2):528–543.

Rabin M, Vayanos D (2010) The gambler's and hot-hand fallacies: Theory and applications. *Rev. Econom. Stud.* 77(2):730–778.

Rakow T, Demes KA, Newell BR (2008) Biased samples not mode of presentation: Re-examining the apparent under-weighting of rare events in experience-based choice. *Organ. Behav. Human Decision Processes* 106(2):168–179.

Rapoport A, Budescu DV (1992) Generation of random series in two-person strictly competitive games. *J. Experiment. Psych.: General* 121(3):352–363.

Rapoport A, Budescu DV (1997) Randomization in individual choice behavior. *Psych. Rev.* 104(3):603–617.

Ren Y, Croson R (2013) Overconfidence in newsvendor orders: An experimental study. *Management Sci.* 59(11):2502–2517.

Rudi N, Drake D (2014) Observation bias: The impact of demand censoring on newsvendor level and adjustment behavior. *Management Sci.* 60(5):1334–1345.

Schweitzer ME, Cachon GP (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Sci.* 46(3):404–420.

Silva JS, Tenreyro S (2006) The log of gravity. *Rev. Econom. Statist.* 88(4):641–658.

Simon HA (1955) A behavioral model of rational choice. *Quart. J. Econom.* 69(1):99–118.

Smith J, Winkler R (2006) The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Sci.* 52(3):311–322.

Sniezek J, Henry R (1989) Accuracy and confidence in group judgment. *Organ. Behav. Human Decision Processes* 43(1):1–28.

Spiegler R (2006a) Competition over agents with boundedly rational expectations. *Theoret. Econom.* 1(2):207–231.

Spiegler R (2006b) The market for quacks. *Rev. Econom. Stud.* 73(4): 1113–1131.

Spiegler R (2011) *Bounded Rationality and Industrial Organization* (Oxford University Press, New York).

Soll JB (1996) Determinants of overconfidence and miscalibration: The roles of random error and ecological structure. *Organ. Behav. Human Decision Processes* 65(2):117–137.

Soll JB, Klayman J (2004) Overconfidence in interval estimates. *J. Experiment. Psych.: Learn., Memory, Cognition* 30(2):299–314.

Soll JB, Mannes AE (2011) Judgmental aggregation strategies depend on whether the self is involved. *Internat. J. Forecasting* 27(1): 81–102.

Su X (2008) Bounded rationality in newsvendor models. *Manufacturing Service Oper. Management* 10(4):566–589.

Surowiecki J (2005) *The Wisdom of Crowds* (Anchor, New York).

Thaler R (1985) Mental accounting and consumer choice. *Marketing Sci.* 4(3):199–214.

Thaler RH (1988) Anomalies: The winner's curse. *J. Econom. Perspect.* 2(1):191–202.

Tong J, Ivantsova A, Feiler D (2016) How random error and intelligent choice yield overoptimism: An experimental examination of the optimizer's curse. Working paper, University of Wisconsin–Madison, Madison.

Tsai CI, Klayman J, Hastie R (2008) Effects of amount of information on judgment accuracy and confidence. *Organ. Behav. Human Decision Processes* 107(2):97–105.

Tversky A, Kahneman D (1971) Belief in the law of small numbers. *Psych. Bull.* 76(2):105–110.

Tversky A, Kahneman D (1973) Availability: A heuristic for judging frequency and probability. *Cognitive Psych.* 5(2):207–232.

Weber EU, Johnson EJ (2009) Mindful judgment and decision making. *Annual Rev. Psych.* 60:53–85.

Yaniv I, Kleinberger E (2000) Advice taking in decision making: Egocentric discounting and reputation formation. *Organ. Behav. Human Decision Processes* 83(2):260–281.

Zarnowitz V, Lambros LA (1987) Consensus and uncertainty in economic prediction. *J. Political Econom.* 95(3):591–621.

Zhang Y, Donohue K, Cui TH (2016) Contract preferences for the loss averse supplier: Buyback versus revenue sharing. *Management Sci.* 62(6):1734–1754.

Zipkin PH (2000) *Foundations of Inventory Management* (McGraw-Hill, New York).