Managerial Decision Making in Censored Environments:

Biased Judgment of Demand, Risk, and Employee Capability

by

Daniel C. Feiler

Department of Business Administration
Duke University

Date:_____

Approved:

_____
Richard Larrick, Supervisor

_____
Dan Ariely

_____
Sim Sitkin

_____
Jack Soll

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the Department of
Business Administration in the Graduate School
of Duke University

2012

ABSTRACT

Managerial Decision Making in Censored Environments:

Biased Judgment of Demand, Risk, and Employee Capability

by

Daniel C. Feiler

Department of Business Administration
Duke University

Date:_____

Approved:

_____
Richard Larrick, Supervisor

_____
Dan Ariely

_____
Sim Sitkin

_____
Jack Soll

An abstract of a dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the Department of
Business Administration in the Graduate School
of Duke University

2012

Copyright by
Daniel C. Feiler
2012

Abstract

Individuals have the tendency to believe that they have complete information when making decisions. In many contexts this propensity allows for swift, efficient, and generally effective decision making. However, individuals cannot always see a representative picture of the world in which they operate. This paper examines judgment in *censored environments* in which a constraint, the *censorship point*, systematically distorts the sample observed by a decision maker. Random instances beyond the censorship point are observed at the censorship point, while instances below the censorship point are observed at their true value. Many important managerial decisions occur in censored environments, such as inventory, risk-taking, and employee evaluation decisions. This empirical work demonstrates a *censorship bias* – individuals tend to rely too heavily on the observed censored sample, biasing their beliefs about the underlying population. Further, the censorship bias is exacerbated for higher degrees of censorship, higher variance in the population, and higher variability in the censorship points. Evidence from seven studies demonstrates how the censorship bias can cause managers to underestimate demand for their goods, overestimate risk in their environments, and underappreciate the capabilities of their employees, all of which can lead to undesirable outcomes for their organizations.

Contents

Abstract iv

List of Tables ix

List of Figures x

Acknowledgements xii

Chapter 1: Introduction, Literature Review, and Theory 1

Introduction 1

Literature Review 4

A Theory of Judgment in Censored Environments 6

Judging the mean of a normal population from a censored sample 7

Behavior in Censored Environments 8

Chapter 2: Biased Judgment of Demand 16

Study 1: Degree of Censorship and Variability of Censorship Point 16

Methods 17

Participants 17

Instructions and procedures 18

Results 19

Comparisons to truth 19

Comparisons to prescriptive heuristic estimate 19

Adjustments from sample mean 20

Study 2: Judgment and Choice with Known versus Unknown Demand 21

Methods 23
Participants 23
Instructions 24
Procedures 24
Results 25
Demand beliefs 26
Stocking decisions 26
Adjustments from sample mean 28
Study 3: Debiasing with Heightened Awareness of Missed Sales 29
Methods 30
Participants 30
Procedures 30
Results 31
Chapter 3: Biased Judgment of Risk 34
Study 4: Risk Perceptions in Sequential Risk-taking 34
The Task 36
Methods 38
Participants 38
Procedures 38
Results 40
Comparisons to observed sample maximum 41

Chapter 4: Biased Judgment of Employee Capability 43
When to Predict Underestimation of Employees 47
Study 5: Deadlines and Employee Variance 49
Methods 50
The Task 50
Procedures 52
Results 53
Comparisons to truth 54
Adjustments from sample mean 55
Ranking accuracy 56
Study 6: Common Work Assignments 58
Method 61
The Task 61
Procedures 63
Results 63
Study 7: Failure and Success Framing 65
Methods 68
Procedures 68
Results 70
Chapter 5: Predictors of Employee Underestimation 74
The Degree of Constraint 74

Employee Disclosure 77
Managerial Awareness 82
Discussion 84
Chapter 6: General Discussion 87
Implications 90
Demand Estimation 90
Risky Choice 92
Employee Capability 92
Improving Learning 94
Conclusion 95
Appendix A 96
Appendix B 97
Appendix C 98
Appendix D 99
References 100
Biography 112

List of Tables

Table 1: Average estimates of mean demand and stocking decisions by condition. The estimates and stocking decisions are relative to true mean demand. * indicates a significant bias below the true mean ($p < .01$). 26

Table 2: Final estimates and stocking decision relative to true mean demand in Study 3. 32

List of Figures

Figure 1: A depiction of observations in an uncensored versus a right censored environment. 3

Figure 2: Censored environments with high (75%) versus low (25%) censorship. 10

Figure 3: A censored environment with low versus high variance. 12

Figure 4: A censored environment with stationary versus variable censorship points. 14

Figure 5: Study 1 mean final demand estimates, observed sample means, and prescriptive heuristic estimates displayed by condition, with standard error bars. The dashed line indicates the true mean demand, 575. 20

Figure 6: Participants' mean final estimates of mean demand and mean observed sample means with standard error bars displayed by censorship condition. Data are from the unknown demand conditions only. Estimates are shown relative to true mean demand. 27

Figure 7: Study 3 mean participant estimates of the number of fish in the pond (N) and the mean maximum number of fish observed by participants displayed by censorship condition, with standard error bars. 39

Figure 8: How deadlines censor employee capability. 45

Figure 9: Depiction of the observed outcomes for employees with high and low variance capabilities when completion times are censored by a deadline. 50

Figure 10: Participants' mean estimates of mean employee capability and mean observed sample means by censorship environment and employee capability variance, with standard error bars. 55

Figure 11: An illustrative graph of participants' ranks of employees relative to their true rank in the censored environment by a median split on employee capability variance. 57

Figure 12: How work assignments censor employee capability. 59

Figure 13: A depiction of how, with common work assignments, the greater censorship and underestimation of high performers leads to a smaller perceived range of employee

capabilities. The arrows indiciate the degree to which each employee is underestimated. 60

Figure 14: An illustration of how the capability distribution was shifted to yield high-, medium-, and low-performers. The normal distribution is used here simply for illustration. 62

Figure 15: Participants' estimates of mean capability across levels of employee capability in Study 6. 65

Figure 16: The estimates and work allocations by condition in Study 6. The true mean capability of the employee was 84. 71

Figure 17: A workforce pyramid expressing the relationship between human capital and employee underestimation. 76

Figure 18: Depiction of the task interface for Study 1. This participant next needs to input a best estimate for the mean of underlying demand (m) for period 5. 96

Figure 19: Depiction of the game interface in Study 2. The demand column only appeared for participants in the uncensored condition. This participant next needs to input a best guess for underlying mean demand, m, and an ordering decision for period 5. 97

Figure 20: A screenshot of the sequential risk taking task in Study 4. 98

Figure 21: A depiction of the first five employee task completion times observed in the censored condition in Study 5. In the uncensored condition, participants could also observe the exact completion times in periods where the employee was finished by the deadline. The deadline for this emploee was set to 81 seconds. 99

Acknowledgements

I thank my co-authors, Jordan Tong and Richard Larrick (committee chair), for their considerable and much-appreciated contributions to this paper. I also thank my other committee members, Jack Soll, Sim Sitkin, and Dan Ariely, who have played an important role in shaping this dissertation. Finally, I especially thank Ayumi Shimokawa for her love, support, and feedback.

Chapter 1: Introduction, Literature Review, and Theory

> "Human rational behavior… is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor."
>
> Herbert Simon (1990, p. 7)

Introduction

Nearly all organizations operate with incomplete and imperfect information. An informed decision often requires one to extrapolate from an imperfect sample of observations to infer the true underlying properties operating in the environment. In this paper I will discuss how a specific form of incomplete information can cause managers to underestimate demand, overestimate risk, and underappreciate employees. The type of environment is a *censored environment*, in which a constraint distorts the observed sample of data. In a censored environment, random draws from an underlying population sometimes lie beyond a specific *censorship point* but are observed at the value of the censorship point. These instances are known to be censored but their exact magnitude is uncertain (see Figure 1). In contrast, random draws that lie below the censorship point are uncensored and observed at their true value. The critical question in such environments is whether individuals can infer the true nature of the underlying population from their censored sample of observations.

Censored environments are interesting to study because they occur in many managerial settings. Consider the following three organizational examples. First, many firms cannot observe sales missed after an inventory stockout. If a procurement officer holds 100 units of inventory in a given period, she can observe exact values of demand

less than 100, but she cannot observe the exact value of demand if demand is greater than 100. The procurement officer's challenge is to infer the nature of the true underlying demand distribution while observing only sales, which is a biased sample of demand observations. Second, a primary task for most managers is to make inferences about the capabilities of their employees. However, while managers can almost always observe when an employee falls short in a task, they often cannot observe how much more employees are capable of doing in periods in which they complete the work assigned to them. This asymmetry provides a biased sample of observations to the manager and makes inference about employees' capabilities difficult. Finally, consider when a manufacturer engages in a policy of precautionary machine replacement after a certain amount of use. In the event that the machine breaks before replacement, he observes its exact lifetime. However, in the event of replacement, he cannot observe the additional time the machine could have functioned, resulting in a biased sample of lifetime observations. In each of these examples, accurate judgment from a censored sample is crucial for effective decision-making. A biased estimation of demand leads to a biased inventory policy, an inaccurate estimate of employee productivity leads to inefficient work assignments and dismissal decisions, and a biased estimation of the lifetime of a machine leads to an inefficient replacement policy.[1]

[1] For a review of other examples of censorship, see Amemiya (1984).

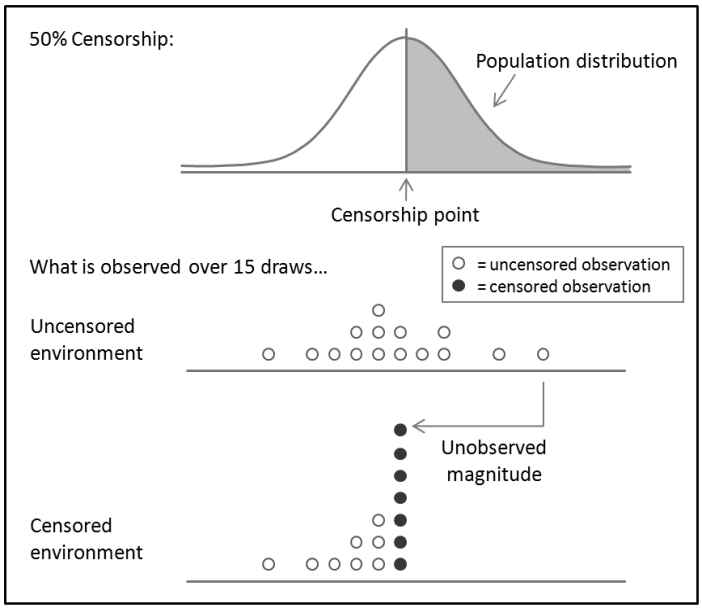

Figure 1: A depiction of observations in an uncensored versus a right censored environment.

In this research, my co-authors[2] and I show that individuals in censored environments exhibit a *censorship bias* – they form beliefs about the underlying population biased in the direction of the observed censored sample. This occurs even with full awareness of the presence of censorship. Further, certain structural dimensions of censored environments predictably determine the degree to which individuals form biased beliefs. In general, this research sheds light on which censored environments are likely to yield the most biased inferences and judgments and therefore are the most important candidates for intervention and improvement.

---

[2] A large portion of this work was done in collaboration with Jordan Tong and Rick Larrick. They have been instrumental to the development of this paper, particularly the narrative and theory development in Chapter 1, the empirical work on demand learning in Chapter 2, and the discussion in Chapter 6. I am tremendously grateful for their contributions.

Literature Review

Previous research suggests that the accuracy of an individual's inference depends on the nature and completeness of the sample of observations they experience (Gilovich 1991; Stewart, Chater, & Brown 2006). Hogarth (2001) suggests that inference is easiest in environments that provide immediate, clear, and unbiased observations (Fiedler 2000; Fiedler & Juslin 2006; Hogarth, McKenzie, Gibbs, & Marquis 1991; Maddox, Ashby, & Bohil 2003). Many researchers have investigated the effects of non-representative samples on individual inference. Past research has primarily examined three possible causes of non-representative samples.

First, an individual may cause an unrepresentative sample by using a biased collection process. For example, in hypothesis testing, individuals may only search for evidence consistent with their theories and neglect to look for disconfirming evidence (Klayman & Ha 1987; Mynatt, Doherty, & Tweney 1977; Nickerson 1998). Relatedly, individuals may face biased samples of recollections from their memories. If recency increases accessibility of memories, then more recent occurrences may be over-represented in hindsight (MacLeod & Campbell 1992; Tversky & Kahneman 1973).

Second, an individual may face an unrepresentative sample when another individual with a specific agenda strategically presents a biased sample of observations (Brenner, Koehler, & Tversky 1996; Juslin 1994; Klayman, Soll, Gonzalez-Vellejo, & Barlas 1999; Koehler & Thompson 2006). For example, Silverman et al. (2010) found that doctors fail to account for conflicts of interest when drawing inferences from clinical trials. Similarly, Koehler and Mercer (2009) showed that mutual fund companies only

advertise their best performing funds but investors respond to the advertised data as if they are representative of company's overall performance. In these cases, individuals generally are not fully aware of the process by which the biased sample has been generated, but they are aware that the samples are presented with the intention of persuasion.

Finally, sometimes simple environmental constraints systematically create a biased sample. For example, if an environment permits the observation of only a chosen alternative's outcome, then an overly negative perception of a foregone alternative cannot be disconfirmed because its true value goes unobserved (Denrell 2005, 2007; Einhorn & Hogarth 1978; March 1996). Einhorn and Hogarth (1978) propose that human resource departments face this problem because they can more easily observe the performance of applicants they hire than the performance of those they reject. Empirical research has found similar results in other organizational contexts. Managers with overly pessimistic beliefs about employee motivations create conditions that make it difficult for their beliefs to be disconfirmed, while overly optimistic managers tend to have their beliefs corrected through experience (Markle 2009, 2011). And negotiators are aware when they claim less value than they expected, but are less aware when they could have claimed more (Larrick & Wu 2007).

A developing body of evidence suggests that many shortcomings in human judgment and decision-making occur as a result of incorrectly treating a biased sample of observations as representative of the true population (Fiedler 2000; Fiedler & Juslin 2006; Hansson, Juslin, & Winman 2008; Juslin, Winman, & Hansson 2007). This

perspective uses the metaphor of the decision maker as a *naïve intuitive statistician*. The metaphor proposes that individuals are optimal cognitive processors of observations but naïvely assume that their observed samples are representative of the population. When observed samples are biased, individuals may form biased inferences about the environment. This perspective has been used to provide alternative causal accounts of judgmental biases that had previously been attributed to ineffective cognitive processing (Fiedler 2000; Juslin, Winman, & Olssen 2000). The present research builds on a central tenet of the naïve intuitive statistician metaphor: the assumption that individuals naïvely treat observed samples as representative of underlying populations. This will be referred to as Sample Naïveté Theory (SNT). Drawing on this theory, this paper examines the judgment and decision-making of individuals faced with misrepresentative samples created by censored environments. The application of this theory to several important managerial contexts reveals dangerous judgment pitfalls that have not been studied prior to this research.

A Theory of Judgment in Censored Environments

Consider the problem of estimating properties of an underlying population given a random sample. The problem is more difficult if the random sample is not representative of the population, as is the case in censored environments. In a censored environment, at least one constraint limits the range of observable values. This section focuses primarily on judgment in "right-censored" environments, in which a constraint prevents exact observations of values that fall to the right of a fixed point, because they are generally

more common in the real-world (e.g., capacity constraints),[3] however, I examine "left-censorship" in Study 5. In this section, however, the term censored environments refers to right-censored environments. For the sake of formalization, this section focuses primarily on a normally distributed population, although I examine inferences about a uniform population in Study 4 and an unknown, naturally occurring distribution in Study 4.

Judging the mean of a normal population from a censored sample[4]

Let $\{d_1, d_2, \dots, d_n\}$ be a random sample of size n from a normal population with known standard deviation σ and unknown mean μ. In a right-censored environment, the *censorship point* c prevents the observation of values greater than c. Thus, define the censored sample of $\{d_1, d_2, \dots, d_n\}$ as $\{(x_1, r_1), (x_2, r_2), \dots, (x_n, r_n)\}$, where

$$(x_i, r_i) = \begin{cases} (d_i, 0) \text{ if } d_i < c \text{ "uncensored observation"} \\ (c, 1) \text{ if } d_i \geq c \text{ "censored observation"} \end{cases} \quad i = \{1,2, \dots n\}$$

For right-censored environments, note that an observation x_i is simply the minimum of the sample d_i and the censorship point c. An environment is uncensored if no observations are censored (i.e., $x_i = d_i$ for all i.) Figure 1 provides a visual depiction of a right-censored and an uncensored environment. For convenience, the observed sample mean is defined as $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and the observed censorship rate $\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i$, which is the proportion of censored observations in the sample. In this paper primarily

---

[3] In a left-censored environment a censorship point prevents the observation of lower values. For example, Chen (2010) investigates a model in which customer type cannot be observed below a threshold. However, since the same principles should apply symmetrically, this section considers only the case of right-censored environments.

[4] The credit for this formalization goes to Jordan Tong.

focuses on the problem of estimating the population mean from a censored sample. That is, given a set of observations $\{(x_1, r_1), (x_2, r_2), \ldots, (x_n, r_n)\}$, what is μ?

In a right-censored environment, it is clear that the observed sample of $x's$, $\bar{x}$, provides a downward-biased estimate of μ. Thus, the difficulty in making judgments about the population from a censored sample is that one must correctly extrapolate from both the observed x's *and* the observed r's to make accurate inferences about the underlying population. Next, I present hypotheses for how individuals make such judgments by drawing on SNT.

Behavior in Censored Environments

As proposed by SNT (Fiedler & Juslin 2006), individuals may naively rely on observed samples to make inferences as if they were representative of the truth. Even when individuals are cognizant of constraints on observations, accounting for the bias in the observed sample often remains a complex task. Individuals may understand the direction in which to adjust their beliefs, but determining the degree to which one should adjust is difficult. As a point of reference, consider the purely naïve intuitive statistician who ignores censorship altogether and treats the observed sample as fully representative. Let e^n denote such a purely naïve estimate. Then e^n is simply the observed sample mean.

$$e^n = \bar{x}$$

If individuals behave consistent with SNT and are naïve to the constraint that limits observations beyond the censorship point, then they will form beliefs about the

population mean that are biased toward the mean of the observed sample. Let e^b denote the behavioral estimate of the population mean μ given a censored sample.

Hypothesis 1a: In a censored environment, estimates of the population mean will be biased low, toward the naïve estimate e^n (i.e., $e^b < \mu$).

Hypothesis 1b: Estimates of the population mean will be lower when decision makers face a censored environment than when they face an uncensored environment.

An advanced decision-maker can use a censored sample to make an accurate estimate of the population mean. Indeed, a prescriptive heuristic based on an approximate maximum likelihood estimate can accomplish this task with great accuracy (Nahmias 1994). The estimate of μ given a censored sample is denoted according to Nahmias (1994) as e^h.

$$e^h = \bar{y} + \frac{\sigma\phi\left(\Phi^{-1}(1-\bar{r})\right)}{1-\bar{r}}$$

Where $\bar{y}$ is the sample mean of the uncensored observations, and ϕ and Φ are the probability density and cumulative distribution functions of the standard normal distribution. In short, this prescriptive heuristic uses sample statistics – it starts with the sample mean of the uncensored observations and adjusts upwards according to the sample censorship rate and the population variance.[5]

As each individual might observe a different random sample drawn from the population, it is arguably fairer to make predictions about an individual's estimate relative to the heuristic's estimate given the same sample rather than to the true

[5] This heuristic cannot be applied in the special case that all observations are censored.

population mean. The prescriptive heuristic yields very near optimal estimates, therefore this prediction follows:

Hypothesis 1c: In a censored environment, estimates of the population mean will be lower than the prescriptive heuristic estimates (i.e., $e^b < e^h$).

Not all censored samples are created, or biased, equally. There are several key dimensions in censored environments that may cause individuals to form beliefs farther from the truth. The effects of these dimensions hinge on one straightforward assertion: as the distance between the observed sample mean and the true mean, $\mu - \bar{x}$, increases, the observed sample is less representative of the population, and individuals will form estimates of μ farther from the truth.

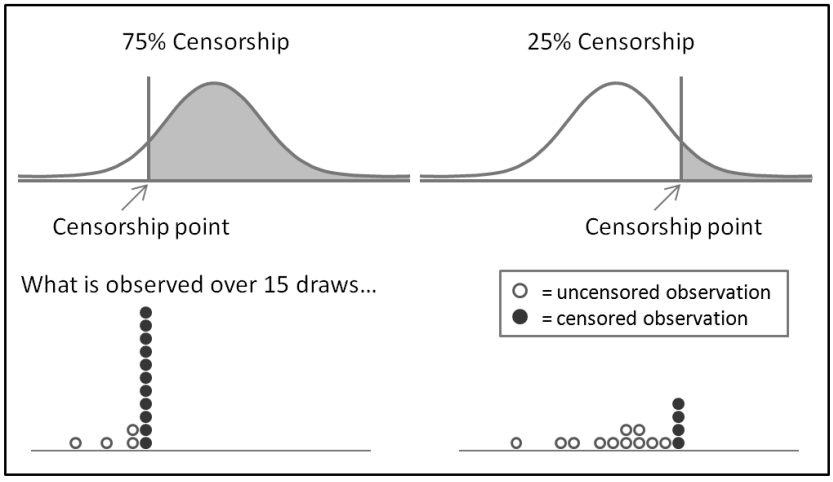

Figure 2: Censored environments with high (75%) versus low (25%) censorship.

This paper discusses two factors in censored environments that cause the observed sample mean to be farther from the true mean: degree of censorship and population variance. First, the degree of censorship increases as the censorship point

moves lower on the population distribution because a greater percentage of random instances will be restricted by the lower constraint (see Figure 2). With greater censorship, the censorship point screens observations more frequently in expectation and censored instances will be observed farther from their true values, on average. Therefore the mean of the observed sample moves farther from the true population mean as the degree of censorship increases.

Hypothesis 2: As the degree of censorship increases, estimates of the population mean will be biased farther from the true population mean (or equivalently, $\mu - e^b$ is decreasing in c).

Second, higher variance in the population causes the mean of the observed sample to lie farther from the population mean (see Figure 3).[6] With higher variance, the decision maker observes more extreme low and high instances; however, the more extreme high instances remain censored by the environment and are still observed at the censorship point. Therefore, holding the censorship point constant, the mean of the observed sample lies farther from μ as the variance of the population increases. As stated previously, as observed samples become more misrepresentative (i.e., as $\mu - \bar{x}$ increases) individuals' estimates of the population mean are expected to be more biased.

Hypothesis 3: As the variance of the population increases, estimates of the population mean will be biased farther from the true population mean (i.e., $\mu - e^b$ is increasing in σ).

---

[6] The actual property considered here is second order stochastic dominance (see Muller and Stoyan 2002), which for the normal distribution is equivalent to the variance.

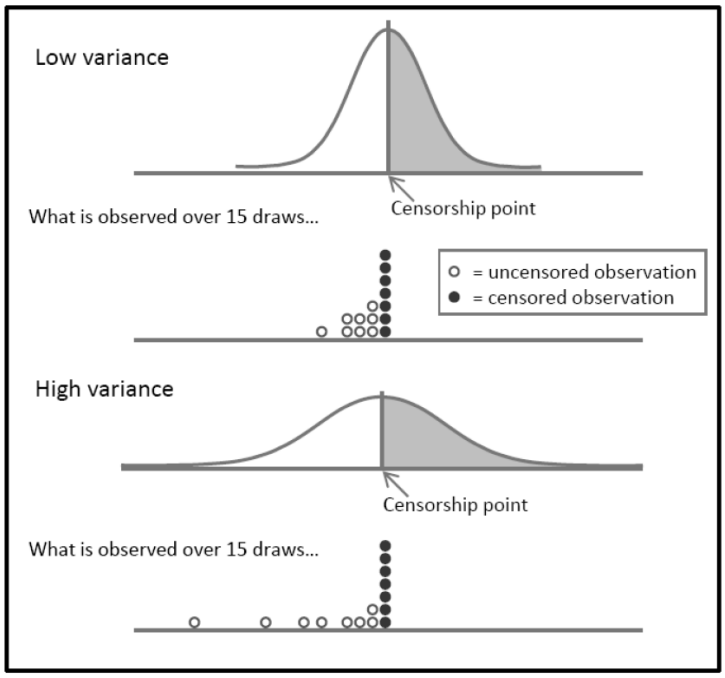

Figure 3: A censored environment with low versus high variance.

Censored environments are particularly interesting to study because individuals are aware of the censorship. Awareness of censorship cues decision-makers into the direction of the sample bias because observations are censored in only one direction. Further, individuals can see exactly how many, and at what point, observations are censored. Therefore, we predict that individuals will use evidence of censorship to adjust their estimates of the population mean in the appropriate direction from the mean of the observed sample, $\bar{x}$. However, past research shows that adjustment from an initial anchor tends to be insufficient (Chapman & Johnson 1999; Epley & Gilovich 2001, 2004, 2006; Tversky & Kahneman 1974; Strack & Mussweiler 1997). While we expect the naïveté prediction of Hypothesis 1 to be the dominant mechanism in censored environments, we

predict that individuals will be better than "pure naïveté," at least partially accounting for the sample bias.

Hypothesis 4: In a censored environment, estimates of the population mean will be greater than the mean of the observed sample (i.e., $e^b > e^n$).

While a greater degree of censorship increases the distance between the sample mean and the true mean, it also gives individuals more evidence of the sample bias by presenting more instances of censorship. As more observations accumulate at the censorship point, it may provide a stronger perceived violation of the expected normality and randomness anticipated by the decision-maker (Kahneman & Tversky 1973). This perceived violation may cause individuals to adjust for the sample bias to a greater extent. This possible secondary "cueing-to-adjust" mechanism of censored samples leads us to make the following prediction:

Hypothesis 5: As the degree of censorship increases, individuals will adjust farther from the observed sample mean in their estimates of the population mean (i.e., $e^b - \bar{x}$ is increasing in $\bar{r}$).

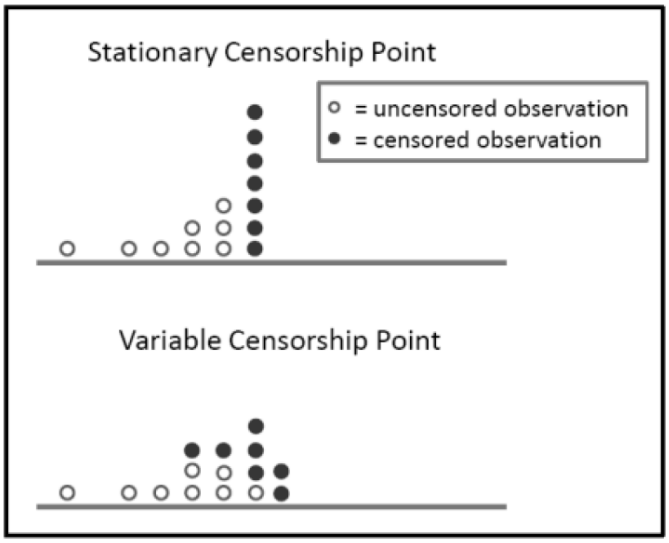

Figure 4: A censored environment with stationary versus variable censorship points.

Another factor that makes censored observations more pronounced, and may consequently spur greater accounting for the sample bias, is the variability of the censorship point. Variability is defined as mean-preserving spread. When the censorship point is stationary the censored observations amass at one specific point. A stationary censorship point therefore yields a single large spike of censored observations at that point. A variable censorship point, on the other hand, will produce censored instances at a number of different points (see Figure 4). Even if the variability is small, it will prevent observations from compiling at a single point. When censored observations are allowed to accrue at a single point (i.e., with a stationary censorship point) they likely become more pronounced to the decision-maker and may cause the decision maker to account for the sample bias to a greater extent.

Hypothesis 6: When facing a stationary, as opposed to variable, censorship point, individuals will adjust farther from the observed sample mean in their estimates of the population mean (i.e., $e^b - \bar{x}$ is larger if c is stationary).

In the next three chapters I will present empirical work that tested these hypotheses and demonstrated the censorship bias across the domains of demand learning, risk taking, and employee evaluation.

Chapter 2: Biased Judgment of Demand

Study 1: Degree of Censorship and Variability of Censorship Point

Study 1 was designed to test for the existence of a censorship bias (Hypotheses 1a-c). It was also designed to test whether the censorship bias is exacerbated by greater censorship (Hypothesis 2) and by variable censorship points (Hypothesis 4). In this task, individuals faced a normally distributed demand with an unknown mean and a known standard deviation. They observed randomly generated sales of a newspaper company and a binary indication of whether the company sold-out in each period. They then made an estimate of the underlying mean demand and subsequently observed the sales of the next period. The sales feedback for all past periods always remained visible and the task included 30 periods. In this task the demand distribution represents the underlying population about which an individual needs to draw inferences. The inventory of the paper company acts as a censorship point for observing demand and the actualized sales are the observed sample.

To manipulate the degree of censorship, the inventory levels (i.e., censorship points) were centered at the 25^{th}, 50^{th}, or 75^{th} fractile on the demand distribution. These fractiles were chosen to be consistent with the large body of literature on stocking decisions with known demand that has primarily studied the 25^{th} and 75^{th} critical fractiles (e.g., Schweitzer & Cachon 2000; Bolton & Katok 2008; Bostian, Holt, & Smith 2008; Lurie & Swaminathan 2008;Ho, Lim & Cui 2010; Chen, Kok, & Tong 2011). The presence of the censorship point at different fractiles alters the proportion of the demand

distribution that is censored. As discussed previously, the greater the censored proportion of the distribution, the greater the expected censorship rate.

The variability of the censorship point was also manipulated. In the stationary condition the inventory available for sale was constant across all 30 periods. In the variable condition the inventory was selected randomly each period from a uniform distribution, ±25 units from the relevant fractile determined by the censorship condition.

This study was a 3 (degree of censorship: high, medium, low) by 2 (censorship point variability: stationary, variable) between-subjects design with within-subject repeated measures for the 30 periods. The underlying unknown demand function for all participants was normal with $\mu = 575$ and $\sigma = 100$. Therefore, in the high, medium, and low censorship conditions the mean inventory levels were 507, 575, and 643 respectively. To control for some of the noise across conditions, each participant was yoked with a participant from each of the other conditions (there were 6 total cells in the study design). These yoked participants faced the same sequence of 30 demand instances. In this manner 18 demand sequences were randomly generated before the study and participants were randomly assigned to a condition and a demand sequence. Since each condition contained the same 18 demand sequences, the possibility of differences across conditions emerging as a result of noise in the random demand instances is reduced.

Methods

Participants

One hundred eight undergraduates at a major American university signed up for the study through an online scheduling system and participated in a computer lab. A

stated prerequisite for participation was having completed at least one college-level statistics course. This prerequisite ensured that participants had previously been exposed to normal distributions. They received $5 for participation and could earn up to an additional $5 based on the mean absolute deviation (MAD) of their demand estimates from the true mean demand. For every 15 units of MAD, participants lost $1 from the bonus until it was exhausted. This relationship was linear and individuals could be awarded fractions of dollars.

Instructions and procedures

The user interface was programmed in Microsoft Excel. Participants read an information sheet explaining the details of the game. They were informed in text and figure that the demand distribution was normal with a stationary mean, *m*, and a standard deviation of 100. They were told that the mean, *m*, was equally likely to be anywhere between 400 and 800. Before the commencement of the study, the instructions were reviewed with each participant to ensure comprehension.

Appendix A contains a snapshot of the user interface. Each period, participants observed sales and a binary indication of whether the paper company sold out. They then made an estimate of the underlying mean demand, *m*. This process continued for a total of 30 periods.

Results

Comparisons to truth

The initial analyses were done in a repeated measures model using residual maximum likelihood estimation and an autoregressive covariance structure. Degree of censorship and censorship point variability were specified as fixed effects in the model. An interaction between the experimental manipulations and period was also included in the model. Period was not significant as a main effect or as a moderator ($p > .25$ for each). The average estimates in each censorship condition were significantly lower than the true population mean ($t(102) > 2.3$, $p < .05$ for each). This provides substantial evidence of a censorship bias as predicted in Hypothesis 1. Also, consistent with Hypothesis 2, there was a significant main effect of censorship ($F(2,102) = 3.83$, $p < .03$) such that mean perceived demand declined as the degree of censorship increased (M_{low} = 567.7, M_{medium} = 547.6, and M_{high} = 530.0; $\mu = 575$).

Comparisons to prescriptive heuristic estimate

An alternative analytical approach is to compare their estimates to what an approximate maximum likelihood heuristic would have estimated given the same observed sample (Nahmias 1994). To use this approach, their final estimates (i.e., their estimates with the full sample of 30 observations) were compared to the prescriptive heuristic estimates given the same sample (see Figure 5). The heuristic's estimates were not significantly different from the true mean in any of the degree of censorship or censorship point variability conditions ($t(102) < 1.7$, $p > .1$ for each). However, participants' final estimates were significantly lower than the heuristic's estimates given

the same sample with a high (75%) or medium (50%) censorship ($t(102) = 7.07$ and 4.29 respectively, ps < .001). Final estimates were not significantly lower than the prescriptive heuristic's estimates with a low (25%) censorship ($t(102) = 1.03$, $p > .3$).

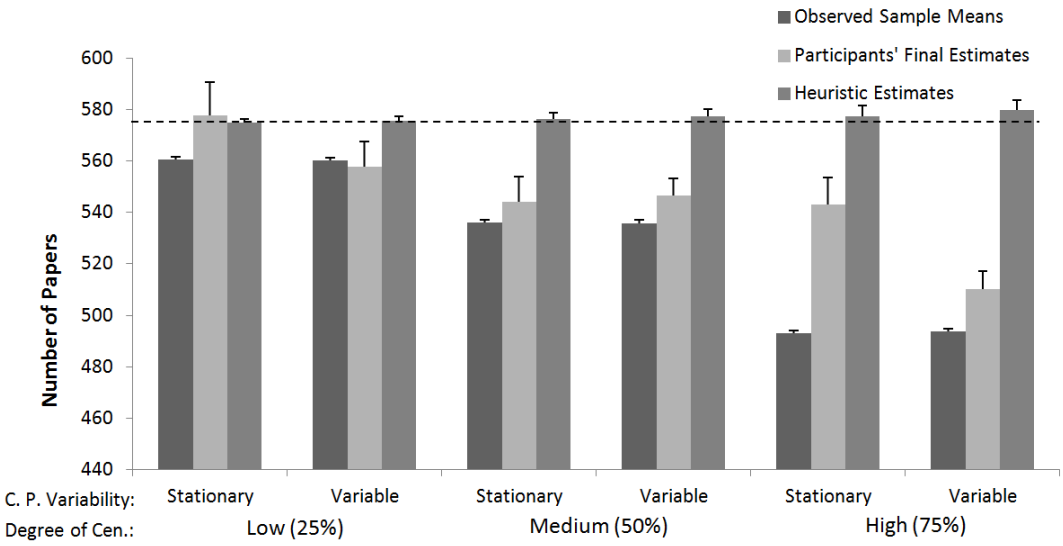

Figure 5: Study 1 mean final demand estimates, observed sample means, and prescriptive heuristic estimates displayed by condition, with standard error bars. The dashed line indicates the true mean demand, 575.

Adjustments from sample mean

While it was predicted that individuals would be naively biased toward the observed sample mean, it was also predicted that they would use cues of censorship to adjust their beliefs. To test this I compared their final estimates to their observed sample mean across conditions. There was no interaction between degree of censorship and censorship point variability. In the high, medium, and low censorship conditions the final estimates were higher than their observed sample means by 33.2, 9.5, and 7.2. They adjusted significantly more with high (75%) censorship than with medium (50%) or low (25%) censorship (ts(102) > 2.39, ps < .02), however the latter two were not significantly

different ($t(102) = 0.22$, $p > .8$). Finally, when individuals faced a stationary, rather than variable, censorship point they adjusted 16.7 units farther from the observed sample mean, a significant difference ($t(102) = 2.07$, $p < .05$). These findings are consistent with the prediction that individuals are partially sensitive to censorship and do use information in the sample to account for the sample bias (Hypothesis 4). Specifically, more censored observations and a stationary censorship point cued greater adjustment from the observed sample mean (Hypotheses 5 and 6).

Study 1 provides support for several predictions. First, estimates of the population mean (mean demand) tended to be downward biased in the censored environment relative to the true mean and the prescriptive heuristic estimate given the same sample. Second, the degree of the censorship bias increased with the degree of censorship. Third, individuals were sensitive to both the degree of censorship and censorship point variability cues causing them to account for the sample bias to a greater extent and adjust farther from the observed sample mean.

Study 2: Judgment and Choice with Known versus Unknown Demand

In Study 2 participants acted as both judges and decision-makers. This study linked biased inference from censored samples to actual decision-making consequences. Second, it explicitly compared judgment in censored versus uncensored environments. Specifically, Study 2 explored how biased inventory orders may result from biased beliefs about demand caused by censorship.

Much behavioral research in the inventory context has focused on the newsvendor ordering task with a *known* demand distribution (e.g., Schweitzer & Cachon 2000). In these studies, participants are informed of a demand distribution and the cost parameters for their good. They are then asked to make a sequence of stocking decisions with incentives to maximize profit. In contrast, this study examines how individuals perform when demand is *unknown*. In Study 2, individuals needed to update their beliefs about demand to inform their inventory decisions in an effort to maximize their expected profit. Based on Study 1, my co-authors and I expected that stocking decisions would be significantly lower when facing censored, as opposed to uncensored, demand observations. However, I expected this stocking bias to be driven by downward biased beliefs about demand and therefore predicted to find this censorship bias only when participants needed to learn an unknown demand, but not when demand was known.

Study 2 was a 2 (environment: censored, uncensored) by 2 (demand knowledge: known, unknown) between-subjects design with within-subject repeated measures for the 30 periods. Censorship was manipulated by allowing some participants to observe only sales each period (censored environment) and others to additionally observe actual demand each period (uncensored environment). Also, I manipulated whether individuals knew the true underlying mean demand. Some participants were told the true underlying mean of demand (known demand), while others were told that their true underlying mean demand was equally likely to be anywhere between 400 and 800 (unknown demand).

In the task, participants purchased newspapers for $1, sold them for $2, and discarded excess inventory at the end of each period at no cost. Participants were told the

overage and underage costs were both equal to one dollar. Given the symmetry of the demand distribution, the optimal policy with known demand was simply to stock the mean of demand. Optimal here assumes expected profit maximization with full information about the demand distribution. The optimal ordering policy without full information is, in fact, even higher due to the value of "stalking information" (e.g., see Lariviere & Porteus 1999).

As in Study 1, participants were yoked across conditions with common demand sequences. Each cell of the design had the same 19 demand sequences, reducing the probability that differences across conditions could emerge due to randomness. Each of the 19 demand sequences were generated from a different underlying demand, with means randomly selected between 500 and 700. Although participants were given a prior of U(400,800) for the selection of the underlying mean, we actually used U(500,700) to avoid ceiling or floor effects that might limit the potential bias in their beliefs.

Methods

Participants

Participants were 76 MBA or business PhD students at an American business school and were randomly assigned to one of the four experimental conditions. All participants had previously taken a basic statistics course and also either an operations management course or an advanced statistics course, making the subject pool quite statistically sophisticated. For each person that participated, $8 was donated to a club or charity of their choice. Further, additional money could be earned based on profit earned in the game that could be either kept by the participant or also donated. For every $2,000

earned in the game they earned $1 in bonus money and fractions of dollars could be earned. In playing the game, participants generally earned $4-9 in bonus money.

Instructions

Participants were given an instruction sheet explaining the details of the game. They were informed that demand was normally distributed with a mean of *m* and standard deviation of 100 and shown a picture of this distribution. They were told that before beginning the game some participants would get to learn their exact *m* while others would learn a range of where *m* might be. They were informed that in either case the mean of demand was stationary and did not change over the course of the study. The cost parameters were then explained and it was explicitly stated that it was just as costly to order one unit too many as it was to order one unit too few.

Procedures

The study was run in a computer lab and was created in a Microsoft Excel interface. At the top of the user interface, participants were informed either of the *m* of their demand distribution (known demand condition) or that the *m* of their demand distribution was randomly chosen with equal likelihood from the range 400 to 800 (unknown demand condition). Each period in the inventory game, participants estimated the underlying mean demand *and* made a stocking decision. Their own stocking decision therefore acted as the censorship point.

See Appendix C for a depiction of the study interface. First, in two columns participants estimated or reported their underlying mean demand and then made a

stocking decision for the period. Sales feedback was then automatically generated and participants in the uncensored condition also observed the actual demand instance that period. The cumulative average of sales (and demand in the uncensored condition) was updated at the bottom of the screen as participants updated demand beliefs, made stocking decisions, and observed sales feedback for 30 periods.

Results

Since each of the 19 demand distributions faced by individuals had a different randomly determined μ, all estimates were analyzed relative to their respective μ. That is, μ was set equal to 0 and estimates greater than their respective μ were positive and estimates less than μ were negative. As in Study 1, analyses were first done with a repeated measures model using residual maximum likelihood estimation and an autoregressive covariance structure. Demand knowledge and censorship environment were specified as fixed effects and an interaction between the experimental manipulations and period was also included in the model. The repeated variable period was not significant as a main effect or as a moderator for demand beliefs or stocking decisions. See Table 1 for a summary of the results.

Table 1: Average estimates of mean demand and stocking decisions by condition. The estimates and stocking decisions are relative to true mean demand. * indicates a significant bias below the true mean ($p < .01$).

Demand knowledge	Censorship environment	Avg. estimates of mean demand		Avg. stocking decisions	
		M	*SD*	*M*	*SD*
Known demand	Uncensored	0.0	0.3	1.7	34.5
	Censored	-1.3	6.0	4.1	16.1
Unknown demand	Uncensored	-4.4	25.7	-7.0	36.1
	Censored	-27.9*	34.6	-31.1*	40.6

Demand beliefs

There was a significant interaction between demand knowledge and censorship environment, $F(1,72) = 11.65$, $p < .002$. When demand was unknown, censoring demand feedback significantly biased demand beliefs downward ($M_{unknown\text{-}censored} = -27.9$ vs. $M_{unknown\text{-}uncensored} = -4.4$), $F(1,72) = 26.17$, $p < .001$. As one would expect, with known demand, censoring demand feedback had no effect on demand beliefs, $F(1,72) = 0.08$, $p > .75$. Further, estimates of the underlying mean demand were not significantly different from the true mean when demand was either known or uncensored ($p > .15$ for each). However, when demand was unknown in a censored environment, average estimates were significantly lower than true mean demand ($M = -27.9$), $t(72) = 8.59$, $p < .001$.

Stocking decisions

The results for stocking decisions were almost identical to those for demand beliefs. There was a significant interaction between demand knowledge and censorship

environment, $F(1,72) = 9.02$, $p < .005$. When demand was unknown, stocking decisions were significantly lower in a censored environment than an uncensored environment ($M_{unknown\text{-}censored} = -31.1$ vs. $M_{unknown\text{-}uncensored} = -7.0$), $F(1,72) = 14.96$, $p < .001$. When demand was known, stocking decisions were not significantly different across censored and uncensored environments, $F(1,72) = 0.14$, $p > .7$. Furthermore, stocking decisions were not significantly different from the optimal level when demand was either known or uncensored ($p > .1$ for each). As expected, when demand was unknown in a censored environment, stocking decisions were significantly lower than optimal**Error! Bookmark not defined.** ($M = -31.1$), $t(72) = 7.07$, $p < .001$.

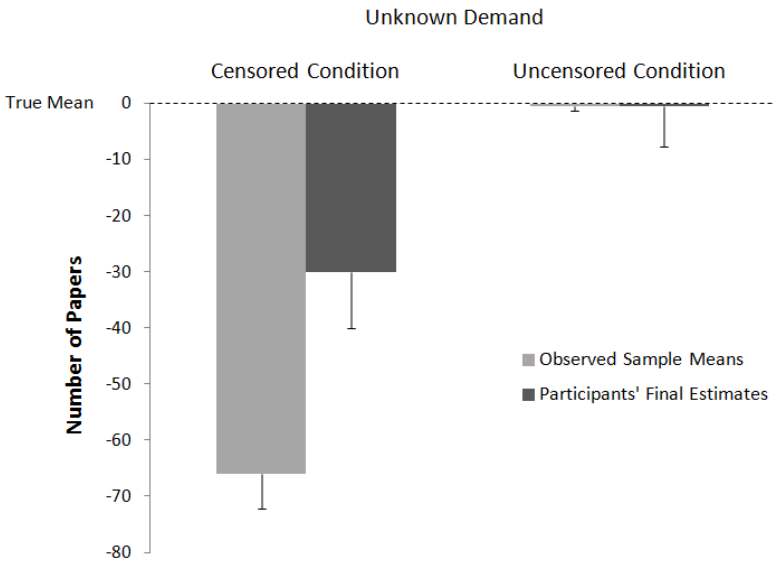

Figure 6: Participants' mean final estimates of mean demand and mean observed sample means with standard error bars displayed by censorship condition. Data are from the unknown demand conditions only. Estimates are shown relative to true mean demand.

Adjustments from sample mean

Next, it was tested whether individuals made estimates of mean demand higher than the mean of their final observed sample (Hypothesis 4). With censorship, both participants' final estimates of mean demand and their final stocking decisions were significantly higher than their 30-period observed sample mean in both demand knowledge conditions (unknown demand: $\bar{x}$ = -66.1, final estimate = -30.0, final stock = -28.9; known demand: $\bar{x}$ = -41.6, final estimate = -0.8, final stock = 6.7; *t*s(72) > 4.5, *p*s < .001). This evidence suggests that individuals were not purely naïve and did significantly account for censorship by adjusting above the sample mean in their estimates of the population mean.

In summary, censorship significantly biased estimates of mean demand below the true mean when individuals needed to infer an unknown demand. Further, stocking decisions were unaffected by censorship when demand was known, but when demand needed to be inferred, censorship caused stocking decisions to be significantly lower. In other words, when individuals formed an accurate understanding of demand—with known demand or with unknown demand in an uncensored environment—their stocking decisions were approximately optimal on average. However, when they formed a biased belief about demand, their stocking decisions were consequently also biased. See work by Rudi & Drake (2009) for some evidence that in some cases censorship may bias stocking decisions even with known demand.

Study 3: Debiasing with Heightened Awareness of Missed Sales

Study 3 examined whether the censorship bias could be mitigated by cueing decision makers to consider the unobserved magnitude of censored observations (i.e., missed sales after a stock-out). Such cues may facilitate greater accounting for the sample bias.

Previous research on counterfactual thinking has found that causing individuals to consider possible alternative outcomes to a past event can improve learning (Roese 1997). Further, consideration of better possible alternative outcomes has been shown to yield more immediate pain but improve learning and future decision making (Reb & Connolly 2009; Roese and Olsen 1995). Similarly, many studies have found that encouraging people to "consider the opposite" improves decision making (Hirt & Markman 1995; Larrick 2004).

To heighten decision makers' awareness of censorship and missed sales, participants were asked to estimate actual demand in each period that they sold out. When a decision maker's awareness of missed sales is heightened, they are more likely to account for a bias in their observed sample of sales. Therefore, if participants are asked to estimate missed sales after a stock-out they are likely to be less susceptible to the censorship bias and form more accurate beliefs about demand. Study 3 tested this prediction.

Methods

Study 3 compared stocking decisions and final demand beliefs across three conditions: uncensored (observable missed sales), censored (unobservable missed sales), and censored with debiasing instructions (unobservable missed sales with estimates of missed sales). Participants faced unknown demand and made stocking decisions for 30 periods. Participants were all randomly assigned to face a true mean demand of 500, 600, or 700 with SD = 100.

Participants

One hundred forty-seven daytime MBA students from a major American business school were randomly assigned to a censorship condition. Every participant had already completed at least one probability and statistics course and one operations management course. In exchange for their participation a donation of $5 was made to a club or charity of their choice and participants could earn additional money based on the amount of profit they earned in the study, which they could have for themselves or also donate.

Procedures

An instruction sheet informed participants that they would be running a fictional newspaper stand for one month and each day they needed to buy newspapers for $1 each to stock in their stand and sell for $2. At the end of each day any excess newspapers would be discarded at no cost. After making an ordering decision they would then observe how many newspapers they had sold that day and the profit they had earned. In

the uncensored condition, participants were also able to observe the actual demand each day.

They were informed that demand was normally distributed with a stationary mean somewhere between 400 and 800 and a standard deviation of 100. Finally, the instructions stated that the goal of the game was to maximize profit and for every $2,000 earned in the game they would earn $1 in bonus money for themselves.

The study was programmed in a user interface created in Microsoft Excel. Participants entered a stocking decision for the day and then sales feedback was generated automatically in an adjacent column based on a demand instance randomly drawn from the demand distribution. In the uncensored condition, participants also observed the actual day's demand. Participants in the debias condition were asked to enter what they believed demand was each day after observing sales. This process was repeated for 30 periods.

After receiving sales feedback for the 30th period, participants were asked to enter their best estimate of the underlying mean demand. They received an extra $1 if their estimate was within 10 units of the true underlying mean demand.

Results

The key dependent variables in the analyses were participants' final estimates of mean demand and their final stocking decisions. The respective true underlying means of demand (500, 600, or 700) were subtracted from each participant's estimate of mean demand and final stocking decision.

Table 2: Final estimates and stocking decision relative to true mean demand in Study 3.

Censorship environment	Final estimates of mean demand		Final stocking decisions	
	M	*SD*	*M*	*SD*
Uncensored	-0.9	23.4	1.7	34.5
Censored	-38.2	40.7	-36.8	46.3
Censored with Debias	-12.5	30.9	5.7	32.6

Censorship significantly affected the estimates of mean demand, $F(2,138) = 17.0$, $p < .001$. Replicating previous studies, demand estimates in the censored condition ($M = -38.2$) were significantly lower than estimates in the uncensored condition ($M = -0.9$), $F(1,138) = 32.60$, $p < .001$. Estimates in the censored condition were significantly lower than the true mean demand, $t(48)=6.57$, $p < .001$, whereas estimates in the uncensored condition were not, $t(48) = 0.28$, $p = .78$.

Another planned contrast found that the estimates of mean demand were significantly higher in the debias condition ($M = -12.5$) than in the censored condition, $F(1,138) = 15.02$, $p < .001$. However, the demand estimates in the debias condition were still lower than those in the uncensored condition, $F(1,138) = 3.37$, $p < .07$, and true underlying mean demand, $t(48) = 2.84$, $p < .01$.

A similar pattern emerged in final stocking decisions. The stocking decisions in the censored condition ($M = -36.8$) were significantly lower than in the uncensored

condition ($M = 5.7$), $F(1,138) = 25.18$, $p < 001$. Final stocking decisions were significantly lower than optimal in the censored condition, $t(48) = 5.57$, $p < .001$, but not in the uncensored condition, $t(48) = 1.23$, $p = .22$.

Final stocking decisions in the debias condition ($M = 12.9$) were significantly higher than those in the censored condition, $F(1,138) = 34.37$, $p < .001$, but not significantly different from stocking decisions in the uncensored condition, $F(1,138) = 0.71$, $p = .40$.

One oddity in the results was that, although estimates of demand in the debias condition were below true mean demand, the final stocking decisions were actually above true mean demand. It is unclear why participants in this condition chose to stock more relative to their estimates of mean demand. It is possible that participants were attempting to make the debiasing task easier for themselves. By reducing the likelihood that they would sellout, they reduced the difficulty of estimating each days demand after observing sales. However, in summary of the Study 3 results, heightening the decision maker's awareness of missed sales partially, but significantly, mitigated the censorship bias. I will return to the implications of demand underestimation in Chapter 6.

Chapter 3: Biased Judgment of Risk

Decision making in the face of risk is a central part of managing a firm (Shapira 1995). However, the degree of risk in the environment is often unknown to managers; instead, they must learn the riskiness from their experiences (Hertwig, Barron, Weber, & Erev 2004). Traditional studies of risky choice provided clearly defined probabilities for possible outcomes and observed whether individuals chose, for example, a certain $5 or a .6 chance of $10 (e.g., Tversky & Kahneman 1981).When probabilities of risky options are explicitly defined, choice is simply driven by risk preferences. In the real world, the probabilities of possible outcomes are usually unlabeled and instead must be learned from experience. When learning from experience, risky choices are driven not only by risk preferences, *but also by risk perceptions* (Sitkin & Pablo 1992; Sitkin & Wiengart 1995). This chapter examines how censorship affects decision making when learning from experience by biasing the perceptions of risk.

Study 4: Risk Perceptions in Sequential Risk-taking

The empirical work in Chapter 2 focused on inferences about the mean of a normal distribution in the domain of learning demand from sales. Study 3 extends these findings in two ways: (1) it tests the censorship bias in a different context, sequential risk-taking (Pleskac 2008; Wallsten, Pleskac, & Lejuez 2005), and (2) it examines inferences about a uniformly distributed population, as opposed to a normally distributed population. Sequential risk-taking consists of a repeated risky choice in which risk changes dynamically and systematically. The risky choice initially has a high probability

of a positive payoff (or *gain*) that continues play and a small probability of a negative payoff (or *bust*) that ends play. After each risky choice that results in a gain, the individual may choose to take another risk in hopes of yet another gain or to quit the round to avoid a potential bust. This decision depends on two critical factors: the risk preference of the decision maker, and the perceived probability of a bust. In this study I examine how censorship affects the latter of these two factors.

Censorship occurs in sequential risk-taking when individuals cannot observe the precise turn on which they would have busted if they had continued taking more risks. With censorship, individuals cannot observe how many total gains they *could have* achieved before busting. On the other hand, in rounds that do end in a bust, they can observe exactly how many gains they were able to achieve before busting. This asymmetry generates censored samples which may cause individuals to underestimate the mean number of risks they can take before a bust.

Clinical psychologists have used sequential risk-taking tasks to simulate risk taking with drugs, alcohol, and other public health risks. Behavior on these laboratory tasks has been shown to correspond well with risk-taking behavior in the real-world (Lejuez et al. 2002, 2003; Wallsten et al. 2005). Sequential risk-taking also occurs in organizations, such as when setting replacement policies (see Chapter 1; Campbell, Whitehead, & Finkelstein 2009). Social dynamics involving potential conflict may also involve sequential risks. For example, an ambitious negotiator may continually fight for as much value as possible at the bargaining table while risking offending the opposing party and ruining a potential deal altogether. Similarly, managers must promote

disagreement and the sharing of diverse opinions among their employees while risking the emergence of interpersonal conflict. This is one of the biggest challenges facing managers today (Behfar, Peterson, Mannix, & Trochim 2008; Jehn, Northcraft, & Neale 1999; Simons & Peterson 2000). The present study examines how individuals infer risk from the outcomes of their sequential risky choices.

The Task

The sequential risk-taking task was adapted from Pleskac's (2008) Angling Risk Task[1]. Participants played a fishing game in which they made 25 trips to a pond (see Appendix D for a picture of the task). Participants were informed that all trips began with the same number of fish in the pond (N), but were not told the precise value of N. One of the fish in the pond was blue and the rest, $N - 1$, were red. Catching the red fish gave them money (i.e., a gain) but catching the blue fish ruined that trip (i.e., a bust). Participants were given a uniform prior for N between 5 and 45. In reality, for all participants the number of fish in the pond was set to 24.

Each trip to the pond entailed the following process. Participants could use a 'Cast and catch a fish' button to randomly catch one fish from the pond. Each time they caught a red fish, 5 cents was placed into a temporary Trip Bank. If they ever caught the blue fish, they lost the money from their Trip Bank and that trip ended. However, at any point after catching a red fish they could use a 'Quit trip and collect' button to end the trip and

[1] The censored condition in this study was very similar to Pleskac's (2008) 'Catch and keep' fishing tournament with cloudy conditions.

move their Trip Bank money to their Permanent Bank. This is how they earned money in the game.

Each trip started with the same number of fish in the pond (24 fish). The chance of catching the blue fish on the next cast increased with each red fish caught, since there was then one fewer red fish in the pond (i.e., sampling without replacement). Also, the amount of money that they would forfeit with a bust increased as they accumulated red fish on a given trip. At the end of each trip, the number of fish in the pond was always reset to 24 and always with only one blue fish. After 25 trips, the game ended and participants could keep the money in their permanent banks. At this point, I asked participants to estimate N, the number of fish in the pond at the start of each trip. In this task busts followed a discrete uniform distribution between 1 and N (or in this case 1 and 24). Therefore, participants essentially estimated the maximum of the uniform distribution of blue fish.

This environment is censored: in trips where participants chose 'Quit trip and collect' to cash in their trip earnings, they could not observe how many more red fish they could have caught before busting. This served as the censored condition. In the uncensored condition, after they quit a trip to collect their trip money, I simulated more casts and told them on which cast they would have caught the blue fish if they had continued casting. Therefore, when uncensored, at the end of each round the participant could observe on which cast the blue fish was or would have been caught. Participants were aware of both possible conditions.

Methods

Participants

The participants were 39 undergraduate students at an American university. They were randomly assigned to a censorship condition. Participants earned a base pay of $4 plus bonus money based on performance in the task. The bonus money was comprised of the money earned from the fishing game and an additional dollar if their estimate of *N* was within 1 of the correct answer. Most participants earned $8-15.

Procedures

Participants were given an instruction sheet that explained the rules of the game as described above. They were not informed of the estimation task (guessing the number of fish in the pond) until after the fishing game was completed. The task was programmed in Authorware and the study was run on computers in a lab. The permanent bank, temporary bank, trip number, result of the most recent cast, and a count of red fish caught on the current trip were displayed on the screen (see Appendix D). When a participant caught the blue fish, a notice popped up showing them the blue fish, the number of casts they made that trip, and the amount of money that they forfeited from their temporary bank. Between trips, participants were reminded that the number of fish in the pond had returned to *N* with one blue fish. When participants clicked 'Quit trip and collect,' they were shown how many casts they made on that trip and how much money they earned. In the uncensored condition, the program simulated more casts and showed them on which cast they would have caught the blue fish had they continued casting. In the censored condition they did not observe this simulation of additional casts.

After the 25th trip, participants were asked to estimate *N*, the number of fish in the pond at the beginning of each trip. Subsequently, I asked them to imagine that they were going to play the game again, but this time with a *known* number of fish in the pond. However, instead of making individual choices, they would need to set a "decision policy" of how many times they would cast before clicking 'Quit trip and collect.' I asked how many casts they would want to do before quitting the trip if they knew that there were 24 fish in the pond (with one blue fish). Participants were then paid based on their permanent bank status and the accuracy of their estimate of *N* (see specific incentives above).

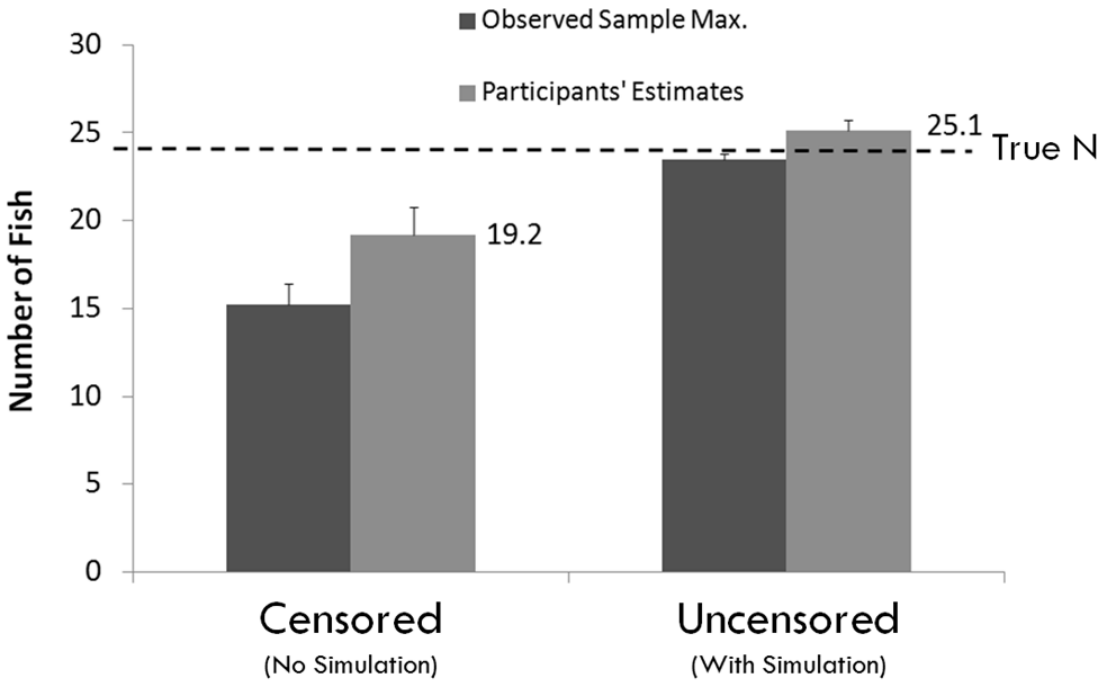

Figure 7: Study 3 mean participant estimates of the number of fish in the pond (N) and the mean maximum number of fish observed by participants displayed by censorship condition, with standard error bars.

Results

The true total number of fish in the pond was 24. Individuals in the censored environment on average estimated that there were 19.2 total fish in the pond (SD = 7.2) while the average estimate of those in the uncensored environment was 25.1 (SD = 2.8), a significant difference between conditions, $F(1,37) = 11.06$, $p < .01$ (refer to Figure 7). The former was significantly lower than the true number of fish in the pond, $t(37) = 3.91$, $p < .001$, but the latter was not, $t(37) = 0.84$, $p > .4$. As an indication of their decision policy for when to stop taking risks, I examined the average number of casts made in trips where participants quit the trip before a bust (Pleskac 2008). Individuals averaged 23% fewer casts in these trips when censored than when uncensored (9.4 vs. 12.2 casts) $F(1,37) = 8.89$, $p < .01$. Overall, including busts, individuals averaged fewer casts when censored than when uncensored (7.9 vs. 9.6), $F(1,37) = 6.96$, $p < .02$. Further, when facing censorship, individuals caught fewer blue fish (11.5 vs. 13.6 blue fish), $F(1,37) = 3.7$, $p < .07$, but earned less money across the 25 trips ($5.90 vs. $6.75), $F(1,37) = 2.47$, $p = .12$, because they quit earlier, although these differences were not significant at the $\alpha = .05$ level.

Individuals across conditions demonstrated very similar preferences, had they known that there were 24 fish in the pond. Individuals in the censored and uncensored conditions on average indicated that they would have set a policy of casting 13.3 and 12.9 times before quitting, respectively (not significantly different, $p > .7$). For individuals in the censored environment, their self-reported preferred policy with 24 fish was to take significantly more risks than the level of risk revealed in their actual behavior in the game

(9.4 casts), $t(37) = 4.16$, $p < .001$. However, for individuals in the uncensored environment, their self-reported preferred policy with 24 fish was not significantly different from their behavior in the game (12.2 casts), $t(37) = 0.77$, $p > .4$.

Comparisons to observed sample maximum

I also compared the censored estimates to a purely naïve estimate given the same sample. The naïve estimate, e^n, in this case was the maximum number of fish observed in any round. If the trip with the highest number of caught fish did not result in a blue fish (e.g., if the round with the most caught fish involved quitting a trip after 15 casts) then I set the naïve estimate to be one greater than the number of fish caught on that trip (e.g., $e^n = 16$) because the naïve judge knows that there is at least one more fish. In the censored condition, individuals significantly adjusted from their maximum number of fish observed in their estimates of the total number of fish in the pond, $M_{samplemax\text{-}censored} = 15.2$ ($SD = 5.4$) vs. $M_{estimate\text{-}censored} = 19.2$, $t(37)=4.87$, $p < .001$ (refer to Figure 7).

In sum, censorship caused individuals to take fewer risks and to underestimate the total number of fish in the pond. This decrease in risk taking caused them to catch fewer blue fish (i.e., achieve fewer busts), but also to earn less money than individuals playing without censorship. Had they known that there were 24 fish in the pond at the start of each trip, they reported that they would have preferred to take more risks. In other words, the censorship bias caused individuals to take fewer risks than they actually would have liked to take. Lastly, although censorship caused individuals to underestimate the number of fish in the pond, individuals were less biased than a purely naïve estimate, suggesting that they did partially account for censorship.

Past research has shown that overly negative perceptions often persist because decision makers abandon options that have resulted in a poor outcome in the past, even if that outcome was an uncharacteristically poor result for that option (Denrell 2005, 2007; March 1996). However, this study demonstrated that even when the decision maker continues to resample, censorship can cause individuals to "learn" to overestimate risk and behave in an overly risk-averse way. Incorrectly learning to overestimate risk may cause managers to under-explore uncertain avenues for firm growth (March 1991; Levinthal & March 1993), over-eagerly replace and upgrade technology (Campbell, Whitehead, & Finkelstein 2009), and behave too conservatively in social exchanges to avoid interpersonal conflict (Simons & Peterson 2000).

Chapter 4: Biased Judgment of Employee Capability

In many cases constraints limit what managers can see about their environments and, as I have demonstrated, this inability to see the whole picture causes them to make crucial misjudgments. In this chapter I explore how constraints in the workplace may limit what managers can see about employees' capabilities.

A 2005 Gallup poll revealed that 55% of U.S employees felt like their time and capabilities were underutilized at work. In 2011, 71% of surveyed American workers stated that they were either "not engaged" or "actively disengaged" at work. (Joyce 2005; Blacksmith & Harter 2011). Work boredom leads many employees to spend work hours playing games, reading, and socializing online. Why are employees so underutilized at work? I suggest that a major contributing factor to underutilization of employees is that managers are forced to infer what their employees are capable of from censored performance outcomes. This censorship causes managers to incorrectly "learn" over time to underestimate their employees. This chapter sheds light on how the interaction between the structure of the work environment and the cognitive limitations of managers leads to biased beliefs about employees.

In the last several decades, the tradeoff of efficiency versus flexibility in organizations has been a focal topic in the management literature (Kang, Morris, & Snell 2007; Levinthal & March 1993; March 1991). Much of this work has pointed to resource slack as a critical component of organizational learning and exploration (Bourgeois 1981; McGrath 2001; Nohria & Gulati 1996; Sitkin, Sutcliffe, & Schroeder 1994). However, past research has largely neglected that work allocations in organizations also affect the

learning of employee capability. Managers must not only make tradeoffs between efficiency and flexibility but also consider how their decisions affect what they can learn about their employees' capabilities. The purpose of this chapter is to consider how work allocations and deadlines act as constraints on what managers can observe and how these limitations affect managers' perceptions of employees. Specifically, I propose that there is an inherent information asymmetry in what managers can observe about employee capability. While unmet employee capability is difficult to see and often ignored, employee failings are observable and salient. That is, managers can easily observe errors of too-high expectations (i.e., over-assigning work) but cannot clearly observe errors of too-low expectations (i.e., under-assigning work).

I interviewed a sporting goods store manager of a national chain and an accountant at one of the largest non-profit organizations in America. In speaking about his employees, the store manager said, "People have their assignments and they do them. To be honest, I only really pay attention if somebody screws up." In speaking about her managers, the accountant said, "They have no way of seeing what I'm doing or when I could be doing more. My direct boss is actually in [another state]." These anecdotes illustrate how this information asymmetry can emerge in organizations. An employee's failures receive more attention and are more easily observed than an employee's unused potential. Over time, such an asymmetry may lead to the underestimation of employees.

To treat the problem more concretely, consider the following example. Imagine that a manager allots 4 weeks for a team to complete a project. If the project actually takes the team 6 weeks to complete, then the manager will see exactly how long it took

the team: the project is submitted 2 weeks late (see case 1 in Figure 8). If the team is actually capable of completing the project in 2 weeks, the manager will only observe that the work is completed by the deadline (see case 2 in Figure 8). Given the team's capability, this time allotment is generous and they can work intermittently, at a slower pace, or procrastinate and complete the work in the final two weeks. While the manager can clearly see the exact magnitude of the failure of the slow team, the excess, untapped potential of the fast team goes unobserved. In this manner, deadlines, or time allotments, can censor observations of the actual time needed to complete a task. In this example, I would predict underestimation of how quickly employees can complete work. Or in other words, managers will believe that employees need more time to complete work than they actually do.

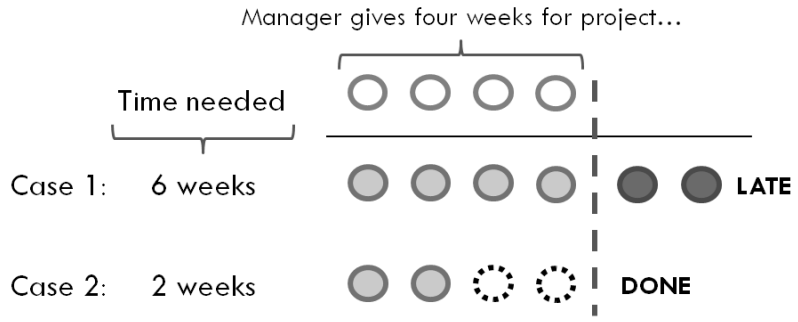

Figure 8: How deadlines censor employee capability.

Some companies recognize the asymmetry in what they can observe about employee capability and try to overcome it. For example, Intel, an American multinational semiconductor chip maker corporation, tries to reduce censorship of employee capability by expecting a certain failure rate in their employee's performance.

Like many companies, Intel requires employees to set objectives at the beginning of the year. Although employees are later rewarded for completing objectives, over the years, employees are expected to meet only 80% of their objectives. If an employee meets their objectives too frequently, it is taken as a sign that they do not set sufficiently ambitious objectives and Intel's management raises the expectations for that employee. While this solution may be vulnerable to gamesmanship—an employee could set a few impossible objectives or intentionally leave a few objectives unfinished—it clearly indicates that Intel recognizes the censorship problem. As a consequence of Intel's tolerance and expectation of failure, employees set more ambitious objectives and the managers are better able to learn what their employees are capable of.

Considerable past work has examined the evaluation of employees through formal measures (e.g., Becker & Gerhart 1996; Denrell, Arvidsson, & Zander 2004; Hogan & Hogan 1989; Ilgen, Barnes-Farrell, & McKellin 1993) but little work has been done to systematically understand how the structure of the work environment shapes what managers learn—or incorrectly learn—about their employees. One of the most fundamental jobs of a manager is to maximize the productivity of an organization. To this end, accurate perceptions of employees' capabilities are critical (Becker & Gerhart 1996; Manzoni & Barsoux 2002). Beliefs about different employees' capabilities are the foundation for many organizational decisions such as work assignment, promotion, compensation, and dismissal decisions. I examine how the work environment shapes what a manager can observe and how a failure of managerial cognition results in predictably biased beliefs about employees.

When to Predict Underestimation of Employees

Clearly not all employees are underestimated by their managers. Centrally, there must be a constraint that limits what a manager can observe about employee capability, either in the form of deadlines or work assignments. When deadlines and work assignments are too conservative, employees are able to easily complete the work and managers are unable to see how quickly or how much more they could have done. On the other hand, when deadlines and assignments are overly ambitious, employees may complete assignments late or be forced to compromise on quality. These failures are readily observable by the manager. In this way, deadlines and work assignments create an asymmetry in what managers can observe about employee capability. Further, there are three key elements that determine the likelihood that, and degree to which, employees are underestimated due to the censorship bias: (1) the degree of constraint, (2) employee disclosure of unused capability, and (3) managerial awareness of censorship. I will discuss these in greater detail in the following chapter.

The first element that determines the likelihood of employee underestimation is the degree of constraint. The degree of constraint will depend on how ambitious or conservative deadlines and work assignments are, which is largely determined by how the firm values employee failure versus employee productivity. In many cases, it is costly for a firm to over-promise and under-deliver to clients. Further, when work is interdependent it may be costly internally to not have work completed when expected. As the cost of failing to complete work increases, less work will be assigned to employees to decrease the probability that they fall short; doing so, however, increases the degree of

constraint and, consequently, the likelihood of employee underestimation. Conversely, when an individual's labor is more valuable, firms give her more ambitious assignments so as not to miss out on such valuable potential productivity. More ambitious assignments reduce censorship and underestimation of employees.

The second element that determines the likelihood of employee underestimation is whether employees choose to self-disclose their unused potential up the organizational hierarchy, which would uncensor their capabilities. Employees may (e.g., in response to incentives) or may not (e.g., due to social pressure from other employees) choose to disclose their unused capability to managers. In this way, the willingness of employees to pursue greater amounts of work also determines the likelihood that a manager underestimates them.

The third element that determines the likelihood of employee underestimation is managerial awareness and consideration of unused capability. To form accurate beliefs about employees' capabilities, managers must sufficiently account for unused capability in the event that employees are able to complete assignments. Therefore, factors that facilitate (hinder) managerial awareness of unused capability will reduce (increase) the likelihood of employee underestimation.

Next, I present three studies that empirically explored how underestimation of employees can emerge with censorship and which types of employees are more likely to be underestimated. Then in the subsequent chapter, I will return to these determinants of employee underestimation and delineate a set of factors that affect the likelihood that, and degree to which, employees are underestimated.

Study 5: Deadlines and Employee Variance

Study 5 examined judgment of employee capability when a deadline censors how long it takes the employee to complete work. It simulated a scenario in which a manager allots a certain amount of time for the completion of a task and returns at the end of the allotted time (i.e., the deadline) to collect the work. If the work is finished, the manager faces a censored observation of the employee's capability; the manager knows only that the employee was able to do the task in the allotted time or less. If the work is unfinished by the deadline, the employee continues working and submits the completed task late. In this case, the manager faces an uncensored observation of the employee's capability; they can observe exactly how long it took the employee to complete the work. The use of deadlines as censorship points provides an example of a left-censored environment. In this case, the censorship bias should cause over-estimation of the average time needed to complete a task.

This study tested Hypothesis 3 (Chapter 1): the censorship bias is exacerbated for higher variance populations. With higher variance in completion times, employees have more extreme good and bad periods. However, when completion times are censored by deadlines, managers can only observe the more extreme bad completion times, while the more extreme good completion times are censored and unobserved—the work is simply done by the deadline. Therefore, higher variance will only be visible to the manager in periods where the employee completes the work late (see Figure 9). Higher variance therefore causes the observed sample to be less representative of the employee's true capability. The mean of the observed performance moves farther from the mean of the

employee's true capability. This will lead to greater overestimation of the amount of time needed to complete a task.

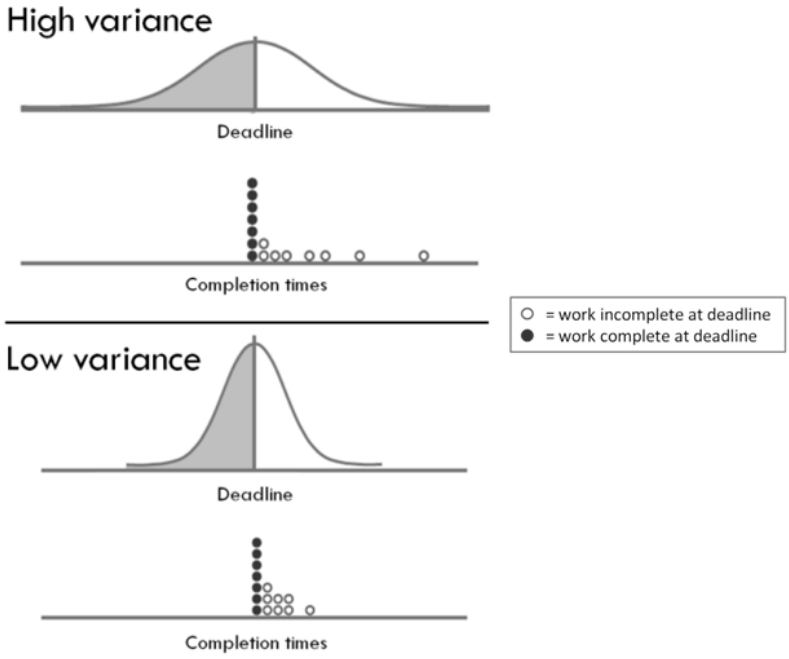

Figure 9: Depiction of the observed outcomes for employees with high and low variance capabilities when completion times are censored by a deadline.

In this study, environment was a between-subjects factor (environment: censored, uncensored) and employee capability variance was a within-subject factor with each participant making judgments of 20 employees who naturally varied in their capability variance.

Methods

The Task

In this study I examined whether censorship can bias an individual's perception of an individual's performance capability. The study occurred in two phases. In phase one, a

set of 20 participants each did 30 word search puzzles as fast as they could, one at a time, and their completion times were recorded for each puzzle. This set of times served as their distribution of times to complete a single puzzle. I then randomly drew five puzzle times with replacement from the set of 30 times, which represented that participant's performance for a single period in the study. The purpose of using sets of five puzzles was to make each person's performance somewhat normally distributed. In this manner, 40 periods of performance results were created for each participant, as five individual puzzle times were randomly drawn for each period. These participants were the "employees" that would be judged in the second phase of the study. Employees differed in their average performance and in the variance of their performance.

In the second phase of the study, participants acted as managers that observed employees performances and needed to make a judgments about their capabilities. Managers were told that employees had done a new task in each period and each task was made up of a set of five puzzles. The managers (participants in phase 2) needed to learn how long it took the employees (participants in phase 1) to complete a set of five puzzles on average based on the performance times they observed.

There were two environment conditions: censored and uncensored. To simulate capability censorship that a manager might experience in the real world, I presented the employee performance times with a censoring deadline which was equal to the mean time for each employee. The manager knew that the employee had done the task as fast as he could, however, in periods in which the employee did the task faster than the deadline, the manager could only observe that the employee was finished by the deadline. On the

other hand, when an employee took longer than the deadline, the manager could observe that he had failed to finish by the deadline and saw the exact time that it took the employee to finish the task (see Appendix D). Setting the censorship point at the mean for each employee created a censorship rate of approximately 50% for each, while the capability variance differed naturally across employees. In the uncensored environment, the manager could observe the employee's completion time in each period.

Managers observed employee completion times and made judgments about how long it takes them to complete the task on average. I predicted that participants in the censored environment would underestimate the capabilities of the employees by overestimating the amount of time it takes them to complete a task. Further, I predicted that higher variance in employee capability would exacerbate the bias, consistent with Hypothesis 3. In the uncensored environment, I expected unbiased estimates of employee average capabilities.

Procedures

Forty-eight individuals participated in phase two of the study. Participants were paid a base rate for participation and could also earn bonus pay. For each participant, one employee was randomly selected at the end of the study. The participant's bonus was determined by the accuracy of their estimate for that one employee, ranging from $5 for being within 1 second of the true mean to $0 for being more than 21 seconds from the true mean.

The study was conducted at a computer laboratory and the estimation task was done with paper and pen. Individuals received a packet of employees' performance

outcomes and an answer sheet on which to make their estimates. In the employee performance packet, each page contained 40 periods of performance outcomes for a single employee and the packet contained performance outcomes for 20 employees. The order of the employees in the packet was randomized.

Participants were told that in each period employees did the task as fast as they could. In the censored condition, it was explained to them that they would observe the employees' performance with respect to a certain deadline—when the employee finished the task before the deadline, the participants would simply see that they had finished. When employees finished late, they would be able to see exactly how long it took them.

For each of the 20 employees, participants estimated how long on average it takes that employee to finish a task when the employee goes as fast as they can. The instruction sheet for the censored environment said explicitly, "You should not simply average the times that you can observe because in some periods the employee was able to finish before the deadline." After making an estimate for each employee, participants were asked in an open-ended follow-up question to explain their strategy for estimating an employee's average capability. The bonus pay was then calculated based on the accuracy of their estimate for a randomly selected employee, as described above.

Results

Employees had their own mean times needed to complete the tasks, which the participants tried to estimate. The average of these means across employees was 84 seconds (SD = 23.1). The average standard deviation within an employee was 26.4 seconds (SD = 9.5).

The data were analyzed using residual maximum likelihood estimation with censorship environment as a between-subjects factor and employee capability variance as a within-subjects factor. For each observation, the employee's actual mean capability was subtracted from the participant's estimate. This was the dependent variable in the analyses. Positive (negative) values indicated that the participant believed the employee's mean time was greater (less) than the employee's true mean time. See Figure 10 for a summary of the results.

Comparisons to truth

There was a significant main effect of environment, $F(1,46) = 305.6$, $p < .001$. The participants in the censored environment overestimated the mean time needed to complete the task by an average of 9.0 seconds ($SD = 9.1$), $t(46) = 26.7$, $p < .001$. The estimates of participants in the uncensored environment were not significantly different from the true mean time ($M = 0.1$, $SD = 7.0$), $t(46) = 26.7$, $p < .001$. Further, there was a significant interaction between environment and employee capability variance, $F(1,910) = 29.3$, $p < .001$. In the censored environment, estimates of employee capability were significantly more biased for high-variance than low-variance employees (censored: $M_{low} = 6.7$, $M_{high} = 11.2$), $F(1,46) = 42.09$, $p < .001$. In the uncensored environment, employee capability variance did not affect estimates of mean employee capability (uncensored: $M_{low} = -0.2$, $M_{high} = 0.4$), $F(1,46) = 0.55$, $p > .45$.

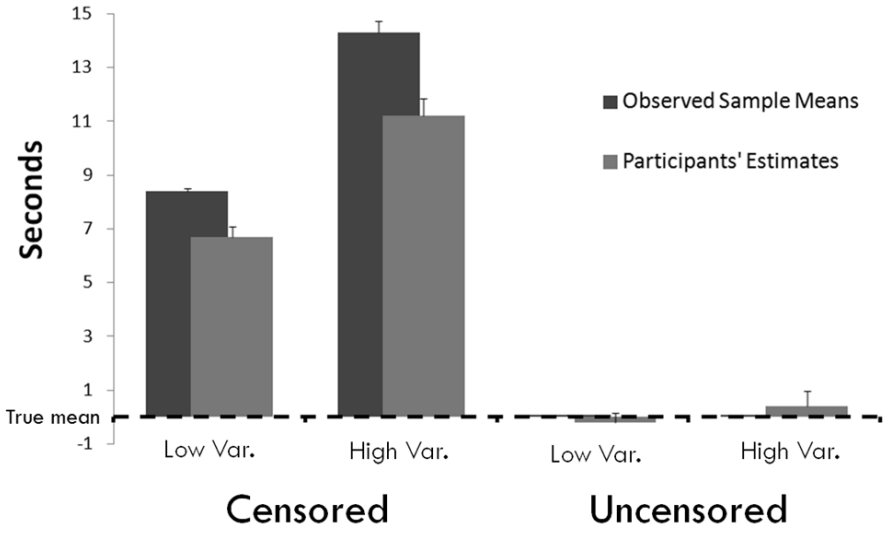

Figure 10: Participants' mean estimates of mean employee capability and mean observed sample means by censorship environment and employee capability variance, with standard error bars.

Adjustments from sample mean

Next, I tested whether participants adjusted from the mean of their observed samples when making estimates of mean employee capability. The sample mean is the average of the times observed for each employee, including censored observations registered at the censorship point. Consistent with Hypothesis 4, I predicted that individuals would not be purely naïve and would partially account for unobserved capability. I expected that individuals would significantly adjust their estimates from the mean of their observed censored samples. As predicted, in the censored environment individuals adjusted significantly from their observed sample mean in their estimates of mean employee capability ($M = -1.8$, $SD = 8.9$; $t(15) = 5.29$, $p < .001$).

Ranking accuracy

Many compensation and promotion decisions rely on a manager's ranking, or relative perception, of employees. To examine how censorship and employee capability variance affect employee ranking, I inferred participants' ranking of employees based on their estimates of each employee's average capability. Each participant's ranking of employees was then compared to the true ranking of employees based on their actual mean capability.

I examined if the number of instances of a mismatch between perceived rank and actual rank differed significantly across censorship environments. On average there were 2.2 more ranking errors per participant in the censored environment than in the uncensored environment, $F(1,46) = 12.3$, $p < .01$. Next, I tested whether capability variance affected over- or under-ranking of employees. There was a significant interaction between censorship environment and employee capability variance, $F(1,941) = 20.60$, $p < .001$. Using a median split on employee capability variance, I found that in a censored environment, higher variance in capability significantly increased the degree to which an employee was ranked worse than their true ranking, $F(1,46) = 16.16$, $p < .001$. Employee capability variance had no significant effect on rankings in the uncensored environment, $F(1,46) = 3.67$, $p > .05$.

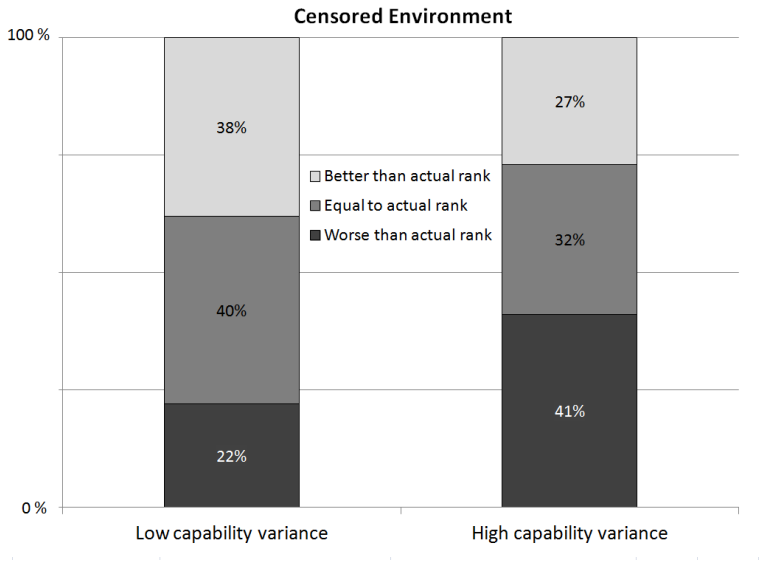

Figure 11: An illustrative graph of participants' ranks of employees relative to their true rank in the censored environment by a median split on employee capability variance.

Figure 11 provides some insight into the nature of this ranking effect in the censored environment. In some of these cases, an employee was misranked by more than one spot. In the censored environment, employees with high capability variance were ranked worse than their true ranking 41% of the time and better than their true ranking only 27% of the time. Employees with low capability variance were ranked worse than their true ranking only 22% of the time and better than their true ranking 38% of the time. These descriptive results illustrate the nature of a prejudice against high variance employees in censored environments. Higher capability variance exacerbates the censorship bias causing high variance employees to be underrated by their managers relative to low variance employees.

This study demonstrated how deadlines can left-censor completion times. Managers can easily observe when an employee completes work late (i.e., when a

deadline is too ambitious), but often cannot observe how much earlier an employee could have completed work that is submitted by the deadline (i.e., when a deadline is to generous). Further, an employee that has higher variance in their capability was shown to be underestimated more. Because higher variance leads to greater underestimation, in some cases a low capability employee that has low variance may be perceived as having higher average capability than a higher capability employee with higher variance. Clearly, ratings of employees should not always be based solely on their average capability, particularly in cases in which consistency is important. However, in many cases understanding an employee's average capability is essential for maximizing an organization's productivity.

Study 6: Common Work Assignments

While the previous study demonstrated how deadlines might censor capabilities, the amount of work available can also censor employee capability. For example, imagine that a manager assigns 10 clients to her employee to be serviced by the end of the week (see Figure 12). If the employee is capable of servicing only 6 clients that week, then the manager observes the failure and the exact number of clients left unserviced. If the employee is capable of servicing 14 clients that week, then the manager simply observes that all 10 clients have been serviced by the end of the week. The potential to have serviced 4 additional clients goes unobserved. In this manner, work assignments, or the quantity of work available, can censor what a manager can observe about the amount of work an employee is capable of completing in a given timeframe.

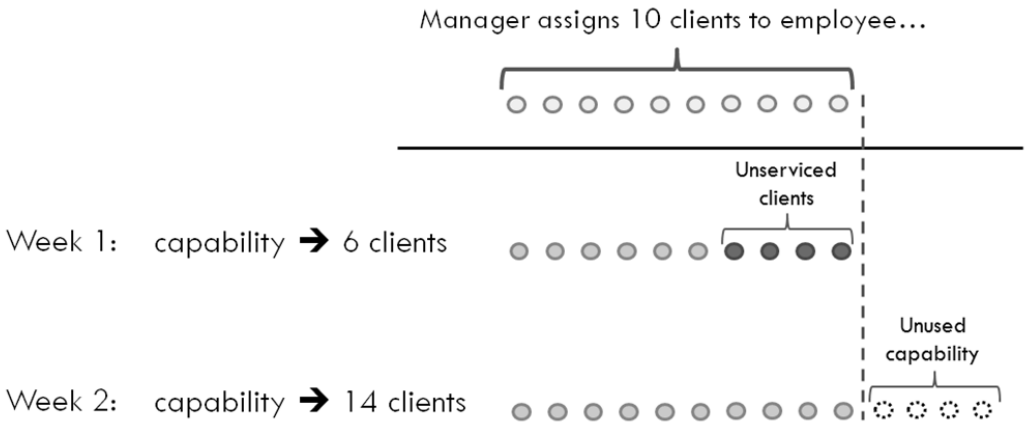

Figure 12: How work assignments censor employee capability.

More generally, the work assignment censors what the manager can observe about the employee's capability. Capability *lower* than the work assignment is observed at its true value, while capability *higher* than the work assignment is censored and observed at the censorship point—the manager can only observe that the assigned work was finished.

The purpose of this study was to examine learning of employee capability when the same amount of work is assigned to each employee. This study built on the existing findings in several ways. First, it empirically demonstrated how the amount of work assigned in a given period can censor employee capability. Second, it demonstrated how common work assignments—giving all employees the same quantity of work—causes managers to underestimate the capabilities of high performers more than low performers. Third, it demonstrated how censorship can cause managers to perceive their employees to be more homogenous in their capabilities than they actually are.

Revisiting Hypothesis 2 in Chapter 1, as the degree of censorship increases, judgment is more biased because the observed sample mean moves farther from the true

mean. When all employees are assigned the same quantity of work, employees with higher capabilities are able to complete the assignment more frequently than employees with lower capability. That is, the capabilities of high-performers will be censored more often than the capabilities of low-performers. Since the degree of underestimation increases with the degree of censorship, I predicted high-performers would be underestimated to a greater extent than low-performers.

As a consequence of high-performers being underestimated more than low-performers, managers will perceive their employees to be more similar to each other in their capabilities than they actually are (see Figure 13). Therefore, a corollary hypothesis is that a manager's perceived distribution of capabilities across employees will be more condensed than the actual distribution.

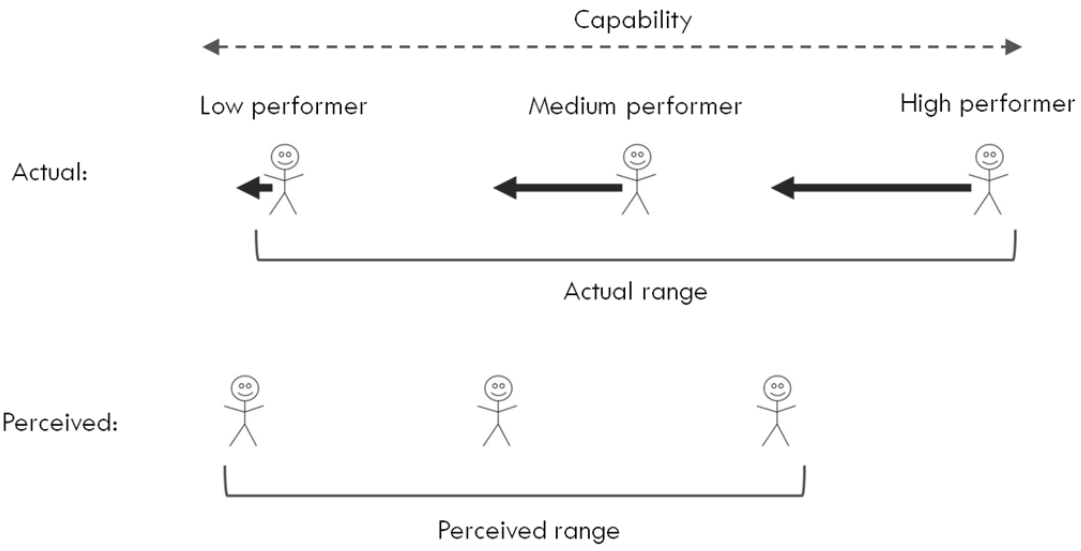

Figure 13: A depiction of how, with common work assignments, the greater censorship and underestimation of high performers leads to a smaller perceived range

of employee capabilities. The arrows indiciate the degree to which each employee is underestimated.

Method

The Task

First, I measured the number of small puzzles a single participant could complete in one minute. I repeated this for a total of 40 minutes. To achieve a greater spread, I multiplied each of the 30 observations by a constant yielding a distribution of capabilities with M = 66.0 and SD = 7.8. This served as the capability distribution representing the number of tasks the employee was capable of completing in a day.

In the second part of the study, participants made judgments about three employees: a high-performer, a medium-performer, and a low-performer. I used the measured capability distribution (based on the distribution of the number of puzzles the participant could finish in a minute) as the medium-performer's capability. To hold capability variance constant but manipulate capability, I shifted the entire set of outcomes by a constant to create distributions for the high- and low-performers. For the high-performer, I added 4 units to each capability observation yielding a mean capability of 70 tasks, and for the low-performer, I subtracted 4 units from each capability outcome yielding a mean capability of 62 tasks. See Figure 14 for an example of the resulting distributions.

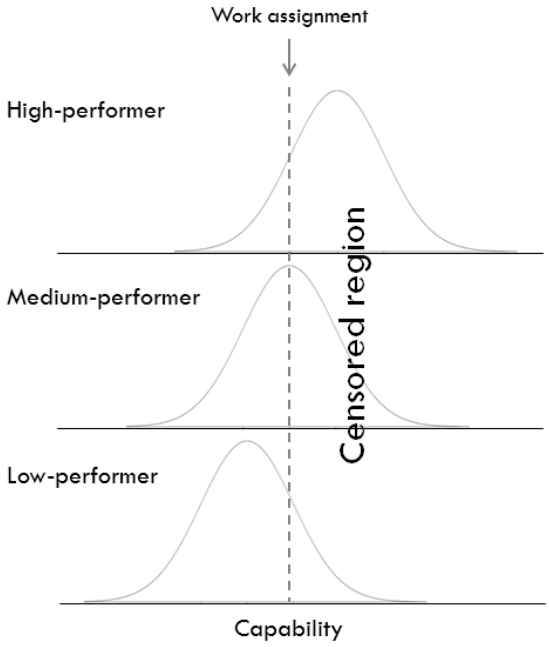

Figure 14: An illustration of how the capability distribution was shifted to yield high-, medium-, and low-performers. The normal distribution is used here simply for illustration.

Next I generated the production outcomes that would be observed by the participants playing the role of the manager. The daily work assignment for each employee was set equal to the medium-performer's average capability, 66 tasks. In a given day, if the number of tasks the employee was capable of was greater than or equal to 66, then manager observed that all 66 tasks were completed. If the number of tasks the employee was capable of was less than 66, then the manager observed how many were left unfinished. These outcomes were generated for the low-, medium-, and high-performer. The order of the observations was randomized for each of the employees so that participants could not recognize that they were derived from a common set.

Procedures

Twenty-two individuals participated as managers in this study. They were paid a base of $5 and could earn bonus money based on the accuracy of their estimates, up to an additional $5. The study was conducted in a computer lab. Each participant received a packet containing performance outcomes for three employees (high-performer, medium-performer, and low-performer) on separate sheets in a random order. For each employee the participant observed 30 days of performance. At the top of each sheet it was highlighted that the employee had been assigned 66 tasks each day. For each day the participant observed whether the employee completed the work assignment and how many tasks the employee was able to complete.

Participants were informed of the following: "In each day employees were assigned 66 tasks and you can see how many tasks they actually completed each day. For any day in which they were not able to complete the work assignment, you can see exactly how much work they were able to complete. For any day in which they were capable of doing more than the work assignment, all you can see is that they finished the 66 tasks for the day." Participants were then asked to estimate the average number of tasks each employee was *capable* of completing in a day if they were never limited by the number of tasks assigned.

Results

See Figure 15 for a depiction of the results. Each employee's actual average capability (62, 66, or 70) was subtracted from the participants' estimates of each. Participants' estimates of the mean capability of low-performers (M = -1.31, SD = 1.35),

medium-performers (M = -2.33, SD = 1.78), and high-performers (M = -3.25, SD = 2.94) were significantly lower than their true mean capabilities, *ts* > 2.8, *ps* < .01. A within-subjects ANOVA found that employee capability (i.e., low, medium, or high-performer) had a significant effect on the estimates of average capability relative to the true average, $F(2,42) = 4.56$, $p < .05$. Follow-up contrast analyses showed that there was significantly greater underestimation of high-performers than low-performers, $t(42) = 3.02$, $p < .01$. The underestimation of medium-performers was not significantly different from that of low-performers ($t(42) = 1.59$, $p = .12$) or that of high-performers ($t(42) = 1.43$, $p = .16$), although it fell between that of the other two conditions, as expected.

The corollary hypothesis was that participants would perceive employees as more similar in their capabilities than they actually are. That is, they would underestimate the range between the mean capability of the low-performer and the mean capability of the high-performer. For each participant, I calculated the range between the participants' estimates of the low-performer and high-performer's mean capabilities. I then compared the perceived range of mean capabilities to the actual range of their mean capabilities (8 tasks) using a one-sample t-test. Participants significantly underestimated the range of employee's average capabilities (M = 6.06, SD = 3.16), $t(21) = 2.88$, $p < .01$. Participants believed that employees were more similar to each other in their capabilities than they actually were.

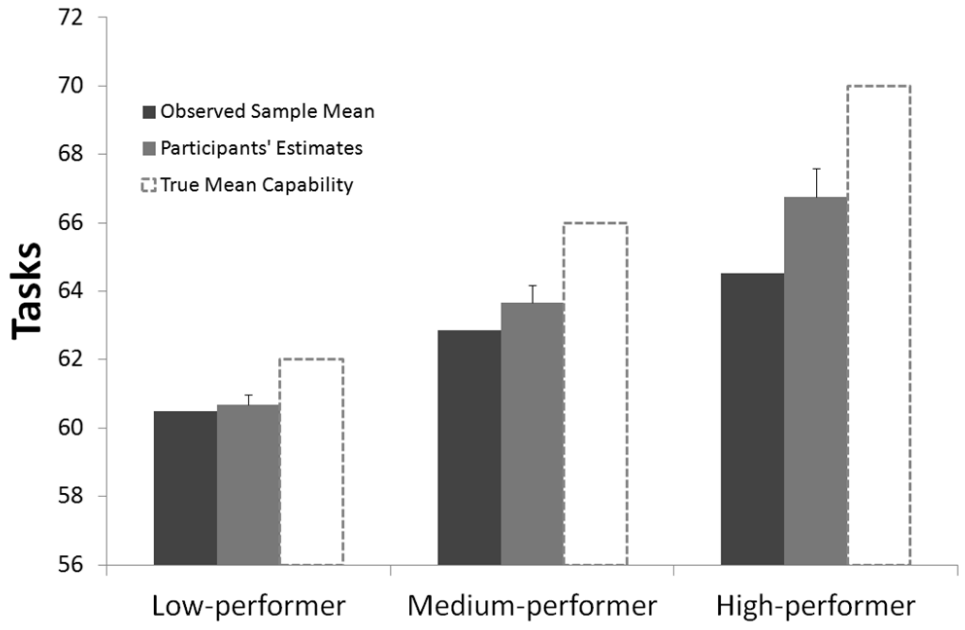

Figure 15: Participants' estimates of mean capability across levels of employee capability in Study 6.

In summary, participants underestimated what the employees were capable of. Further, common work assignments caused high-performers to be underestimated more than low performers. As an additional consequence, participants underestimated the spread of their employees' capabilities. They believed that employees were more similar to each than they actually were. Organizations and managers that fall prey to the censorship bias will underestimate and undervalue high capability employees the most. This tendency may lead to a talent drain, in which high capability employees leave to join companies that better appreciate their capabilities.

Study 7: Failure and Success Framing

Study 7 explored two things. First, it examined the effect of framing outcomes as successes and failures on the perceptions of employee capability. Second, it explored how

a manager's decisions shape what they are able to observe, and ultimately learn, about their employees.

Organizational outcomes are often framed in terms of success or failure (Levinthal & March 1993). Past research has shown that failure causes managers to more readily embrace change whereas success tends to lead managers to maintain the status quo (Audia, Locke, & Smith 2000; Audia & Brion 2006). Many scholars have suggested that exploration, experimentation, and occasional failure provide better learning opportunities than the pure pursuit of immediate successes (March 1991; Sitkin, Sutcliffe, and Schroeder 1994; Cohen & Levinthal 1990). Sitkin (1992) suggested that managers should seek out small failures to ensure that one's organization is accumulating knowledge. Building on these ideas, this study holds the actual outcomes constant and examines the effect of framing results in terms of failures and successes.

This study explored whether framing censored outcomes as "success" or uncensored outcomes as "failure" affects how managers learn employee capability. Failure and success framing might affect learning of employee capability in two ways. It may (1) alter the perceptions of the costs of failure and benefits of success and consequently affect the degree of constraint, or (2) affect the attention paid to unused capability and consequently affect managerial awareness of censorship.

Failure and success framing may affect learning of employee capability by shifting the manager's perception of the value of outcomes. Labeling an employee's inability to complete an assignment as a failure, even if the costs of such a failure are known, may cause that outcome to be perceived as more aversive. If these outcomes are

psychologically registered as worse than they are, managers will seek to reduce their incidence, a behavior I call *failure avoidance*. Similarly, labeling an employee's completed assignment as a success may cause that outcome to be perceived as more positive, irrespective of its actual payoff. When the successful completion of assignments is overvalued, managerial attention may be drawn away from profit maximizing and towards actions that can increase the incidence rate of successes, a behavior I call *success chasing*. To reduce the rate of failure and increase the rate of success, the manager assigns less work to employees. More conservative work assignments will increase the rate at which employee capability is censored, and ultimately increase the degree to which they are underestimated. In this way, a manager's decision making may be the engine of his own destruction and failure or success framing may fuel that engine.

Secondly, failure or success framing may affect the manager's awareness of unused employee capability; however, there are competing hypotheses here. First, a failure framing may simply attract greater attention to uncensored capability observations (i.e., when the employee fell short), while success framing may attract greater attention to censored observations (i.e., when the employee could have done more). Since effective learning of employee capability requires greater managerial attention on unused employee capability in censored periods, it is possible that failure framing exacerbates underestimation of employees, while success framing mitigates it.

On the other hand, failure framing might shift a manager's focus to the tradeoffs of under- vs. over-assigning work. Explicit failures in one direction (for low capability periods) may provide a cue to similar errors in the other direction (for high capability

periods). Conversely, labeling completed assignments as successes may place managers in a satisfied mindset, making them content with the status quo. This might hinder consideration of unused employee capability by giving the manager a sense of satisfaction with the outcome when work is completed. Therefore, failure framing might mitigate underestimation of employees, while success framing might exacerbate it.

The first prediction is regarding how individuals react to failure or success framing in their work assignment decisions. The second set of predictions contains competing hypotheses regarding how individuals interpret censored performance outcomes even before making work assignment decisions. Study 6 tested these hypotheses.

Methods

Procedures

In this study, I used a normal distribution to generate a team's capability, instead of actual measured outcomes. The team's capability was randomly drawn from a normal distribution with M = 84 and SD = 10. Eighty-one participants were seated in a computer lab. They were told that they would be playing the role of a manager of a team that services clients and given the following explanation:

> "The team has good days and bad days, so the number of clients the team is capable of handling in a day varies like a bell-curve. Specifically, their capability is normally distributed with standard deviation 10. So on average they are capable of a certain number, but sometimes they are capable of more and sometimes less. The number of clients actually serviced by the team depends not only on the team's capability, but also on the number of clients assigned to the team that day. Over the last six weeks, the team was assigned 84 clients each day. On days when the team's capability was lower than 84, the

> team serviced as many as they could and left some clients unserviced. On days when the team's capability was greater than 84, they simply serviced all of the clients and then stopped. The team earns $1 for each client that they service. However, for each client that they are assigned and then cannot service, it costs the team $1."

They were then told that, based on what they could see about their team's performance over the last 6 weeks, they needed to decide how many clients to assign to their team each for the *next* 6 weeks. They received up to $5 of bonus pay based on the profit earned by their team in the next 6 weeks. They were also told that an evaluator would judge their decision making ability and competence as a manager based on their decision and the performance of their team. This was done to add evaluative pressure on their decisions which might increase the likelihood that success and failure framing have an effect on learning and decision making.

All participants faced censored performance outcomes and were randomly assigned to one of three framing conditions: control, success, failure. The difference across conditions was in how the daily performance outcomes were presented to the participant. In the control condition, for each of 30 days participants observed the number of clients serviced, the number of clients *not* serviced, and the daily profit. In the success condition, each day that all of the assigned clients were serviced, the word "Success" was shown in green, bold-faced. In the failure condition, each day that the team failed to service all of the assigned clients, the word "Failure" was presented in red, bold-faced. All other elements were identical across conditions.

During the past 6 weeks (30 work days) the team had been assigned 84 clients each day. Participants reviewed these past results and then decided how many clients to assign to their team per day for the next 6 weeks. After making their decision, they were asked to estimate how many clients their team was capable of servicing in a day if they were never limited by the number of clients assigned to them.

The participants then received the results for the next 6 weeks based on their work assignment decision. They were asked to review these results and then update their estimate of their team's average capability.

Results

See Figure 16 for a summary of the results for Study 7. First, I analyzed the effect of framing condition on initial estimates of employee capability. There was no significant difference across framing conditions in the initial estimates of the team's average capability ($M_{control} = 82.33$, $M_{failure} = 82.15$, $M_{success} = 82.11$), $F(2,79) = 0.05$, $p > .9$. Participants in all three framing conditions significantly underestimated the team's mean capability (ts > 2.9, ps $< .01$). Therefore, success and failure framing appeared to have no effect on participants' consideration of unused employee capability and all conditions demonstrated an initial underestimation of employee capability.

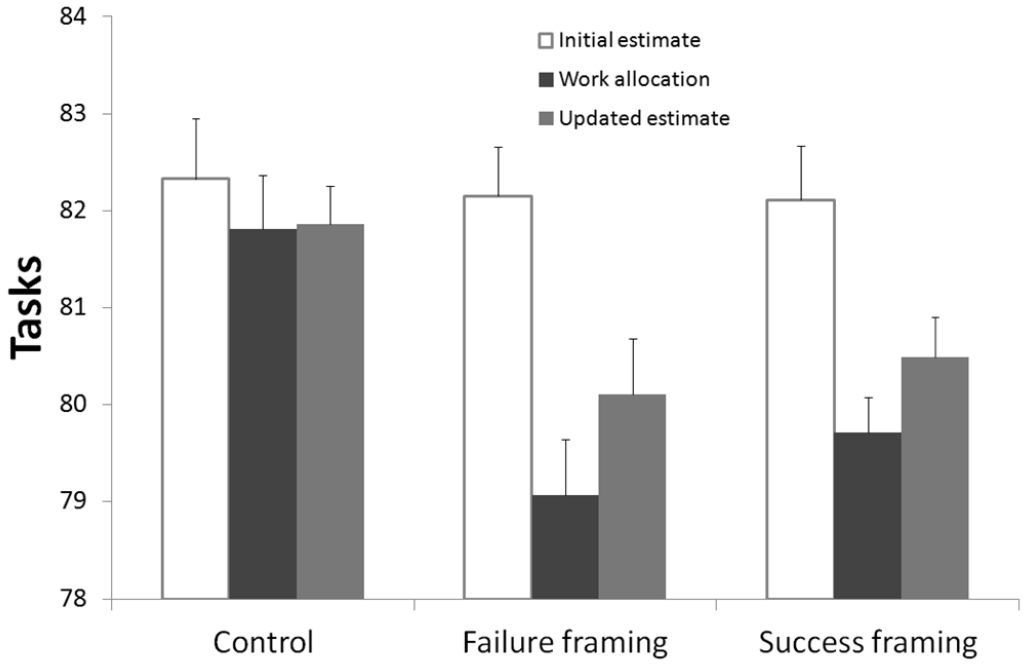

Figure 16: The estimates and work allocations by condition in Study 6. The true mean capability of the employee was 84.

However, there was a significant effect of framing condition on the number of clients participants chose to assign to their team, $F(2, 79) = 3.90$, $p < .001$. Participants in the failure condition ($M_{failure} = 79.07$) assigned significantly fewer clients to their team than participants in the control condition ($M_{control} = 81.81$), $t(79) = 3.81$, $p < .001$. Similarly, participants in the success condition ($M_{success} = 79.71$) also assigned significantly fewer clients to their team than participants in the control condition, $t(79) = 2.95$, $p < .01$. There was no significant difference in assignment between the failure and success framing conditions, $t(79) = 0.90$, $p > .35$.

After observing the outcomes for the next 6 weeks based on the number of clients they chose to assign to their team, there was a significant difference across framing conditions in their updated estimates of the team's average capability, $F(2,79) = 3.83$, $p <$

.05. The updated estimates of mean employee capability were significantly lower in the failure condition ($M_{failure} = 80.11$) than in the control condition ($M_{control} = 81.86$), $t(79) = 2.63$, $p < .05$. Updated estimates were also significantly lower with the success-framing ($M_{success} = 80.49$) than in the control condition, $t(79) = 2.08$, $p < .05$. There was no significant difference in the updated estimates between the failure-framing and success-framing conditions, $t(79) = 0.57$, $p > .5$.

This series of results suggests that failure and success framing did not affect the degree to which individuals considered unused capability in censored periods. However, the failure and success framings caused individuals to be significantly more conservative in their work assignments. These conservative choices created more censorship of employees, subsequently leading to greater underestimation of employees' capabilities. To explicitly test this causal path, I conducted a bootstrap mediation analysis (Preacher & Hayes 2004; Zhao, Lynch, & Chen 2010). The analysis tested if there was a significant indirect effect of failure or success framing on estimates of employee capability through work assignment decisions. Using 20,000 bootstrap resamples, the mean indirect effect was -1.13 and the 95% confidence interval did not contain zero (-1.93 to -0.48), meaning statistical significance for the test of mediation. There was no significant effect of failure or success framing on updated estimates of employee capability when controlling for the work assignment, $t(79) = .77$, $p > .4$. This supports the contention that failure and success framing affects beliefs about employee capability through causing individuals to make conservative work assignments.

In summary, failure and success framing caused individuals to be more conservative in their work assignments, presumably to reduce the incidence of failure (i.e., failure avoidance) and to increase the incidence of success (i.e., success chasing). These conservative work assignments increased the degree to which employee capability was censored, subsequently causing greater underestimation of employee capability.

Chapter 5: Predictors of Employee Underestimation

Organizations provide rich, social, dynamic settings in which to examine behavior. This makes studying decision making and learning in organizations interesting, but also complex. In the previous chapter I briefly foreshadowed the three key mechanisms that affect the likelihood of employee underestimation: the degree of constraint, employee disclosure, and managerial awareness. In this chapter I discuss each mechanism in greater depth. I also develop predictions about individual, social, and organizational factors that affect underestimation of employees through these mechanisms.

The Degree of Constraint

As a constraint limits more of employee capability, the observed sample of employee performance becomes less representative of the employee's true capability. As the mean of the observed sample moves farther from the true mean of the population, the decision maker must adjust one's beliefs to a greater extent to form accurate beliefs about the underlying population. Therefore, factors that make observed employee performance less representative of employee capability will lead to greater underestimation of employees.

The greatest determinant of the degree of constraint is the extent to which work assignments are conservative or ambitious. When work assignments are easily achieved, managers have more restricted observations of what the employee can achieve. As mentioned in the previous chapter, when the cost of employee failure is high, work assignments will be more conservative. However, conservative work assignments mean

that in most periods an employee will finish the work assigned to them—perhaps easily. In the extreme, if an employee is *always* able to complete the assigned work with sufficient quality, then the manager can never observe their true capability. The employee might be capable of twice as much work but the manager has no way of observing that unused capability. Therefore, while it is rational to lower work assignments as the cost of failure increases, doing so also increases the degree of constraint (i.e., degree of censorship), increasing the likelihood of employee underestimation (Hypothesis 2 in Chapter 1).

Conversely, when the value of labor is high (i.e., high human capital), the work assignments will be more ambitious to ensure that the firm does not miss out on such valuable potential productivity. As an individual's human capital increases, they are able to provide more value per unit of time worked. To ensure that they are able to harvest every possible unit of productivity from high capital employees, firms give them extremely ambitious deadlines and large quantities of work. More ambitious assignments reduce censorship as these employees are then pushed to their limits in almost every period. (In the extreme, employees may end up consistently sacrificing quality in their work as they are always pushed to the limits of their capabilities – e.g., overworked and fatigued doctors have more attentional failures, Lockley et al. 2004[1].) Therefore, I would expect to find a tendency of less underestimation of employees at the top of the

---

[1] In the event that employees *under-report* their working hours, as sometimes occurs in medicine or consulting, managers may actually begin to overestimate what employees are capable of in a given amount of time.

workforce pyramid – in such jobs as lawyer, doctor, or consultant – and more underestimation of employees at the bottom of the workforce pyramid (see Figure 17). In sum, the degree of the constraint will be determined by the costs to the firm of employee failure and the benefits to the firm of employee labor. The greater the degree of constraint, the higher the likelihood that managers will underestimate their employees.

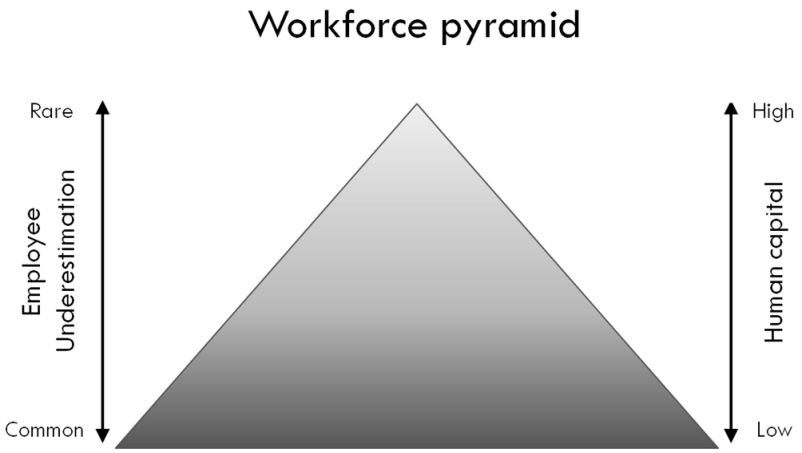

Figure 17: A workforce pyramid expressing the relationship between human capital and employee underestimation.

Proposition 1: The likelihood that, and degree to which, an employee is underestimated will be positively related to the costs of employee failure.

Proposition 2: The likelihood that, and degree to which, an employee is underestimated will be negatively related to the value of the employee's labor.

Over-routinization hinders learning because it only allows managers to observe organizational outcomes under a single set of conditions. Therefore, experimentation is a critical element of how managers and organizations learn (Huber 1991; Levinthal & March 1993, March 1991, McGrath 2001; Sitkin 1992). Experimenting with an employee's work assignments and deadlines provides the opportunity to have a

manager's perceptions of the employee disconfirmed. Periodically providing more ambitious goals, both in terms of assignments and deadlines, allows employees to demonstrate how much more of or how much faster they could have completed certain tasks (Sitkin et al. 2011). Reducing structure by assigning more flexible tasks also allows employees to demonstrate their capabilities more freely. However, for employees to exceed current expectations with flexible assignments, they must also be given some degree of resource slack (Sitkin, Sutcliffe, Schroeder 1994). Flexible assignments and experimentation with ambitious goals allow managers to see more uncensored observations of employee capability and at higher levels. In using these strategies, managers give employees the opportunity to prove their capability beliefs wrong.

Proposition 3a: The likelihood that, and degree to which, an employee is underestimated will be negatively related to the frequency with which the manager assigns flexible tasks with resource slack.

Proposition 3b: The likelihood that, and degree to which, an employee is underestimated will be negatively related to the frequency with which the manager experiments with ambitious work assignments and deadlines.

Employee Disclosure

The second mechanism through which factors can affect employee underestimation is employee disclosure of unused capability. The social nature of this mechanism means that it can be affected by individual, social, and organizational factors.

When an employee is being underestimated, one way to debias a manager's perception is to approach the manager and communicate that one's capability is not being

fully utilized. However, in such a case the employee would be actively seeking more work and higher expectations; even an ambitious employee may be hesitant to raise the expectations of the manager, so as to avoid any possibility of falling short of difficult future assignments. Further, self-disclosure requires boldness on the part of the employee because communicating excess capability implies that the manager made a mistake in assigning too little work. The possible contentiousness of expressing underutilization to one's manager may leave only the most aggressive and self-promoting individuals to do so. Research suggests that individuals differ in their willingness to act in self-promoting ways (Jones & Pittman 1982; Thomas, Whitman, & Viswesvaran 2010; Hackman & Oldham 1976). I expect the most self-promoting individuals to be the most aggressive in self-disclosing underutilization and will therefore be underestimated less by their managers.

Past research has found that women tend to be less self-promoting than men. For example, negotiation research has found that men tend to be more aggressive than women in their demands in salary negotiations which partially accounts for the gender wage gap. Women also tend to have greater faith that systems are meritocratic and that managers will recognize capabilities and reward performance fairly (Babcock & Laschever 2003, 2009). Some scholars have argued that the primary reason why women tend to be less self-promoting than men is that they face large social costs for acting in an agentic way (Bowles, Babcock, & Lai 2007; Glick, Zion, & Nelson 1988). There are often strong social pressures for women to conform to what society declares is prototypical behavior for women (Eagly & Karau 2002; Heilman 1983, 2001; Rudman & Glick 2001).

Specifically, female aggressiveness is frowned upon while female niceness is deemed appropriate (Rudman 1998). When women engage in actions that are inconsistent with this prototype then they are often disliked and discredited at work. Therefore, social pressures discourage women from being self-promoting.

If a key method for overcoming censorship of one's capabilities is to express to the manager that one is capable of more (i.e., to promote oneself), and women tend to be less self-promoting than men, then one would expect that women tend to be underestimated by their manager more than men.

Proposition 4: The likelihood that, and degree to which, an employee is underestimated will be negatively related to the degree to which the employee is self-promoting.

Proposition 5: Women will tend to be underestimated by their managers more than men.

In some organizations there are strong social norms among employees that discourage any self-disclosure of underutilization to the manager. The term "rate-buster" refers to an employee that performs up to their capability, above the socially accepted level of performance among the employees. Such behavior is often discouraged through persuasion and coercion in an effort to bring this individual's productivity back to the generally achieved level. Rate-busting is frowned upon by other workers because it can lead to raised expectations for all employees (Blau 1957). In such an environment, employees that are able to appear hard-working but complete work slowly are contributing the most to the well-being of other employees, while employees performing

above and beyond expectations are hurting other employees (Block 1976). These social pressures reduce the likelihood that employees will seek to disclose to managers when they are capable of more than what is being asked of them.

Proposition 6: The likelihood that, and degree to which, an employee is underestimated will be positively related to the degree to which rate-busting is discouraged among employees.

For an employee to disclose that they are being underutilized, there must be opportunities for upward feedback in the organization and the work environment must feel safe for the employee to do so. Many managers fear negative feedback and feel threatened when they receive it (Swann & Read 1981). They may try to avoid negative feedback (Ashford & Cummings 1983), or when another individual presents a criticism they may try to discredit the feedback or the source (Ilgen, Fischer, & Taylor 1979). Research has shown that leaders differ in their willingness to receive and learn from upward feedback (Owens & Hekman 2012). When an organization fosters an environment of safe communication and employee empowerment, employees will be more likely to disclose to the manager when they are capable of more than what is currently asked of them (Morrison & Milliken 2000; Morrison & Phelps 1999; Edmondson 1996). In an organization that provides structured opportunities for upward feedback and a safe environment for bottom-up communication, employees will be more likely to disclose unused capability, and managers will ultimately underestimate their employees less.

Proposition 7: The likelihood that, and degree to which, an employee is underestimated will be negatively related to the degree to which an organization structures upward feedback opportunities and provides a safe environment for employees to give feedback to managers.

Organizations, and jobs within organizations, differ in their incentive structures. When an employee has large personal incentives to complete additional work, then they are more likely to disclose to the manager when they are underutilized. Therefore, underestimation of employees through the censorship bias is less likely to occur where employees are paid based on piece production and is more likely to occur where employees are paid by a fixed salary or by the hour. With piece-rate pay, if an employee is given insufficient work then their own income is being harmed when they are unable to reach their maximum potential. However, with salary or hourly pay, employees may be content to be underutilized and may choose not to seek more work (i.e., be effort-averse).

Particularly ambitious employees may notice the censorship of their capabilities and choose to self-disclose their excess capability to a manager. However, a dearth of promotion opportunities in an organization would leave even ambitious employees with little incentive to debias a manager's perception of their capability. Therefore, I expect incentive schemes and prevalence of promotion opportunities affect the likelihood and degree of employee underestimation.

Proposition 8a: The likelihood that, and degree to which, an employee is underestimated will be greater with hourly or salary-based pay than with piece-rate pay.

Proposition 8b: The likelihood that, and degree to which, an employee is underestimated will be negatively related to the prevalence of promotion opportunities.

Managerial Awareness

The third mechanism through which factors affect employee underestimation is managerial awareness of unused capability. Censored environments provide evidence to the decision maker that information is being missed. The formation of accurate perceptions of employee capability depends on the manager's ability to consider and account for unused potential in the event that the employee is able to complete assignments and meet deadlines. Factors that cue managers to pay greater attention to what could have been achieved are likely to reduce underestimation of employees. Factors that draw managerial attention away from unobserved employee capability are likely to exacerbate underestimation of employees.

For example, Study 1 found that slightly variable censorship points led to less accounting for censorship than stationary censorship points, because the latter produce large spikes of censored observations at a single point. These spikes provide stronger cues that the observed sample is biased. This effect operates through managerial awareness. In terms of capability learning, when work assignments are identical across periods, the manager can observe that the employee can consistently complete the work at that level, providing a strong cue that the employee is underutilized. This draws the manager's attention to the employee's unused capability and causes them to raise their expectations. Such an observation might be more difficult to make if assignments vary slightly from period to period.

A manager's span of control, specifically the number of employees reporting to the manager, affects the manager's ability to consider the unused potential of employees (Bell 1967; Blau & Schoenherr 1971; Keren & Levhari 1979; Ouchi & Dowling 1974; Simon 1945). Urwick (1956) suggested that the optimal number of employees for a manager to be in contact with is between 4 and 6. When the number of employees reporting to a manager is large, then the manager has less attention to spend on deliberative consideration of what could have been for each employee. Accounting for capability that is not readily observable requires effortful deliberation by the manager. A cognitively overloaded manager, with fewer mental resources to allocate to each employee, is more likely to take employee performance outcomes at face value and less likely to consider when employees could have done more than what was asked of them. Therefore, the more employees a manager's attention is divided across, the more likely the manager is to underestimate the employees.

Proposition 9: The likelihood that, and degree to which, an employee is underestimated will be positively related to the number of employee's under that manager (i.e., managerial span of control).

Individuals in managerial positions may come to be managers by different avenues. Some individuals are promoted to a managerial position from within and manage individuals that serve the role that they formerly occupied. These managers have personal experience with the work that their subordinates face and are less likely to underestimate the capabilities of their employees because they can remember how their capability was censored when they were in that position. On the other hand, some

individuals come to be managers from external organizations or other areas within the same organization. These managers do not have personal experience with the work faced by their subordinates. I predict that individuals with personal experience with the work of their employees are more aware of censored capability than managers without such personal experience. However, if this effect relies on the memory of personal experience as an employee, then over time the value of past experience as an employee may decay and managers may begin to underestimate their employees to a greater extent.

Proposition 10: The likelihood that, and degree to which, an employee is underestimated will be lower when the manager has been internally promoted or previously occupied a role similar to the employee's.

Discussion

Deadlines often prevent managers from observing how much faster an employee could have completed the work. However, even if a manager sees that an employee has completed an assignment just before the deadline, it does not necessarily imply that the employee could not have completed the work any faster. For instance, the deadline may have been extremely generous, but employees may have procrastinated before finishing the task just in time (Ariely & Wertenbroch 2002; O'Donoghue & Rabin 2001). If the employee had been given less time to accomplish the work, they may have simply procrastinated less. Similarly, an employee might finish the task just before a lenient deadline if the employee works intermittently, or at a slower pace than they are capable of (Hasija, Pinker, & Shumsky 2010; Latham & Locke 1975; Parkinson 1955). Therefore,

evidence that an employee has finished a task just before the deadline is not indicative that they were pushed to limit of their capability.

Some researchers have found that managers' evaluations of employees tend to be less favorable than employees' evaluations of themselves (e.g., Campbell & Lee 1988; Gioia & Sims 1985). In past research this difference has been attributed to self-serving biases, where employees believe that they are better than they actually are. However, my findings suggest that this difference may, in part, result from an asymmetry in what the manager can observe about an employee's capability. Censorship may cause the manager to underestimate the employee's capability, while the employee can better see how much more they could have done. Therefore, the difference between the manager's evaluation and the employee's self-evaluation may result from the underestimation of employees caused by the censorship bias.

While the primary focus of Chapters 4 and 5 has been on learning the capability of a single employee, the same predictions can apply to learning a team's capability. However, there are reasons to believe that underestimation is more or less likely to emerge when making judgments of teams, as opposed to individuals. In teams there may be a diffusion of responsibility which makes employees less likely to disclose to the manager when the team is capable of more (Darley & Latané 1968; Mynatt & Sherman 1975; Wallach, Kogan, & Bem 1964; Whyte 1991). Furthermore, cohesive teams may have strong social norms that discourage team-members from seeking additional work for the team (Blau 1957). Conversely, it is possible that the censorship bias is less likely to emerge for teams because it only takes one member of the team to disclose to the

manager that the team has excess capability. If each member of the team has an independent probability *p* of disclosing the team's excess capability, then the disjunctive probability that at least one employee discloses the team's unused potential will be greater than *p*. Therefore, the likelihood of the censorship bias emerging for perceptions of teams will largely depend on the social norms within the team and the independence of disclosure likelihoods of the team members.

Chapter 6: General Discussion

This research has demonstrated a *censorship bias*– individuals in censored environments tend to rely too heavily on their observed sample, biasing their beliefs about the underlying population. First, I explored how censorship affects learning demand. Study 1 examined the task of estimating the mean of a normally distributed unknown demand from sales. This study provided evidence of the censorship bias: individuals with censorship underestimated mean demand. Further, the censorship bias was exacerbated for higher degrees of censorship and when the censorship point was variable. Study 2 linked biased inferences from censored samples to behavior by studying simultaneous judgment and decision-making in a task that involved dynamic demand learning and inventory choices. Censorship caused individuals to underestimate demand and stock less inventory than optimal. Study 3 demonstrated that the censorship bias could be partially mitigated by drawing greater attention to the unobserved magnitude of censored observations (e.g., missed sales).

Next, I examined how perceptions of risk can be shaped by censorship. Study 4 tested the effect of censorship in a sequential risk-taking task where individuals attempted to avoid a negative outcome that was uniformly distributed with an unknown upper boundary. Individuals with censorship took fewer risks and underestimated the upper boundary of the distribution. Had they known the upper boundary, they reported that they would have preferred to take more risks than they did.

Finally, I examined how censorship affects the ability of a manager to learn about employee capability. Study 5 demonstrated the censorship bias with actual performance

outcomes, rather than random draws from a pre-defined probability distribution and found that higher variance in the underlying population exacerbated the bias. With censorship, individuals playing the role of the manager overestimated the average time employees needed to complete a task, and to a greater extent for those with high variance in their capabilities. In Study 6, managers assigned the same amount of work to employees that differed in their average capability. High-performers were underestimated more than low-performers causing managers to believe employees were more similar in their capabilities than they actually were. Study 7 examined the effect of framing censored outcomes as successes and uncensored outcomes as failures. These framings did not affect people's consideration of unused employee capability, but they realigned people's perceptions of the costs of failure and benefits of success. Failure and success framing caused individuals to make conservative work assignments in an effort to avoid failures and chase successes. Their conservatism censored employee capability in the following period to a greater extent and ultimately led to greater underestimation of employees.

Consistent with the naïve intuitive statistician metaphor (Fiedler & Juslin 2006; Juslin et al. 2007), the censorship bias was greater when the censored sample was less representative of the true population. However, individuals did use evidence of censorship to adjust from the observed sample to form their beliefs about the underlying population. Nevertheless, their adjustments fell short of a theoretically attainable heuristic strategy. In Study 1, individuals' estimates were compared to the estimates of a simple MLE-based heuristic estimate (Nahmias 1994) given the same censored samples. The

prescriptive heuristic's estimate given the same censored samples greatly outperformed the estimates of individuals, suggesting that individuals could benefit from simple decision aids in censored environments.

These empirical results also suggest that judgment in censored environments may be driven by the use of the censored sample as an initial anchor for estimating characteristics of the underlying population (Chapman & Johnson 1999; Epley & Gilovich 2001, 2004, 2006; Tversky & Kahneman 1974; Strack & Mussweiler 1997). Since individuals tend to naïvely believe biased samples are more representative of populations than they actually are (Fiedler 2000; Juslin et al. 2007), they insufficiently adjust from the observed sample and form biased beliefs about the population.

This research demonstrates that greater censorship and higher variance in the population can make beliefs about the population more biased. Other factors may also affect judgment in censored environments by making the observed sample less representative of the underlying population. A negative correlation between censorship points and population draws makes the observed sample more misrepresentative of the underlying population because high draws become censored at even lower points. For example, a negative correlation may occur if a firm has more limited access to inventory in periods where customer demand is high. On the other hand, a positive correlation between censorship points and sample observations makes the sample more representative of the population because high draws coincide with high censorship points. For example, if a firm can partially anticipate random demand realizations and

appropriately adjust inventory levels, then their beliefs about the mean of the demand distribution are likely to be less biased.

These results suggest that skewness in the population may also cause censored samples to be more misrepresentative. A right-censored environment would censor the long tail of a right-skewed population. Therefore, the true value of the high observations in the long tail would go unobserved causing the observed sample to be less representative of the population. Future research may explore when these and other factors determine the extent to which individuals form biased beliefs in censored environments.

Implications

Demand Estimation

The empirical findings on learning an *unknown* demand complement the large amount of inventory ordering research that assumes a *known* demand distribution (e.g., Schweitzer & Cachon 2000; Bolton & Katok 2008; Croson & Donohue 2003, 2006; Su 2008; Ho et al. 2010). Demand beliefs directly inform inventory decisions. Therefore, even if an inventory policy is already determined and optimized for certain cost parameters, firms using past sales data may under-order because the censorship bias causes them to underestimate demand.

Furthermore, the results suggest that the censorship bias will be more problematic in several predictable circumstances. The effect of the degree of censorship suggests that the censorship bias in demand estimation is likely to be a greater problem with low-profit margin products, for which it is in the manager's interest to maintain lower inventory and

incur more frequent stock-outs; the resulting high censorship rate is likely to cause greater underestimation of demand. Conversely, for high profit-margin products, greater inventory is maintained and fewer stockouts occur, so a censorship bias is likely to be smaller. Similarly, a cash or space-constrained firm may be forced to hold lower levels of inventory, which is likely to cause greater underestimation of demand.

The effect of higher variance in the population suggests that high demand variance may be costly not only because of expected inventory-demand mismatch costs, but also because of a greater demand estimation bias. Finally, the effect of a variable censorship point suggests that demand inference is likely to be more accurate in cases where inventory levels remain constant for multiple periods of time before adjustment. Consistency in inventory levels may facilitate better demand learning by increasing the salience of stockouts as they amass at a single point (for related research, see Bolton & Katok 2008; Lurie & Swaminathan 2008).

Van Nieuwerburgh & Veldkamp (2006) have suggested that some massive economic patterns may actually be driven by asymmetric learning. Specifically, they argued that the asymmetry in business cycles (Sichel 1993; Veldkamp 2005)—rapid recession and slow growth—emerges as a consequence of what can be learned from productivity. During a boom, firms engage in high investment and productivity. When the boom ends, firms have precise evidence of the downturn and decisively decrease investment, yielding low productivity. However when growth resumes, the low productivity allows only noisy signals of improvement, which slows learning and makes recovery more gradual. In the terms of this paper, the productivity level censors what

firms can observe about the market. Downturns yield more uncensored observations allowing swift recognition of a shift, while upturns produce more censored observations which only provide imprecise evidence of improvement. In this way, it is conceivable that the micro-level findings in this paper relate to macro-level learning processes (Van Nieuwerburgh & Veldkamp 2006).

Risky Choice

In domains of risky choice, individuals often cannot observe what would have happened if they had taken more risks (March 1996). Since individuals tend to rely on their observed sample of experiences to make judgments, they fail to appreciate potential benefits that were missed as a result of conservative decision making. As in Study 4, censorship may cause people to overestimate risk and consequently avoid risky options that they actually would have liked to take. Therefore, it may be particularly important for managers to reconsider their perceptions of risk when they use conservative risky choice policies, which censor outcomes to a greater extent. Similarly, having insurance when taking risks not only reduces the variance of outcomes, but also realigns incentives to make exploratory risk-taking more attractive. Therefore, moral hazard may improve learning.

Employee Capability

The findings in this paper also suggest that managers are likely to underestimate their employees when there are constraints, such as deadlines or work assignments, that limit the observation of their capability. Traditionally, management & organization

research has focused on how work assignments involve the tradeoff of efficiency and slack and how this tradeoff affects innovation and firm survival in the long-run (e.g., Cohen & Levinthal 1990; Levinthal & March 1993; March 1991). However, this literature has largely neglected that managers must simultaneously learn the capabilities of their employees and that work assignments may systematically limit what employees can achieve and managers can observe. I encourage future research to examine how managers should integrate intra-organizational learning goals into their work assignment decisions, while balancing the tradeoff between efficiency and flexibility.

How can organizations reduce the likelihood that managers underestimate employees? An important first step is to redefine the nature of success and failure. Successes should be interpreted not solely relative to a previously held aspiration point, but also relative to what could have been achieved if other courses of action had been taken. Organizations have a tendency to remain inert after success (Audia, Locke, & Smith 2000); however, to form accurate beliefs in censored environments, a manager must have the ability to consider what more could have been achieved, even in the face of success. Attempting to find shortcomings in the face of success is likely to improve organizational learning and performance.

In light of the present research, I concur with past research that has suggested that organizations should not single-mindedly avoid failure, but should, perhaps counter-intuitively, seek it out (e.g., Sitkin 1992). Failure avoidance leads to conservative decision making that limits the range of outcomes a manager can observe. By tolerating, or even encouraging, some amount of failure, managers are able to form more accurate

impressions of the world around them and are more likely to be able to maximize the productivity of their workforce. Redefining success and failure as outcomes on a continuum rather than absolute states may help organizations develop greater ability and willingness to change (Audia, Locke, & smith 2000; Helfat & Peteraf 2003).

Improving Learning

The empirical findings in this paper also suggest that censored environments may be good candidates for intervention with cognitive repairs (Heath, Larrick, & Klayman 1998), decision aids, or, if practically and financially possible, optimization tools. For example, Study 3 found that asking individuals to explicitly estimate the true value of each censored observation can significantly improve the accuracy of the estimates of the population mean. Further, the excellent performance of Nahmias'(1994) prescriptive heuristic in Study 1 demonstrates the value of implementing even very simple decision tools for improving judgment in censored environments, although more complicated solutions should be implemented when censorship points change dramatically across periods.

There is a large body of organizational research arguing that managers and firms experiment too little, leading them to persist with a narrow set of beliefs and strategies (March 1991). When there are constraints that systematically limit what one can see, managers may consider implementing greater experimentation to increase learning. In the case of censored environments, optimal experimentation requires sacrificing some short term profitability by acting in a way that reduces censorship and reveals more true values of otherwise censored instances (e.g., Harpaz, Lee & Winkler 1982, Larivierre &

Porteus 1999). However, the findings in Study 1 suggest that such experimentation may be more effective for human decision-makers when censorship points are systematically set at a stationary value for multiple periods than when they are adjusted each period. An important direction for future research is to examine whether decision-makers recognize the need to explore in environments with constraints on information. The evidence in this paper suggests that individuals do not do so optimally. Further, research could examine whether individuals are capable of exploring effectively even when experimentation is of little or no cost.

Conclusion

While previous research has developed statistical tools for coping with censored data, little attention has been given to how managerial intuition may be biased by censorship. This paper provides insights into how individuals make judgments in censored environments which can be applied to various managerial settings. Individuals in censored environments tend to rely too heavily on their observed sample, causing them to form biased beliefs about the underlying population. Systematic aspects of the environment—such as the degree of censorship, population variance, and censorship point variability—increase the degree of bias. The censorship bias can cause sub-optimal decision-making which may be costly for organizations. An important challenge faced by managers is the need to build flexible organizations that can recognize and harvest gains in high potential-performance periods, while avoiding excessive vulnerability in low potential-performance periods.

Appendix A

Period	Sales	Sold out?	Estimate m (mean demand)
1	580	Yes	600
2	502	No	558
3	573	Yes	578
4	521	No	554
5	545	No	?

Figure 18: Depiction of the task interface for Study 1. This participant next needs to input a best estimate for the mean of underlying demand (m) for period 5.

Appendix B

Period	What is *m*?	Order	Sales	Demand	Profit
1	600	600	600	673	600
2	650	645	525	525	405
3	575	580	580	630	580
4	615	615	604	604	593
5	?	?			

Figure 19: Depiction of the game interface in Study 2. The demand column only appeared for participants in the uncensored condition. This participant next needs to input a best guess for underlying mean demand, m, and an ordering decision for period 5.

Appendix C

Figure 20: A screenshot of the sequential risk taking task in Study 4.

Appendix D

Period	Status after 81 seconds	(If NOT finished by deadline) Total time taken in seconds
1	Finished	
2	Not finished	85
3	Finished	
4	Not finished	98
5	Finished	

Figure 21: A depiction of the first five employee task completion times observed in the censored condition in Study 5. In the uncensored condition, participants could also observe the exact completion times in periods where the employee was finished by the deadline. The deadline for this emploee was set to 81 seconds.

References

Amemiya, T. 1984. Tobit models: A survey. *J. Econometrics* **24** 3–61.

Ariely, D., K. Wertenbroch. 2002. Procrastination, deadlines, and performance: Self-control by precommitment. *Psych. Sci.* **13**(3) 219-224.

Ashford, S. J., L.L. Cummings. 1983. Feedback as an individual resource: Personal strategies of creating information. *Org. Beh. Human Perform.* **32** 370-398.

Audia, P.G., S. Brion. 2007. Reluctant to change: Self-enhancing responses to diverging performance measures. *Org. Beh. Human Dec. Proc.* **102**(2) 255-269.

Audia, P.G., E.A. Locke, K.G. Smith. 2000. The paradox of success: An archival and laboratory study of strategic persistence following radical environmental change. *Academy of Management J.* **43**(5) 837-853.

Babcock, L., S. Laschever. 2003. *Women Don't Ask: Negotiation and the Gender Divide*. Princeton and Oxford: Princeton University Press.

Babcock, L., S. Laschever. 2009. *Ask for it: How women can use the power of negotiation to get what they really want*. New York: Bantam.

Becker, B., B. Gerhart. 1996. The impact of human resource management on organizational performance: Progress and prospects. *Acad. Management J.* **39**(4) 779-801.

Behfar, K., R. Peterson, E.A. Mannix, W. Trochim. 2008. The critical role of conflict resolution in teams: A close look at the links between conflict type, conflict management strategies, and team outcomes. *J. Applied Psych.* **93** 170-180.

Bell, G.D. 1967. Determinants of the span of control. *Amer. J. Sociol.* **73** 100-109.

Belsley, D.A., E. Kuh, R.E. Welch. 1980. *Regression diagnostics: Identifying influential data and source of collinearity*. John Wiley, New York.

Blacksmith, N., J. Harter. 2011. "Majority of American workers not engaged in their jobs." (October 28), The Gallup Organization. http://www.gallup.com/poll/-150383/majority-american-workers-not-engaged-jobs.aspx

Blau, P.M. 1957. Formal organization: Dimensions of analysis. *Amer. J. Sociology* **63**(1) 58-69.

Blau, P.M., R. A. Schoenherr. 1970. *The Structure of Organizations*. New York: Basic Books.

Block, W. 1976. *Defending the undefendable*. San Fransisco: Fox & Wilkes.

Bloomfield, R.J., J. Hales. 2002. Predicting the next step of a random walk: experimental evidence of regime-shifting beliefs. *J. Financial Econom.* **65**(3) 397-414.

Bolton, G.E., E. Katok. 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.

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

Bourgeois, L.J. 1981. On the measurement of organizational slack. *Academy Management Rev.* **6** 29-39.

Bowles, H.R., L. Babcock, L. Lai. 2007. Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Org. Beh. Human Dec. Proc.* **103** 84-103.

Brenner, L.A., D.J. Koehler, A. Tversky. 1996. On the evaluation of one-sided evidence. *J. Behavioral Decision Making* **9** 59-70.

Campbell, D.J., C. Lee. 1988. Self-appraisal in performance evaluation: Development versus evaluation. *Academy Management Rev.* **13**(2) 302-314.

Campbell, A., J. Whitehead, S. Finkelstein. 2009. Why good leaders make bad decisions. *Harvard Business Review* Feb 60-66.

Chapman G.B., E. J. Johnson. 1999. Anchoring, activation, and the construction of values. *Organ. Behav. Human Decision Processes* **79**(2) 115-153.

Chen, L. 2010. Markov decision processes with censored observations: analysis and applications in dynamic pricing. Working Paper, Duke University, The Fuqua School of Business.

Chen, L., A.G. Kok, J.D. Tong. 2011. The effect of payment schemes on inventory: the role of mental accounting. Working Paper, Duke University, The Fuqua School of Business.

Cohen, J., P. Cohen, S.G. West, L.S. Aiken. 2003. *Applied multiple regression/correlation analysis for behavioral sciences*. Lawrence Erlbaum Associates, Mahwah, NJ.

Cohen, W.M., D.A. Levinthal. 1990. Absorptive capacity: A new perspective on learning and innovation. *Admin. Sci. Quart.* **35**(1) 128-152.

Croson, R., K. Donohue. 2003. Impact of POS data sharing on supply chain management: An experimental study. *Production and Oper. Management* **12**(1) 1-11.

Croson, R., K. Donohue. 2006. Behavioral causes of the bullwhip effect and the observed value of inventory information. *Management Sci.* **52**(3) 323-336.

Darley, J.M., B. Latané. 1968. Bystander intervention in emergencies: Diffusion of responsibility. *J Personality Social Psych.* **8**(4) 377-383.

Denrell, J.C. 2005. Why most people disapprove of me: Experience sampling in impression formation. *Psych. Rev.* **112** 951-978.

Denrell, J.C. 2007. Adaptive learning and risk taking. *Psych. Rev.* **114** 177-187.

Denrell, J.C., N. Arvidsson, U. Zander. 2004. Managing knowledge in the dark: An empirical study of the reliability of capability evaluations. *Management Science* **50**(11) 1491-1503.

Eagley, A.H., S.J. Karau. 2002. Role congruity theory of prejudice toward female leaders. *Psych. Rev.* **109**(3) 573-598.

Edmondson, A.C. 1996. Learning from mistakes is easier said than done: Group and organizational influences on the detection and correction of human error. *J. Applied Behavioral Sci.* **32**(1) 5-28.

Einhorn, H.J., R.M. Hogarth. 1978. Confidence in judgment: Persistence of the illusion of validity. *Psych. Rev.* **85** 395-416.

Epley, N., T. Gilovich. 2001. Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psych. Sci.* **12**(5) 391-396.

Epley, N., T. Gilovich. 2004. Are adjustments insufficient? *Personality Social Psych. Bull.* **30** 447–460.

Epley, N., T. Gilovich. 2006. The anchoring and adjustment heuristic. *Psych. Sci.***17** 311–318.

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

Fiedler, K., P. Juslin. 2006. Taking the interface between mind and environment seriously. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition*. 3–32. Cambridge University Press, New York.

Gilovich, T.R. 1991. *How we know what isn't so: The fallibility of human reason in everyday life*. Free Press, New York.

Gioia, D.A., H.P., Jr., Sims. 1985. Self-serving bias and actor-observer differences in organizations: An empirical analysis. *J. Applied Social Psych.* **15**(6) 547-563.

Glick, P., C. Zion, C. Nelson. 1988. What mediates sex discrimination in hiring decisions? *J Personality Soc. Psych.* **55**(2) 178-186.

Hackman, J.R., G.R. Oldham. 1976. Motivation through the design of work: Test of a theory. *Org. Beh. Human Dec. Proc.* **16** 250-279.

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

Harpaz, G., W.Y. Lee, R.L. Winkler. 1982. Learning, experimentation, and the optimal output decisions of a competitive firm. *Management Sci*. **28**(6) 589-603.

Hasija, S., E. Pinker, R.A. Shumsky. 2010. Work expands to fill the time available: Capacity estimation and staffing under Parkinson's law. *Manufacturing Service Oper. Management* **12**(1) 1-18.

Heath, C., R.P. Larrick, J. Klayman. 1998. Cognitive repairs: How organizational practices can compensate for individual shortcomings. *Res. Organ. Behav.* **20** 1–37.

Heilman, M.E. 1983. Sex bias in work settings: The lack of fit model. *Res. Organizational Beh.* **5** 269–298.

Heilman, M.E. 2001. Description and prescription: how gender stereotypes prevent women's ascent up the organizational ladder. *J Soc. Issues* **57**(4) 657-674.

Hekman, D.R., B.P. Owens. 2012. Modeling how to grow: An inductive examination of humble leader behaviors, contingencies, and outcomes. *Academy Management J.* (forthcoming).

Helfat, C.E., M.A. Peteraf. 2003. The dynamic resource-based view: Capability lifecycles. *Strategic Management J.* **24**(10) 997–1010.

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

Hirt, E.R., K.D. Markman. 1995. Multiple explanations: A consider-an-alternative strategy for debiasing judgments. *J. Personality Soc. Psych.* **69**(6) 1069-1086.

Larrick, R.P. 2004. Debiasing. In D. Koehler, N. Harvey, eds. *Blackwell Handbook of Judgment and Decision Making.* Blackwell, Oxford 316-337.

Ho, T-H., N. Lim, T.T. Cui. 2010. Reference-dependence in multi-location newsvendor models: A structural analysis. *Management Sci.* **56**(11) 1891-1910.

Hogan, J., R. Hogan. 1989. How to measure employee reliability. *J. Applied Psych.* **74**(2) 273-279.

Hogarth, R.M. 2001. *Educating Intuition*. Chicago: University of Chicago Press.

Hogarth, R.M., C.R.M. McKenzie, B.J. Gibbs, M.A. Marquis. 1991. Learning from feedback: Exactingness and incentives. *J. Experiment. Psych.: Learning, Memory, and Cognition* **17** 734-752.

Huber, G.P. 1991. Organizational learning: The contributing process and the literatures. *Org. Sci.* **2** 88-115.

Ilgen, D.R., J.L. Barnes-Farrell, D.B. McKellin. 1993. Performance appraisal process research in the 1980s: What has it contributed to appraisals in use? *Org. Beh. Human Dec. Proc.* **54** 321-368.

Ilgen, D.R., C.D. Fisher, M.S. Taylor. 1979. Consequences of individual feedback on behavior in organizations. *J. Applied Psych.* **64** 349-371.

Jehn, K., G. Northcraft, M. Neale. 1999. Why differences make a difference: A field study of diversity, conflict, and performance in workgroups. *Admin. Sci. Quarterly* **44** 741–763.

Jones, E. E., T.S. Pittman. 1982. Toward a general theory of strategic self-presentation. In J. Suls (Ed.), *Psychological perspectives on the self* (231-261). Hillsdale, NJ: Lawrence-Erlbaum.

Joyce, A. 2005. "Boredom numbs the work world." The Washington Post (August 10).

Juslin, P. 1994. The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organ. Behav. Human Decision Processes* **57** 226–246.

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

Juslin, P., A. Winman, H. Olsson. 2000. Naive empiricism and dogmatism in confidence research: A critical examination of the hard–easy effect. *Psych. Rev.* **107** 384–396.

Kahneman, D., A.Tversky. 1973. Subjective probability: A judgment of representativeness. *Cognitive Psych.* **3** 430-454.

Kang, S.C., S.S. Morris, S.A. Snell. 2007. Relational archetypes, organizational learning, and value creation. *Academy Management Rev.* **32** 236-256.

Klayman, J., Y.W. Ha. 1987. Confirmation, disconfirmation, and information in hypothesis testing. *Psych. Rev.* **94**(2) 211-228.

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

Koehler, J.J., M. Mercer. 2009. Selection neglect in mutual fund advertisements. *Management Sci.* **55**(7) 1107-1121.

Koehler, J.J., W.C. Thompson. 2006. Mock jurors' reactions to selective presentation of evidence from multiple-opportunity searches. *Law Human Behav.* **30** 455-468.

Kremer, M., B. Moritz, E. Siemsen. 2010. Demand forecasting behavior: system neglect and change detection. Working paper, Penn State University.

Larrick, R.P., G. Wu. 2007. Claiming a large slice of a small pie: Asymmetric disconfirmation in negotiation. *J. Personality Social Psych.* **93**(2) 212-233.

Lariviere, M.A., E. L. Porteus. 1999. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Sci.* **45**(3) 346-363.

Latham, G.P., E.A. Locke. 1975. Increasing productivity with decreasing time limits: A field replication of Parkinson's law. *J. Applied Psych.* **60**(4) 524-526.

Lejuez, C. W., W.M. Aklin, H.A. Jones, J.B. Richards, D.R. Strong, C.W. Kahler, et al. 2003. The Balloon Analogue Risk Task (BART) differentiates smokers and nonsmokers. *Experiment. Clinical Psychopharm.* **11** 26–33.

Lejuez, C. W., J.P. Read, C.W. Kahler, J.B. Richards, S.E. Ramsey, G.L. Stuart, et al. 2002. Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *J. Experiment. Psych.: Applied* **8** 75–84.

Levinthal, D.A., J.G. March. 1993. The myopia of learning. *Strategic Management J.* **14** 95-112.

Lockley, S.W., J.W. Cronin, E.E. Evans, et al. (2004) Effect of reducing interns' weekly work hours on sleep and attentional failures. *New England J. Med.* **351**(18) 1829-1837.

Lurie, N.H., J.M. Swaminathan. 2008. Is timely information always better? The effect of feedback frequency on decision making. *Organ. Behav. Human Decision Proc*esses **108** 315-329.

MacLeod, C., L. Campbell. 1992. Memory accessibility and probability judgments: An experimental evaluation of the availability heuristic. *J. Personality Social Psych.* **63**(6) 890-902.

Maddox, W.T., F.G. Ashby, C.J. Bohil. 2003. Delayed feedback effects on rule-based and information-integration category learning. *J. Experiment. Psych.: Learning, Memory, Cognition* **29**(4) 650-662.

Manzoni, J.F., J.L. Barsoux. 2002. *The set-up-to-fail syndrome: How good managers cause great people to fail*. Boston: Harvard Business School Press.

Massey, C., G. Wu. 2005. Detecting Regime Shifts: The Causes of Under- and Overreaction. *Management Sci.* **51**(6) 932-947.

March, J.G. 1991. Exploration and exploitation in organizational learning. *Organization Sci.* **2** 71-87.

March, J.G. 1996. Learning to be risk averse. *Psych. Rev.* **103** 309-319.

Markle, A. 2009. Asymmetric disconfirmation in managerial beliefs about employee motivation. Unpublished doctoral dissertation, University of Chicago.

Markle, A. 2011. Dysfunctional learning in decision processes: The case of employee reciprocity. *Strategic Management J*. **32** 1411-1425.

McGrath, R.G. 2001. Exploratory learning, innovative capacity, and managerial oversight. *Academy of Management J.* **44** 118-131.

Morrison, E.W., F.J. Milliken. 2000. Organizational silence: A barrier to change and development in a pluralistic world. *Academy Management Rev.* **25**(4) 706-725.

Morrison, E.W., C.C. Phelps. 1999. Taking charge at work: Extrarole efforts to initiate workplace change. *Academy Management J.* **42**(4) 403-419.

Muller, A., D. Stoyan. 2002. *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons: New York.

Mynatt, C.R., M.E. Doherty, R.D. Tweney, 1977. Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly J. Experiment. Psych.* **29** 85-95

Mynatt, C., S.J. Sherman. 1975. Responsibility attribution in groups and individuals: A direct test of the diffusion of responsibility hypothesis. *J. Personality Social Psych.* **32**(6) 1111-1118.

Nahmias, S. 1994. Demand estimation in lost sales inventory systems. *Naval Res. Logistics* **41** 739-757.

Nickerson, R.S. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. Gen. Psych.* **2** 175–220.

Nohria, N., R. Gulati. 1996. Is slack good or bad for innovation? *Academy Management J.* **39**(5) 1245-1264.

O'Donoghue, T., M. Rabin. 2001. Choice and procrastination. *Quarterly J. Econom.* **116**(1) 121-160.

Ouchi, W.G., J.B. Dowling. 1974. Defining span of control. *Administrative Science Quarterly* **19**(3) 357-365.

Pleskac, T. J. (2008). Decision making and learning while taking sequential risks. *J. Experiment. Psych.: Learning, Memory, and Cognition* **34** 167–185.

Reb, J., T. Connolly. 2009. Myopic regret avoidance: Feedback avoidance and learning in repeated decision making. *Organizational Behavior and Human Decision Processes* 109(2) 182-189.

Roese, N.C. 1997. Counterfactual Thinking. *Psychological Bulletin.* **121**(1) 133-148.

Roese, N.J., J.M. Olsen. 1995. *What might have been: The social psychology of counterfactual thinking*, Lawrence Erlbaum Associates, Mahwah, NJ.

Rudi, N., D. Drake. 2009. Level, adjustment and observation biases in the newsvendor model. Working paper, INSEAD.

Rudman, L.A. 1998. Self-promotion as a risk factor for women: The costs and benefits of counter-stereotypical impression management. *J. Personality Soc. Psych.* **74** 629–645.

Rudman, L.A., P. Glick. 2001. Prescriptive gender stereotypes and backlash toward agentic women. *J. Soc. Issues* **57**(4) 743-762.

Schweitzer, M.E., G.P Cachon. 2000. Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Sci.* **46**(3) 404-420.

Shapira, Z. 1995. *Risk taking, a managerial perspective.* New York: Russell Sage Foundation.

Sichel, D.E. 1993. Business cycle asymmetry: A deeper look. *Economic Inquiry* **31**(2) 224-236.

Silverman, G.K., G.F. Loewenstein, B.L. Anderson, P.A. Ubel, S. Zinberg, J. Schulkin. 2010. Failure to discount for conflict of interest when evaluating medical literature: a randomized trial of physicians. *J. Medical Ethics* **36** 265-270.

Simon, H.A. 1945. *Administrative Behavior*. New York: Free Press.

Simon, H.A. 1990. Invariants of human behavior. *Annual Review of Psychology* **41** 1-19.

Simons, T.L., R.S. Peterson. 2000. Task conflict and relationship conflict in top management teams: The pivotal role of intragroup trust. *J. Applied Psych.* **85** 102–111.

Sitkin, S. 1992. Learning through failure: The strategy of small losses. *Res. Organ. Behav.* **14** 231-266.

Sitkin, S., A.L. Pablo. 1992. Reconceptualizing the determinants of risk behavior. *Acad. Management Rev.* **17**(1) 9-38.

Sitkin, S.B., K.E. See, C.C. Miller, M.W. Lawless, A.D. Carton. 2011. The paradox of stretch goals: Organizations in pursuit of the seemingly impossible. *Academy Management Rev.* **36**(3) 544-566.

Sitkin, S., K.M. Sutcliffe, R.G. Schroeder. 1994. Distinguishing control from learning in total quality management: A contingency perspective. *Academy of Management Rev.* **19**(3) 537-564.

Sitkin, S.B., L.R. Wiengart. 1995. Determinants of risky decision-making behavior: A test of the mediating role of risk perceptions and propensity. *Academy Management J.* **38** 1573-1592.

Strack, F., T. Mussweiler. 1997. Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *J. Personality Social Psych.* **73** 437-446.

Stewart, N., N. Chater, G.D.A. Brown. 2006. Decision by sampling. *Cognitive Psych.* **53** 1–26.

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

Swann, W.B., S.J. Read. 1981. Acquiring self-knowledge: The search for feedback that fits. *J. Personality Soc. Psych.* **41** 1119-1128.

Thomas, J.P., D.S. Whitman, C. Viswesvaran. 2010. Employee proactivity in organizations: A comparative meta-analysis of emergent proactive constructs. *J Occupat. Org. Psych.* **83** 275-300.

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

Tvesky, D., D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* **185** 1124-1130.

Tversky, A., D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* **211** 453-458.

Urwick, L.F. 1956. The manager's span of control. *Harvard Business Review* **34** 39-47.

Van Nieuwerburgh, S., L.L. Veldkamp. 2006. Learning asymmetries in real business cycles. *J. Monetary Econom.* **53** 753-772.

Veldkamp, L.L. 2005. Slow boom, sudden crash. *J. Economic Inquiry* **124** 230-257.

Wallach, M.A., N. Kogan, D.J. Bem. 1964. Diffusion of responsibility and level of risk taking in groups. *J. Abnormal Social Psych* **68**(3) 263-274.

Wallsten, T. S., T.J. Pleskac, C.W. Lejuez. 2005. Modeling behavior in a clinically-diagnostic sequential risk-taking task. *Psych. Rev.* **112** 862–880.

Whyte, G. 1991. Diffusion of responsibility: Effects on the escalation tendency. *J Applied Psych.* **76**(3) 408-415.

Winman, A., P. Hansson, P. Juslin. 2004. Subjective probability intervals: How to cure overconfidence by interval evaluation. *J. Experiment. Psych.: Learning, Memory, and Cognition* **30** 1167–1175.

Zhao, X., Lynch, J.G., Jr, Q. Chen. 2010. Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *J. Consumer Res.* **37** 197–206.

Biography

Daniel Feiler was born in 1985 and raised in Pittsburgh, PA. He received his B.S. degree from Carnegie Mellon University (May, 2007) in Economics and Decision Science. He has a paper published in *Climatic Change* and papers forthcoming in *Management Science* and the *Journal of Experimental Social Psychology*. While at Duke, he was a recipient of the James B. Duke fellowship.