Probability Weighting and the Newsvendor Problem: Theory and Evidence

Xiao Huang
John Molson School of Business, Concordia University, Montreal, QC, Canada
xiao.huang@concordia.ca

Mahesh Nagarajan
Sauder School of Business, University of British Columbia, Vancouver, BC, Canada
mahesh.nagarajan@sauder.ubc.ca

Shanshan Guo
School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai, China
guo.shanshan@sufe.edu.cn

Behavioral studies have consistently reported the pull-to-center (PTC) effect observed in laboratory studies of the newsvendor problem. We examine whether this and some other observed effects can be reconciled under the general framework of prospect theory without special assumption on reference point. Specifically, we allow decision makers to value each potential outcome with decision weight rather than the actual probability and further validate through some experiments if the PTC effect is correlated with the shape of the weighting function. The results confirm that the general framework of prospect theory can explain the PTC effect without special assumption on reference point. The only stipulation is a general set of probability weighting functions that that admit underweighting of small probabilities. The proposed model is inclusive, robust, and can explain a number of prominent newsvendor behavioural observations with reasonable benchmark predictions. Our finding suggests that the systematic, suboptimal decisions in behavioural newsvendors could be due to over-focusing on the big picture and neglecting rare events. Accordingly, both academicians and practitioners should revisit the design of decision support systems as well as competitive strategies that involve human decision making against large number of uncertain future outcomes.

Key words: newsvendor; risk and uncertainty; prospect theory; decision weight; probability weighting; pull to center; behavioral operations;

History: October 6, 2021

---

1. Introduction

The newsvendor problem is a cornerstone of operations management. A simple and well studied instance of the model involves a newsvendor (the decision maker) facing unit cost c and retail price r determining an order quantity q of an item against an uncertain future demand with cumulative distribution function $F(\cdot)$. Under expected utility theory, a risk neutral newsvendor will place an order q^* that maximizes the expected profit by solving:

$$F(q^*) = 1 - \frac{c}{r} \tag{1}$$

where the RHS, known as the critical fractile, reflects the optimal no-stockout probability.

A large body of literature in behavioral operations examines how human subjects respond to this decision problem in laboratory settings (e.g., Schweitzer and Cachon 2000, Bolton and Katok 2008, Bostian et al. 2008, Kremer et al. 2010, and Ockenfels and Selten 2014). These studies consistently report a *pull-to-center* (PTC) effect. That is, decision makers tend to place an order $\hat{q}$ in between the mean of the demand μ, and the quantity q^* that maximizes the expected profit (thereafter referred to as "the newsvendor optimum"). Specifically, in the *high-profit* scenario (c/r is low) where q^* is greater than μ, people will usually *underorder* ($\mu < \hat{q} < q^*$); and in the *low-profit* scenario (c/r is high) where q^* is less than μ, people will normally *overorder* ($q^* < \hat{q} < \mu$).

Various explanations have been proposed as an explanation for this PTC effect including waste or stockout aversion, underestimating opportunity cost, a desire to reduce ex post inventory error (Schweitzer and Cachon 2000), bounded rationality (Su 2008), and overconfidence (Ren and Crosen 2013). However, most (if not all) fail to accommodate individual level heterogeneity and deliver consistent within-subject prediction at the same time (Uppari and Hasija 2019). More paramountly, a systematic explanation that rests on a theoretical model validated by experiments is missing. Of particular interest is the question whether behavioral frameworks such as prospect theory (Kahneman and Tversky 1979), which entails the role of *reference point*, *value* and *decision weight* in one's utility, could explain the PTC effect. Schweitzer and Cachon (2000) somewhat surprisingly find that part of their observation cannot be properly explained by prospect theory. This conclusion was arguably insufficient, as the role of *decision weight* was ignored in their arguments. Weighting functions are an integral part of prospect theory. Empirical evidence (e.g., Tversky and Kahneman 1992, Tversky and Wakker 1995, and Gonzalez and Wu 1999) suggests that subjects tend to overweight low probability events and underweight high probability events. A more recent paper (Nagarajan and Shechter 2014) explicitly incorporates *decision weights* that overweight small probabilities, and a *value* function satisfying the principal of diminishing sensitivity, into the newsvendor model. They find that the original observation in Schweitzer and Cachon (2000) that prospect theory cannot fully explain the results from the laboratory experiments still holds.

Noting that both Schweitzer and Cachon (2000) and Nagarajan and Shechter (2014) use *status quo* as the *reference point*, a number of recent studies attempt to consider a variant on *reference points* with a break on *decision weight*. For instance, Long and Nasiry (2015) prove the PTC effect in various scenarios by allowing the newsvendor to use a reference point based on a weighted average of the maximum and minimum profits resulting from an order quantity. The reference point is then a non-state-dependent value varying across different prospects. Their model is assessed to be dominated by a state-dependent reference point model in Uppari and Hasija (2019), which suggests that a decision maker would evaluate each prospect based on the profit it would otherwise obtain by ordering at the mean. Since the demand is stochastic, so is the reference point. However, empirical evidence does not lend much support to exogenous reference points in these specific formations – static reference points such as *status quo* are found to be most commonly used (Baillon et al. 2020), while stochastic reference points (Sugden 2003, Köszegi and Rabin 2006, Köszegi and Rabin 2006) have more grounds as endogenized rational expectations (e.g., Crawford and Meng 2011, De Giorgi and Post 2011, Lien and Zheng 2015, Baron et al. 2015). Therefore, it remains an open question whether the general framework of prospect theory can (or, why it could not) explain the PTC effect.

This paper takes an important step towards answering the above question. We show that the generalized framework of prospect theory, without any special assumption on the *reference point* or *value* function, can explain the PTC effect in the newsvendor problem. The only stipulation is on the *decision weight*, where the restrictive use of functions that overweight small probabilities is relaxed. That is, we allow low probabilities to be overweighted as well as underweighted. This relaxation of weighting functions has theoretical foundations based on recent set of documented studies which we will describe below in the next section. Further, as will be discussed later, this fits well with the environments of the laboratory studies of the newsvendor problem that are reported in the operations management literature.

Our contributions are manifold. First, our theoretical results show that under prospect theory with the generalized weighting functions, the PTC and some other effects observed in lab experiments can be reconciled. As will be clear from later sections of this paper, we make little to no assumptions on the functional forms that specify the decision weights other than their shapes. We show that when the weighting functions are of S shape, PTC effect holds for both positive prospects (with no assumption on utility function or demand distribution) and mixed prospects (under linear utility with mild conditions on the demand distribution and loss aversion parameters). When the weighting functions are inverse-S shaped, the PTC effect ceases to exist in general. These two sets of theoretical results allows us to explain both the PTC effect at the aggregate level and heterogeneous decisions at the individual level observed in laboratory newsvendor experiments.

Second, we test using a set of laboratory experiments if the PTC effects are indeed correlated to the shape of the weighting functions as per the above theoretical findings. To do this, we expose newsvendor subjects to decision paradigms that would impact the shape of weighting functions as recent behavioural studies have found. In parallel, we elicit parameters that would describe the shape of the weighting function for individual subjects and use it to reconcile the above findings. The results indicate that the above explanation is quite plausible.

Lastly, we conduct robustness analysis and confirm that our model holds under a set of alternative reference points and provides more reasonable and inclusive benchmark predictions than existing models in literature.

2. Probability Weighting: Overweighting vs. Underweighting

Under the framework of prospect theory, individuals choose among prospects by multiplying their values of potential outcomes with decision weights rather than the stated probabilities. Empirical evidence based on lottery experiments suggests that low probabilities are likely to be overweighted while high probabilities are underweighted; the probability weighting functions thereby exhibit an inverse-S shape (Kahneman and Tversky 1979, Tversky and Kahneman 1992, Tversky and Wakker 1995, Gonzalez and Wu 1999).

In context beyond pure lotteries, the evidence is mixed. In marketing, consumers were found to act in contrary to overweighting-small-probabilities when confronted with uncertain prices (e.g., Danziger et al. 2014, Gaertig and Simmons 2018). Similar behaviors have been reported in finance and risk management when subjects were asked to forecast stock price changes (Durbach and Montibeller 2018) or to prepare for crisis (Chaudhry et al. 2018). Investment communities are found to often underestimate the occurrence of "tail risk" events (DeCambre 2020). Popular press credited the potential underweighting of rare events in part leading to many US hospitals under stocking critical supplies needed in the event of a pandemic like COVID19 (de Puy Kamp 2020).

The disparity in probability weighting has been heavily discussed in the realm of *description-experience gap* in behavioral decision theory (e.g., Hertwig et al. 2004, Barron and Yechiam 2009, Camilleri and Newell 2011), which finds that small probabilities could be underweighted or overweighted depending on how information is coded and presented. Specifically, when the prospects are provided in a direct, straightforward and *descriptive* manner as those in the lottery experiments, subjects tend to overweight small probabilities. Otherwise, when the probabilities are learnt from one's own *experience* or encounters, such as an exhaustive sampling of binary balls in Barron and Ursino (2013), underweighting small probabilities has a stronger presence.

The underweighting small probabilities in the *experience* paradigm is robust to the extent that studies have ruled out a number of common causes such as under-sampling, recency or judgement errors (e.g., Ungemach et al. 2009, Hertwig and Erev 2009, Barron and Ursino 2013) leaving open to other plausible explanations such as tallying (Hills and Hertwig 2010). Furthermore, underweighting rare events also has a footprint in some *descriptive* settings when the presentation of the prospects deviates from that in pure lotteries (Palma et al. 2014).

Subsequent to these findings, a stream of research further quantifies decision weights in the *description-experience* paradigms and identifies empirical evidence that the probability weighting functions may also exhibit S shape (e.g., Ungemach et al. 2009, Camilleri and Newell 2013, Frey et al. 2015, Lejarraga and Müller-Trede 2017, Dai et al. 2019). Although the support is not unanimous (e.g., Abdellaoui et al. 2011 and Lejarraga et al. 2016), the common sentiment is that "inverse-S (weighting function) ... is certainly not universal" (van de Kuilen and Wakker 2011), and that there exists systematic difference between the two paradigms, leading to either underweighting or less overweighting of rare events in the *experience* paradigm (Wulff et al. 2018).

2.1 Decision Paradigms in Laboratory Newsvendors

Over the arch of the *description-experience* paradigms, we find the newsvendor problem and related laboratory studies leaning towards the latter. In real life, a newsvendor's profit prospects can hardly be *described* as pure lotteries. Managers typically need to exert efforts reviewing historical sales and profits, understanding current cost and market structures, in making a rational ordering decision. This course of actions is similar to an exhaustive sampling of potential outcomes in Barron and Ursino (2013) thus heavy in *experience*.

For the above reason, laboratory newsvendor experiments inevitably bear similar characteristics. A review of literature suggest that most, if not all, laboratory newsvendor experiments carry some *experience* traits – although participants were informed of demand distributions, the prospects of profit or final state of wealth were still unclear and called for further deliberation. For instance, in most studies, the connection between an order decision and its profit prospect is not straightforward. In Schweitzer and Cachon (2000), participants could order any quantity between 1 and 300, and they needed to solicit among a large pool of information to learn the profit distribution associated with each order quantity. In Bolton and Katok (2008) the subjects were provided with historical profits based on past demand realization, but one still had to refer to instructional formula and did the algebra to obtain a full picture of potential profits and associated probabilities. In other experiments such as Bostian et al. (2008) and Ockenfels and Selten (2014), the assistance was even less; similar to real life setting, subjects were merely informed of the context and had to rely on their own means for effective decision making.

The experiments by Kremer et al. (2010) are arguably the most *descriptive* among all as they had provided the subjects with an explicit table containing profit prospects associated with each choice (ordering quantity). However, there are distinctions between the profit table and pure lotteries. For one thing, the profit table presents much larger numbers of prospects and scenarios for a decision maker to digest than pure lotteries. For another, instead of describing a prospect as (€8.6, 6/7; €4.9, 1/7) as a pure lottery would do, the profit table and the accompanying description had to elaborate that one would receive a payoff of €4.9 if a random number turns out to be between 1 and 14 (with probability 1/7), €8.6 if the random number is between 43-56 (with probability 1/7), €8.6 if the random number is between 85-100 (with probability 1/7), so on and so forth. This process of "unpacking" occurrence of outcomes proportional to their true probabilities carries similar effect to the *experience* paradigm as summarized in Palma et al. (2014).

In addition, the relation between the earned profit and one's final payoff can be obscure. For example, in Schweitzer and Cachon (2000), de Véricourt et al. (2013) and Rudi and Drake (2014), only one or few out of all participants may receive a final award, the chance or the amount of which was commensurate to their earned profits in the experiment. In Kremer et al. (2010), only a random 2 out of 60 rounds of the decisions will count towards one's final pay off. Therefore, even if the participants were well motivated to maximize their profits, it is hard to articulate to what extent one's choice (order quantity) may affect its future payoff.

In summary, newsvendors (in both real life and laboratory environment) have to deal with non-trivial *experience* elements in the course of their decision making. We henceforth conjecture that it is reasonable to consider a broader spectrum of decision weights than those elicited from purely *descriptive*, lottery-based experiments.

2.2 Generalized Weighting Functions

In light of the above discussion, we re-examine the applicability of prospect theory in newsvendor experiments, where *value* and *reference point* remain generic but the *decision weight* is broadened to accommodate both the inverse-S and S shapes. The inverse-S shaped function, as depicted in the grey curve in Figure 1, overweights small probabilities and underweights large probabilities. In the *description-vs-experience* framework, such can be configured for the *description* paradigm. Conversely, the S shaped weighting function in dark curve underweights small probabilities and overweights large probabilities hence is suitable for the *experience* paradigm.

In literature, the weighting functions can be obtained by properly configuring some one- or two-parameter weighting functions. One frequently used weighting function was proposed in Tversky and Kahneman (1992), that $w^+(p) = \frac{p^\gamma}{(p^\gamma+(1-p)^\gamma)^{1/\gamma}}$ can be used for the gains and $w^+(p) = \frac{p^\delta}{(p^\delta+(1-p)^\delta)^{1/\delta}}$ for the loss. When the parameters γ, δ are less than 1, the weighting function is of inverse-S shape that overweights small probabilities and underweights large probabilities; when the parameters are greater than 1, it becomes S shaped and underweights small probabilities.

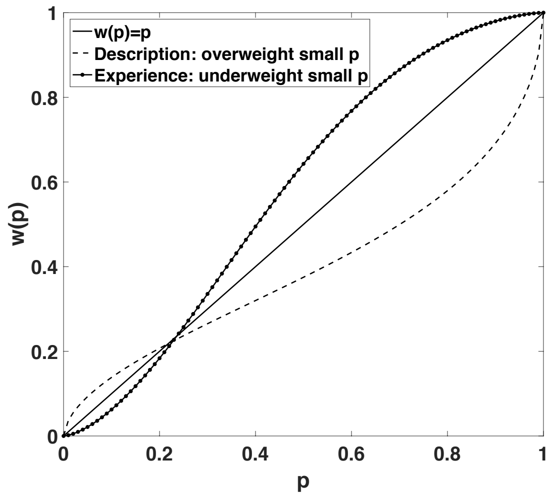


Figure 1 **Decision Weights under Different Paradigms.**

Ungemach et al. (2009) calibrate a set of *experience*-based experimental results to the said weighting function and find best fit for parameters over 1. By assuming $\gamma = \delta$, Camilleri and Newell (2013) also identify similar pattern. Lejarraga and Müller-Trede (2017) estimate parameter sets for the averages in each paradigm and obtained weighting function shapes similar to Figure 1 in one of their experiments.

Other options that tolerate both the S- and inverse-S shape weighting functions include $w(p) = e^{-\delta(-\ln p)^{\gamma}}$ proposed in Prelec (1998) — the simplified form ($\delta = 1$) of which has been often used, and the "linear in log odds" function introduced in Goldstein and Einhorn (1987) and further assessed by Gonzalez and Wu (1999): $w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma}+(1-p)^{\gamma}}$. In both cases, $\gamma > 0$ controls the curvature while $\delta > 0$ controls the elevation, and underweighting small probability is more likely to appear among larger γ. In the *experience paradigm*, Frey et al. (2015) apply the function in Prelec (1998) to estimate parameters for younger vs. older adults which, albeit different, are both above 1. More recently, Dai et al. (2019) use the "linear in log odds" function and estimate the γ to be over 1 in the *experience* paradigm and less than 1 in the *description* paradigm of their experiments.

Before diving into our analysis, we describe the main theoretical results, as summarized in Table 1. First, with S shaped weighting functions, when all prospects are positive, we find that PTC holds when subjects are risk neutral or risk averse assuming no specific functional forms. When the prospects are mixed, and subjects are loss averse, we show that PTC holds under some mild assumptions. This analysis can be found in §3.1. Next, when weighting functions are inverse-S shaped, we show in §3.2 and Appendix B that PTC does not systematically hold.

Table 1 Summarization of Theoretical Results

Exhibit PTC Effect?	**S shaped** Weighting	**Inverse-S shaped** Weighting
Positive Prospects		
Risk Neutral	Yes	No
Risk Averse	Yes	Not systematically but plausible in high profit scenarios.
Mixed Prospects		
Risk Neutral/Low Loss Aversion	Yes	No
Risk Neutral/High Loss Aversion	Not systematically but plausible in high profit scenarios.	
Risk Averse	*Similar insights supported by numerical experiments.	

3. Newsvendor under Generalized Weighting Functions

We now revisit the classical newsvendor problem under the framework of prospect theory. Assume that the demand is stochastically distributed on $[\underline{d}, \bar{d}]$, where $0 < \underline{d} < \bar{d}$, with cumulative distribution function $F(\cdot)$ and probability density function $f(\cdot)$. The reference point is *status quo*. The value function $u(x)$ satisfies $u(0) = 0$, $u^{'} \geq 0$, and the following principles:

- *risk-averse for the gains*: $u^{''}(x) \leq 0$ for all $x \geq 0$;
- *risk-seeking for the losses*: $u^{''}(x) \geq 0$ for all $x < 0$;
- *loss aversion*: $u^{'}(x) < u^{'}(-x)$ for all $x > 0$.

Consider a generalized weighting function $w(\cdot)$ and denote $\dot{w}(p) = \partial w(p)/\partial p$ and $\partial^2 w(p)/\partial p^2 = \ddot{w}(p)$. The weighting function is assumed to satisfy the following properties:

A1: $w(0) = 0$ and $w(1) = 1$;

A2: $\dot{w}(p) \geq 0$ for any $p \in [0, 1]$;

A3: there exists a unique $p_0 \in (0, 1)$ such that

- *S shape*: $w(p) \leq p$ and $\ddot{w}(p) \geq 0$ for $p \in [0, p_0]$; $w(p) \geq p$ and $\ddot{w}(p) \leq 0$ for $p \in [p_0, 1]$;
- *Inverse-S shape*: $w(p) \geq p$ and $\ddot{w}(p) \leq 0$ for $p \in [0, p_0]$; $w(p) \leq p$ and $\ddot{w}(p) \geq 0$ for $p \in [p_0, 1]$.

The first two conditions entail the consistency of the weighting functions such that higher probability always receives higher weight. The third condition ensures the shape of the weighting function with overweighting and underweighting switching at p_0. Each outcome will be weighted by the difference in the weights of respective cumulative probabilities following the Cumulative Prospect Theory (CPT) proposed in Tversky and Kahneman (1992), which will be elaborated during the analysis. All technical proofs can be found in Appendix A.

3.1 Newsvendor under S Shaped Weighting Functions

We first analyze whether prospect theory with S shaped weighting function may explain the PTC effect observed in laboratory newsvendor experiments.

3.1.1 Strictly Positive Prospects In line with existing literature (Schweitzer and Cachon 2000, Nagarajan and Shechter 2014), we first examine the scenario where all possible profit realizations are positive, i.e., $r\underline{d} - c\bar{d} \geq 1$. According to CPT, in a prospect with n positive, ordered outcomes $(x_1, p_1; x_2, p_2; ...; x_n, p_n)$, the i^{th} outcome x_i will receive weight $w(\sum_{j=i}^{n} p_j) - w(\sum_{j=i+1}^{n} p_j)$. The value out of order quantity q is then given by

$$V(q) = \int_0^q u(rx - cq)[-w'(\bar{F}(x))]dx + u(rq - cq)w(\bar{F}(q)), \tag{2}$$

where $w'(\cdot) = \partial w(\bar{F}(x))/\partial x < 0$. The first and second order derivatives with respect to the order quantity q are:

$$V'(q) = \int_0^q cu'(rx - cq)w'(\bar{F}(x))dx + (r - c)u'(rq - cq)w(\bar{F}(q)), \tag{3}$$

$$\begin{aligned} V''(q) = &\int_0^q -c^2u''(rx - cq)w'(\bar{F}(x))dx + cu'(rq - cq)w'(\bar{F}(q)) \\ &+(r-c)^2u''(rq - cq)w(\bar{F}(q)) + (r - c)u'(rq - cq)w'(\bar{F}(q)). \end{aligned} \tag{4}$$

Since $u' \geq 0, u'' \leq 0$ and $w' < 0$, there is $V''(q) < 0$. Thus a subject's value is concave in q and uniquely maximized at $\hat{q}^*$ where $V'(\hat{q}^*) = 0$. For risk-neutral subjects, we have the following results:

Proposition 1. **(Pull-To-Center under Risk Neutrality)**

(i) A risk-neutral subject ($u' \equiv k$ for some $k > 0$) orders at

$$w(\bar{F}(\hat{q}^*)) = \frac{c}{r} := 1 - \textit{critical fractile}. \tag{5}$$

(ii) There is overorder ($q^* < \hat{q}^* < q_0$) in the low profit scenario ($c/r > p_0$) and underorder ($q_0 < \hat{q}^* < q^*$) in the high profit scenario ($c/r < p_0$), where $q^* = \bar{F}^{-1}\left(\frac{c}{r}\right)$ and $q_0 = \bar{F}^{-1}(p_0)$.

(iii) The pull-to-center effect is minimized at $c/r = p_0$.

Therefore, as opposed to the classical newsvendor who directly sets the stockout rate $\bar{F}(q^*)$ to $(1 -$ critical fractile), a risk-neutral subject will instead configure the weight of the stockout rate $w(\bar{F}(\hat{q}^*))$ to the same fraction, $1 -$ critical fractile. The anchoring effect is subsequently confirmed among risk-neutral subjects, thus PTC holds even in absence of risk aversion.

We next investigate the scenario with risk averse utility functions. In exploring this issue, we apply the Arrow-Pratt measure of *absolute risk aversion* (Pratt 1964, Arrow 1965), namely,

$ARA(u) = -u''/u'$, in assessing the degree of risk aversion among the subjects. That is, a subject is more risk averse if it possesses a higher $ARA(u)$. The following result concerns the impact of risk aversion on the subjects' order quantity $\hat{q}^*$:

Lemma 1. *$\hat{q}^*$ decreases in the degree of absolute risk aversion.*

Therefore, as risk aversion develops among the subjects, the order quantity $\hat{q}^*$ reduces itself accordingly. The subsequent question is whether the PTC effect retains during this process. We summarize our findings in the following proposition:

Proposition 2. **(Pull-To-Center under Risk Aversion)** *For a strictly risk-averse subject ($u'' < 0$),*

(i) there exists a $z_0 \in (p_0, 1)$ such that the pull-to-center effect is minimized at $c/r = z_0$, i.e., $\hat{q}^* = q^* = \bar{F}^{-1}(z_0)$;

(ii) there is overorder ($q^* < \hat{q}^* < \bar{F}^{-1}(z_0)$) in the low profit scenario ($c/r > z_0$) and underorder ($\bar{F}^{-1}(z_0) < \hat{q}^* < q^*$) in the high profit scenario ($c/r < z_0$).

Lemma 1 and Proposition 2 together suggest that even though the entire order will shift "left" as the subject becomes more risk averse, such shifts are bounded so that the anchoring effect still holds. Unless the subject is facing the unlikely moderate profit scenario ($c/r = z_0$) in which the behavioral and the theoretical optimum coincide with each other ($\hat{q}^* = q^*$), the behavioral optimum is always higher than the theoretical optimum ($\hat{q}^* > q^*$) in a low profit scenario ($c/r > z_0$), or lower ($\hat{q}^* < q^*$) in a high profit scenario ($c/r < z_0$).

3.1.2 Mixed Prospects We next look into the general scenarios where profit realizations can be both positive and negative, i.e., $r\underline{d} - c\bar{d} < 0$. In this case, the CPT decision weights are subject to a small modification due to the existence of negative prospects. That is, in a prospect with n positive and m negative outcomes, ordered by $(x_1^-, p_1^-; \ldots x_m^-, p_m^-; x_1^+, p_1^+; \ldots; x_n^+, p_n^+)$, the i^{th} positive outcome x_i^+ will receive weight $w(\sum_{j=i}^n p_j) - w(\sum_{j=i+1}^n p_j)$, and the the k^{th} negative outcome x_k^- will receive weight $w(\sum_{j=1}^k p_j) - w(\sum_{j=1}^{k-1} p_j)$.

For each order quantity q, the break-even point is cq/r. The value of the order quantity q is then given by:

$$V(q) = \int_0^{cq/r} u(rx - cq)w'(F(x))dx + \int_{cq/r}^{q} u(rx - cq)[-w'(\bar{F}(x))]dx + u(rq - cq)w(\bar{F}(q)) \quad (6)$$

where $w'(\cdot) = \partial w(\cdot)/\partial x$. Following a similar analysis, the subject's order can be characterized as follows:

LEMMA 2. *A subject's order quantity $\hat{q}^*$ can be determined by:*

$$w(\bar{F}(\hat{q}^*)) = \frac{c}{r-c}\int_0^{c\hat{q}^*/r} \frac{u^{'}(rx - c\hat{q}^*)}{u^{'}(r\hat{q}^* - c\hat{q}^*)} w^{'}(F(x))dx + \frac{c}{r-c}\int_{c\hat{q}^*/r}^{\hat{q}^*} \frac{u^{'}(rx - c\hat{q}^*)}{u^{'}(r\hat{q}^* - c\hat{q}^*)}[-w^{'}(\bar{F}(x))]dx. \quad (7)$$

In the scenario where the subjects are risk-neutral, but suffer more from loss than from the same gain, i.e., $\forall x > 0, \lambda u^{'}(x) = u^{'}(-x) \equiv k$ for some $k > 0$ and $\lambda > 1$, by substituting $u^{'}(rx - cq)/u^{'}(rq - cq) = \lambda$ into (7), the order quantity $\hat{q}^*$ satisfies:

$$w(\bar{F}(\hat{q}^*)) = \frac{c}{r}\left[\lambda w\left(F(\frac{c\hat{q}^*}{r})\right) + w\left(\bar{F}(\frac{c\hat{q}^*}{r})\right)\right]. \quad (8)$$

In general, the ordering behavior is more complicated under mixed prospects mainly due to the dynamics introduced by loss aversion. When the subjects are not severely loss averse, we can show that the same over (under) order fashion will hold in the low (high) profit scenario after imposing some mild technical assumptions on the demand and weighting functions.

PROPOSITION 3. **(Pull-To-Center under Low Loss Aversion)** *Assume that the demand distribution is symmetric with increasing generalized failure rate (IGFR), i.e.,* $\frac{qf(q)}{\bar{F}(q)}$ *increases in* q*, the weighting function* $w(\cdot)$ *preserves the IGFR property, i.e.* $\frac{qf(q)\dot{w}(\bar{F}(q))}{w(\bar{F}(q))}$ *increases in* q*, and* $w(p) + w(1-p)$ *is convex on* $p \in [0,1]$*. Then, for a risk-neutral subject there exists a* $\bar{\lambda} \geq 1$ *such that* $\forall \lambda \leq \bar{\lambda}$*,*

(i) the pull-to-center effect is minimized at $\frac{c}{r} = z_m$*, i.e.,* $\hat{q}^* = q^* = \bar{F}^{-1}(z_m)$*, for some* $z_m \in (0, p_0)$*;*

(ii) there is overorder ($q^* < \hat{q}^*$) *in the low profit scenario* ($c/r > z_m$) *and underorder* ($\hat{q}^* < q^*$) *in the high profit scenario* ($c/r < z_m$)*;*

(iii) the pull-to-center effect is evident when $\frac{c}{r} \leq \frac{\bar{F}^{-1}(\bar{z})}{\bar{F}^{-1}(z_m)}$*, i.e.,* $q^* < \hat{q}^* < \bar{F}^{-1}(z_m)$ *in the low profit scenario* ($c/r > z_m$) *and* $\bar{F}^{-1}(z_m) < \hat{q}^* < q^*$ *in the high profit scenario* ($c/r < z_m$)*, where* $\bar{z} = \arg\max z\bar{F}^{-1}(z)$*.*

We note that that the assumptions on demand and weighting functions are very mild. IGFR has been commonly adopted in operations literature (Lariviere 2006). Together with symmetric distribution, it covers all the demand distributions that have been used in newsvendor laboratories thus far (e.g., uniform, normal). It can be verified that all the all the candidate weighting functions discussed in §2, e.g., Tversky and Kahneman (1992), Prelec (1998) and Goldstein and Einhorn (1987), preserves the IGFR property as well as the convexity assumption. Specifically, The convexity assumption also leads to $w(p) + w(1-p) \leq 1$, which was referred to as "supercertainty" in Kahneman and Tversky (1979).

The result shares similar implications with Proposition 1, but applies to a more general setting with both positive and negative prospects. It characterizes a threshold loss aversion level $\bar{\lambda}$, under which there is under/over order in the high/low profit scenarios respectively. Specifically, the high vs low profit scenario is distinguished by a benchmark profitability level z_m characterized in part (i), where the behavioral-theory gap is minimized.

In addition to the under/over order behavior, part (iii) also identifies the "center" that the subject will anchor to when the profitability level is not too low. We wish to note that the condition $\frac{c}{r} \le \frac{\bar{F}^{-1}(\bar{z})}{\bar{F}^{-1}(z_m)}$ is more sufficient than necessary, and that the threshold frequently goes over 1, allowing the PTC statement to be valid for any $c/r \in [0,1]$. When the threshold $\frac{\bar{F}^{-1}(\bar{z})}{\bar{F}^{-1}(z_m)}$ is less than 1, part (iii) suggests that PTC is more likely to sustain under high profit scenarios. As will be shown in the following proposition, this is consistent with our findings for settings with strong loss aversion.

Proposition 4. **(Pull-To-Center Reduced under Strong Loss Aversion)**

(i) A risk-neutral, strongly loss-aversion subject orders at $\hat{q}^*|_{\lambda\to\infty} = \min\{\frac{r}{c}\underline{d}, \bar{F}^{-1}[w^{-1}(\frac{c}{r})]\}$.

(ii) The pull-to-center effect only sustains in high profit scenarios, i.e., $c/r \le \underline{z}$ *for some* $\underline{z} \in [0,1]$.

Without additional assumption on the demand or weighting function, Proposition 4 predicts that the PTC effect may or may not be present when subjects are extremely loss averse. Significant disutility upon loss severely punishes any order beyond $r\underline{d}/c$ and will henceforth result in a rather conservative ordering behavior. Unless the unit profitability is high (i.e., $c/r \le \underline{z}$), in which case the order will be similar to that in the case of all prospects being positive, as characterized in Proposition 1, the subject will basically stay within the safe zone that guarantees a gain by anchoring its order proportionally to the minimum demand ($\hat{q}^* = r\underline{d}/c$).

Even though the analysis for more general demand and risk attitude can be intractable, numerical simulations suggest the same anchoring effect persists among extensive circumstances. Consider the settings used in Bolton and Katok (2008), where $c = 9, r = 12, \underline{d} = 50, \bar{d} = 150$ for the low-profit, low-safety-stock scenario, and $c = 3, r = 12, \underline{d} = 0, \bar{d} = 100$ for the high-profit, high-safety-stock scenario, such that $q^* = 75$ for both scenarios. Applying Lemma 2 with the family of weighting functions and power utility functions used in Tversky and Kahneman (1992):

$$u(x) = \begin{cases} x^{\alpha} & \text{when } x \ge 0 \\ -\lambda(-x)^{\beta} & \text{when } x < 0, \end{cases}$$

where $0 < \alpha \le 1$, $0 < \beta \le 1$ and $\lambda \ge 1$, we find the same over- vs. under-ordering in the low- vs. high-profit scenario respectively. For example, at $\alpha = 0.8, \beta = 0.6, \lambda = 1.2, \gamma = 2, \delta = 1$, there is $\hat{q}^* = 85.4 > 75 = q^*$ in the low-profit scenario and $\hat{q}^* = 61.4 < 75 = q^*$ in the high-profit scenario. The same pull-to-center effect also holds when loss aversion is increased to $\lambda = 3$, whereas $\hat{q}^* = 83.8 > 75 = q^*$ in the low-profit scenario and $\hat{q}^* = 60.6 < 75 = q^*$ in the high-profit scenario. Similar observations hold true with the generic linear weighting functions as well; the results are omitted here due to the similarity in structure. Calibrating the exact or approximate value for these parameters calls for a more extensive experimental design and implies meaningful future research directions.

3.2 Newsvendor under Inverse-S Shaped Weighting Functions

We now investigate whether prospect theory with the inverse-S shaped weighting function can explain the results obtained in laboratory experiments of the newsvendor problem. Recall that by applying a specific form of utility function and weighting function, Nagarajan and Shechter (2014) had shown that prospect theory cannot explain the PTC effect when all prospects are positive. Our model confirms this view by systematically analyzing settings with both positive and mixed prospects under general functional forms, and offers additional insights for each specific setting. All technical details can be found in Appendix B.

In general, subjects will frequently overorder in high profit scenarios and underorder in low profit scenarios. This is opposite to the PTC effect. In fact, risk neutral subjects universally exhibit such ordering behavior under strictly positive prospects, or mixed prospects with low loss aversion. When subjects are risk averse with strictly positive prospects, they will always underorder in low profit scenarios, but may overorder or underorder in high profit scenarios. Thus PTC is consistently absent in low profit scenario and may only be plausible in the high profit scenario. For risk aversion subjects under mixed prospects, the opposite of PTC effect is dominant in numerical experiments.

Even though prospect theory with inverse-S-shaped weighting functions systematically fails to support the PTC effect, one cannot rule out its applicability on individual subjects. This pertains to the fact that PTC is more of a group effect out of aggregated statistics, whereas individual decisions are subjects to much more heterogeneity and may possess no PTC or even “pull-to-extreme” effect (Bolton and Katok 2008, Lau et al. 2014). In general, allowing the decision weight to take both the S- and inverse-S shapes empowers the prospect theory framework to support the laboratory newsvendor observations on both the aggregate and individual levels.

3.3 Further Notes

We conclude this section by noting that broadening the weighting functions to include the S shaped weights aligns well with the observation that routine managerial decisions are less likely to be driven by the plausibility of rare events. It also fits with the idea mentioned earlier that business decisions are fundamentally different than how subjects view lotteries presented to them without context. Indeed, managers have been trained to “focus on the big picture” than “micromanagement.” This mindset is also common among many risk management tools (e.g., dual sourcing, quick response, business insurance, contingency plans) which entail preparing for more probable events proactively and reacting to rare events in a responsive manner. Our model resonates with this logic.

4. Empirical Validation

In this section, we describe a preliminary set of lab experiments (and their results) that provide initial but reasonable evidence to some of the main assumptions that drive the theoretical results. As mentioned in §2, the shape of weighting functions is impacted by whether the settings behind experiments are *described* rather than *experienced*. In particular, *described* settings correspond to the inverse-S shaped weighting function, and the *experienced* setting corresponds to the S shaped function. Also, we keep in mind that the *described* setting is hard to achieved beyond pure lotteries as the weighting function can be rather sensitive to information presentation especially for context-heavy problems like newsvendor. Hence instead of attempting to frame a newsvendor problem into a pure *described* setting, we test if subjects exhibit differences in their order quantities when faced with *more description* versus *more experience* paradigm. We then reconcile this with our theoretical results which show that as one moves to a *more experienced* setting, the PTC effect is stronger.

We report experiments and results from a subject pool of 191 students in the university of one of the co-authors. All students were business or economics majors with a basic knowledge of probability and calculus, i.e., with more than adequate knowledge to be able to do calculations involved with a newsvendor problem. The subjects were divided into 6 groups going through 4 simuli, as summarized in Table 2. As will be elaborated in what follows, Simuli 1 and 3 are newsvendor experiments with different (*description* or *experience*) treatments; Simuli 2 and 4 use pure lotteries to elicit weighting function parameters from the participants. All the simuli are provided in Appendix C.

4.1 Description versus Experience

The first set of experiments (Simuli 1) that each group received is a newsvendor problem. Groups 1 and 2 received a low profit ("LP") scenario and Group 3 and 4 received a high profit ("HP") scenario, all under a discrete uniform demand $D \sim U\{20, 40, 60, 80, 100\}$ (thereafter referred to as the "5-point demand"). The newsvendor optimum q^* is 40 or 80 under LP or HP, respectively. On the other hand, Group 5 and 6 received the continuous uniform demand used in Bolton and Katok (2008), i.e., $U(50, 150)$ for LP and $U(0, 100)$ for HP (thereafter referred to as the "BK08 demand") where $q^* = 75$ for both scenarios. Notice that Group 1 and 2 were exactly the same, the only difference being that Group 2 had a detailed table that describes potential profits and

Table 2 Experiment Episodes

Group	Simuli 1		Simuli 2	Simuli 3	Simuli 4
	Baseline newsvendor experiment	Profit table	Elicit weighting function parameters	"Experience" Simuli 1	"Experience" Simuli 2
1	5-Point Demand/LP	✗	✓	✓	✓
2		✓	✓	✗	✗
3	5-Point Demand/HP	✗	✓	✓	✓
4		✓	✓	✗	✗
5	BK08 Demand/LP	✗	✓	✓	✓
6	BK08 Demand/HP	✗	✓	✓	✓

related probabilities for each order decision. The same applies to Group 3 and 4 with Group 4 being awarded the profit table. This between-subjects treatment creates a *more described* environment for Group 2 and 4 compared to Group 1 and 3, following which we classify the latter as the *baseline* setting and the former as the *more description* setting. Notice that the descriptive table would be rather cumbersome for BK08 demand, the provision of such was thus omitted for Group 5 and 6.

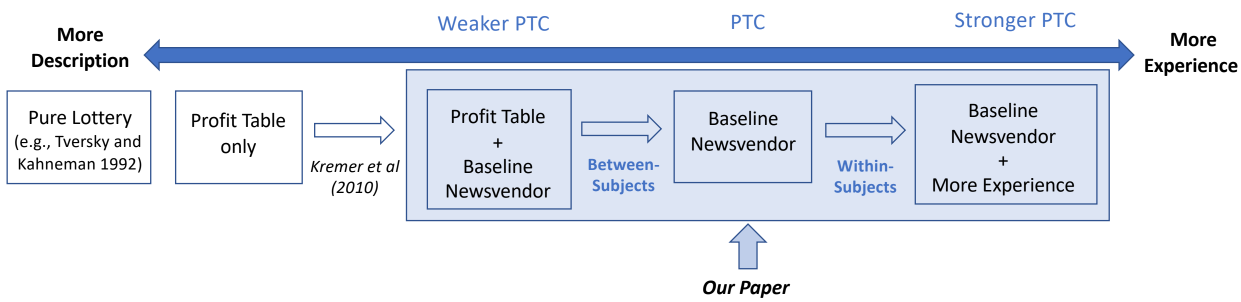
In a separate set of experiments (Simuli 3), all students except for those in Group 2 and 4 were allowed to have a greater degree of experience with the same exact problem. This involved numerous simulations of various possible scenarios, e.g., practicing answers to questions such as "What is the profit when an order was placed for Q units," demonstrations of probabilities using a virtual dice, etc. Much of these elements was unstructured and left to the participants and an assistant that coordinated the experiments to explore. This set of experiments constitutes Simuli 3, which in essence captures the same situation in Simuli 1 but adds *more experience* within-subjects.

The idea of the above newsvendor experiment design is illustrated in Figure 2. According to our theory, possessing a profit table increases the level of *description*, therefore should result in a lower degree of PTC effect compared to the *baseline* experiment. On the other hand, allowing *more experience* should lead to a higher degree of PTC. For instance, a *more description* group in Kremer et al. (2010) dealing with profit table only had exhibited less PTC effect than the other group who also had to absorb the newsvendor storyline (hence *more experience*).

4.2 Elicit Parameters

In a parallel set of experiments (Simuli 2 and 4), we elicit parameters for weighting functions as the paradigm shifts. The weighting function we are interested in estimation is one of those mentioned in §2, i.e. the two-parameter function used in Gonzalez and Wu (1999): $w(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma}$, where δ

Figure 2 **Summary of Newsvendor Experiments and Findings**



reflects the elevation of the function and γ the curvature. When δ is relatively small, higher values of γ, i.e., $\gamma > 1$, correspond to an S-shaped weighting function. To calibrate the weighting function, we assume that the utility function takes the familiar exponential form, i.e., $u(x) = x^{\sigma}$. Note that although we estimate the value of σ, δ and γ, our main interest is the value of γ.

The estimation procedure that we use is inspired by Tanaka et al. (2010) and followed by many other studies such as Toubia et al. (2013) and Gopalakrishnan et al. (2020). The main idea is that by carefully designing three series of lotteris, one can calculate the three CPT parameters (σ, δ, γ) from the three switching points subsequently. Specifically, in Series 1 lotteries, subjects are given a choice between two outcomes and they have to choose one of them (left table in Figure C3 in Appendix C). By the design of the lotteries, we look for the first switching point of the subject. This corresponds to a unique set of $(\sigma,\ \delta)$, which can be looked up from the right table in Figure C3. Series 2 lotteries are similar and we record once again where the respondents switch in this lottery set. The design of these two series of lotteries allows us to solve simultaneously for feasible values of σ and δ. Once σ and δ are obtained, we can elicit the value of γ from Series 3 lotteries.

These three lottery series are the main features of Simuli 2 and 4. In Simuli 2, participants were given the three lottery series when doing Simuli 1. In Simuli 4, participants were given only the third lottery series when doing Simuli 3, along with the opportunity to simulate the outcome of the lotteries, thus adding *experience* compared to the same lotteries in Simuli 2. Notice that to test if *experience* affects the value of γ, it suffices to test if by *experiencing*, the participants switch later in Simuli 4 as compared to Simuli 2. For the sake of tractability, we only provide Series 3 lotteries in Simuli 4. That is, we assume that *experiencing* does not alter the value of σ and δ.

4.3 Results

Among the 191 participants, we rule out those who had provided incomplete responses (e.g., those who had been part of Simuli 1 but not Simuli 3) or submitted erratic answers (e.g., "20 or 40"). Our discussion of the results is then based on responses from the remaining 174 respondents.

We first look into the within-subjects effect of *experience* on individual order quantities by comparing the *baseline* and the *more experience* settings (i.e. Simuli 1 and 3 for Groups 1, 3, 5 and 6). Since we only run one round of experiment for each participant in a simuli, we aggregate the order quantities in each group and report the difference between the average order quantity $\bar{\bar{q}}$ and the newsvendor optimum q^* in Table 3.

Table 3 **Average behavioral-theory gap ($\bar{\bar{q}} - q^*$) for different groups**

	5-Point Demand		BK08 Demand	
	LP	HP	LP	HP
Simuli 1 (Baseline)	18.03	−15.07	14.26	−13.59
Simuli 3 (More Experience)	23.17	−23.50	22.45	−16.44
t-statistics[i]	1.83†	−1.86†	5.27**	−2.45*

[i] The t-test compares the difference in individual order quantities between Simuli 3 and Simuli 1.
† p-value $\approx 7\%$
* $0.01\% \leq p$-value $< 5\%$
** p-value $< 0.01\%$

As we can see, PTC effect is universal across the simuli – the LP's over-ordered on average and the HP's under-ordered on average. The impact of *experience* is highlighted by the change between the first two rows. By merely comparing the averages, the order quantities are more distorted in Simuli 3 than Simuli 1. We then verify the differences between these two simuli using paired-sample t-tests. The results are reported in the last line of Table 3. We first note that under BK08 demand, individual orders vary significantly from Simuli 1 to Simuli 3 ($p < 0.01\%$ for LP and $p < 5\%$ for HP). This represents a higher degree of PTC and confirms the influence of *experience* as we have predicted. For the 5-point demand, similar patten holds albeit at a marginal significance level (p around 7%). Considering that the 5-point demand is much more coarse than the continuous BK08 demand, it is natural that any incremental PTC effect can be less obvious. We are therefore in the opinion that the 7% significance level under 5-point demand supports the prediction that *experience* leads to a stronger PTC effect.

We next discuss the between-subjects effect of *experience* on the same behavioral-theory gaps. Recall that Group 2 and 4 received the *more description* treatment and Group 1 and 3 were with the *baseline* and *more experience* paradigms in LP and HP, respectively. The summarization data in Table 4 suggests that in both scenarios, the average order quantities are closer to the newsvendor optimum (i.e., less PTC) as the the paradigm carries less *experience* (or more *description*).

To highlight this effect, we compare the order quantities between the *more description* vs. *more experience* paradigms (i.e., Group 1 vs 2 for LP, and Group 3 vs 4 for HP). Results from the two-sample t-tests are reported in the last column of Table 4. The statistics imply that orders in the same scenario vary significantly between these two paradigms ($p \leq 0.1\%$ for LP and $p \leq 5\%$ for HP) in the same way as predicted by our model.

Table 4 Average behavioral-theory gap ($\bar{\bar{q}} - q^*$) in different paradigms

	More Description	*Baseline*	*More Experience*	t-statistics[ii]
LP	8.67	18.03	23.17	3.36^{**}
HP	-14.22	-15.07	-23.50	-1.97^{*}

[ii] The t-test compares the order quantities between *More Description* vs *More Experience* paradigms under 5-point demand (Group 1 vs 2 for LP and Group 3 vs 4 for HP).
* $0.1\% < p$-value $\leq 5\%$
** p-value $\leq 0.1\%$

We need to point out that if the two-sample t-tests are instead performed between the *more description* vs *baseline* paradigms, then only the difference related to LP is at 5% significance level. This less than obvious effect under HP could again be due to the coarse nature of 5-point demand. The asymmetry between LP and HP may also be linked to similar observations in Schweitzer and Cachon (2000) and Bostian et al. (2008), where they find a stronger PTC effect under LP.

In the set of newsvendor experiments that have been conducted, we consistently observe that as the setting moves from *more description* towards *more experience*, there is a tendency that the PTC effect is reinforced. This provides some support to the view that *experiencing* may explain what is observed in several recorded lab experiments in the literature. Along the same time, we also see a notable increase in γ when moving from Simuli 2 to Simuli 4. For instance, for Group 1 the average γ value in Simuli 4 was 0.89. Indeed, it is less than 1 (recall that we are still in a relatively *descriptive*, lottery environment) but greater than 0.69 for Simuli 2. Those for Group 3, 5 and 6 have similar shifts. Assuming that σ and δ remain unchanged, this implies the same trend of movement from inverse-S to S shaped weighting functions. This preliminary finding may offer a mechanism that partially explains some of the observations in the literature.

4.4 Limitations

Our experiments are only a preliminary investigation to test if some of the machinery behind our theoretical results have behavioural underpinnings. Our findings thus far indicate the plausibility of our claims. Nevertheless, we wish to point out several limitations of our experimental approach.

The first is the subject pool. It would be wise to replicate this set of experiments over a larger sample size. This will allow one to validate more of these insights through between-subject experiments. It will also be interesting to run the experiments across subjects in different cultural backgrounds (e.g., Cui et al. 2013) to gain a robust view of the impact of *experience*.

The second is the design. Since newsvendor experiments are heavily context dependent and involve substantially more prospects and higher degree of randomness than pure lotteries, it is hard to frame the experiments as pure *descriptive*. The fact that we were not able to compensate the subjects with performance-based incentives also adds to its relevancy in the *experience* paradigm. Therefore, what we have been focusing on studying is the marginal effect of the paradigms. And this is done by granting (or controlling) a subject with more or less *experience* element. Future experiments may approach reduced-form newsvendor experiments for a better alignment with pure lotteries and include proper incentive schemes to further validate these effects.

In addition, to preserve parsimony and experimental elegance, we made several assumptions during eliciting parameters. For instance, we exploited the fact that at $p = 0.5$, the weighting function nicely breaks down, which lets us design easy lotteries. This comes with a caveat in that a 50% chance lottery may result in unreliable outputs. Further, we assumed that the values of δ and σ do not change with more *experience*, which calls for more rigorous test.

Despite these limitations, we believe that these initial results are intriguing and work as a valuable add-on to the theory that supports our model foundations. It also offers the opportunity to test new observations using the *experience* versus *described* paradigms.

5. Robustness

As an endnote, we discuss the robustness of our model – how well it can accommodate other reference points and produce reasonable benchmark predictions that align with laboratory observations.

5.1 Alternative Reference Points

Even though our main analysis applies *status quo* as the reference point, it can be verified that the same insights hold for some other reference points. We show case of such through another commonly used reference point identified in Baillon et al. (2020) – a security level that represents the maximum of the minimal outcomes of the prospects; namely *MaxMin*.

In the newsvendor model considered in §3, the minimal outcome of any prospect (order quantity $q \in [\underline{d}, \bar{d}]$) is given by $r\underline{d} - cq$. Therefore, the maximum among them is achieved at $q^{MaxMin} = \arg\max_q (r\underline{d} - cq) = \underline{d}$. The MaxMin reference point in the newsvendor problem is then:

$$R^{MaxMin} = (r - c)\underline{d}.$$

The utility considered in (6), accounting the *MaxMin* reference point, can then be rewritten as:

$$\begin{aligned}V(q) &= \int_0^{cq/r} u(rx - cq - R^{MaxMin})w^{'}(F(x))dx + \int_{cq/r}^{q} u(rx - cq - R^{MaxMin})[-w^{'}(\bar{F}(x))]dx \\ &\quad + u(rq - cq - R^{MaxMin})w(\bar{F}(q)) \\ &= \int_{\hat{x}\le\frac{c}{r}\hat{q}} u(r\hat{x} - c\hat{q})w^{'}(F(x))dx + \int_{\frac{c}{r}\hat{q}<\hat{x}\le\hat{q}} u(r\hat{x} - c\hat{q})[-w^{'}(\bar{F}(x))]dx + u(r\hat{q} - c\hat{q})w(\bar{F}(q)),\end{aligned}$$

where $\hat{x} = x - \underline{d}$ and $\hat{q} = q - \underline{d}$.

As can be seen, the newsvendor essentially needs to deal with the same problem as when the reference point is *status quo* but demand following the same shape of distribution within $[0, \bar{d} - \underline{d}]$. Following the same argument, it is straightforward to verify that all results in §3 hold for any static reference point within $[0, R^{MaxMin}]$.

Recall that Uppari and Hasija (2019) also assess *MaxMin* (referred to as "sure-shot profit" in their paper) capable to predict PTC. On this matter alone, our finding is fairly consistent with literature. However, the support of *MaxMin* rather than *status quo* (Schweitzer and Cachon 2000) is not consistent as the two will merge when the minimum demand turns zero. To certain extent, our model reconciles this conflict through the consideration of decision weights and support for any static reference point within an open space between the *status quo* and *MaxMin* inclusively.

5.2 Benchmark Predictions

In what follows, we discuss some benchmark implications of our model and comment on its alignment with empirical evidence.

5.2.1 Extreme critical fractiles. As the critical fractile approaches to 0 or 1, it follows from Lemma 2 that $\hat{q}^*_{\frac{c}{r}=1} = \underline{d}$ and $\hat{q}^*_{\frac{c}{r}=0} = \bar{d}$. Thus our model produces the reasonable prediction that a subject will order at the floor (ceiling) demand when the margin is rather low (high), which also conforms to expected profit maximization. In other words, PTC effect is minimized when $\frac{c}{r}$ is 0 or 1. This is supported by empirical evidence from Ockenfels and Selten (2014) and employed as a critical criteria in evaluating candidate models in Uppari and Hasija (2019).

These properties, albeit naive, do not necessarily hold in all models that predict PTC effect. Consider the overconfidence model in Ren and Crosen (2013) for instance, which suggest that the subjects make their decisions based upon a transformation of the true demand, i.e., $D_O = \phi D + (1 - \phi)F(\mu, 0)$ where $D(\mu, \sigma)$ is the true demand and $F(\mu, 0)$ is a zero variance distribution. For Bostian et al. (2008) with demand distribution $U[0, 100]$, the overconfidence model estimates the ϕ as 0.64 for the high profit scenario and 0.44 for the low profit scenario. Therefore, the transformed demand is $U[18, 82]$ for the high profit scenario and $U[28, 72]$ for the low profit scenario.

According to Proposition 1 in Ren and Crosen (2013), the subjects will then order at 28 or 82 when the critical fractile is 0 or 1, respectively, suggesting a counterintuitive (also against empirical evidence) conclusion that the PTC effect persists when the margin is extremely low or high. Indeed as Ren et al. (2017) comment, one should "expect overprecise newsvendors to make particularly large errors in both highly profitable markets and very low-profit markets."

5.2.2 Moderate critical fractiles. Proposition 1-3 in our main analysis suggest that there exists a unique, moderate critical-fractile level at which PTC will be minimized, and above (below) which there will be under- (over-) order. In connection with the prediction for 0 or 1 critical fractiles, it resonates some other behavioural patterns found in laboratory newsvendor, notably the *convex-concave pattern* for the average order quantities found in Ockenfels and Selten (2014) and the *M-shaped performance gap* depicted in Uppari and Hasija (2019). Both patterns suggest that the behavioral decision $\hat{q}$ and the newsvendor optimum q^* coincide at the two ends of the critical fractiles as well as in the middle, and further apart elsewhere. To the best of our knowledge, only the *mean-demand stochastic reference point* model ("M5" in Uppari and Hasija 2019) and our model support these patterns.

5.2.3 Risk/Loss neutral. Following Proposition 3, our model continues to support PTC even when the subjects are risk neutral ($u'' = 0$) and loss neutral ($\lambda = 1$). This support is somehow missing in other prospect theory based models that do not consider decision weights, e.g., M2-M5 in Uppari and Hasija (2019) predict that a risk/loss neutral subject will order at the newsvendor optimum. Indeed, if a subject faithfully weights each potential outcome with respect to their true probabilities without risk or loss aversion, there is no reason why it would not order at a quantity that maximizes expected profit. However, thus far there has been no empirical evidence supporting this (i.e., the PTC effect is absent within risk neutral and loss neutral subjects). For one thing, the rare existence of risk and loss neutral subjects could make such hard to validate. More importantly, the measurement of risk and loss aversion is based on a reference point and existing parametric estimates typically employ *status quo* as the reference point (e.g., Abdellaoui et al. 2007). The literature is not loud on whether (or, to what extent) the estimates may vary under a different reference point. This makes the risk/loss neutral cornercase even more challenging to validate for models based on, say, stochastic reference point. It is therefore in our opinion that a model framework more inclusive on behavioural outcomes in this particular case is more preferable.

5.2.4 Strong loss aversion. Our model suggests that PTC may disappear for subjects possessing strong loss aversion in low profit scenarios (Proposition 4). This prediction is intuitive; otherwise, it would be hard to argue why some one severely suffering from loss would still order towards mean even though such does not warrant a sure gain. Not all models in literature come with the same prediction though. For instance, applying model M5 in Uppari and Hasija (2019) for strongly loss averse newsvendors, it "... predicts μ (mean demand) as the optimal order quantity." Perhaps this discrepancy is forgivable, as laboratory subjects are unlikely to be extremely loss averse (e.g., Tanaka et al. 2010 find the loss aversion λ typically around 2.5 and rarely exceed 3) and that the notion of loss aversion is tightly attached to the reference point itself. However, since existing estimates of loss aversion are primarily drawn from *status quo* and the superiority of M5 (comparing to other models in Uppari and Hasija 2019) was based on a wide range of loss aversion parameter ($\lambda \in [0, 50]$), this distinction may nevertheless be considered when evaluating the openness and robustness of a model.

6. Concluding Remarks

In conclusion, our paper shows that the general framework of prospect theory can explain the phenomena observed in behavioral lab studies of the classical newsvendor problem, the only stipulation being a generalized set of probability weighting functions. Our theoretical analysis is possibly the most open one in this literature. First, it holds under an inclusive and open PT framework without any special assumption on *reference-point*. This model has both theoretical and empirical foundation, and the parameters can be configured and estimated across a wide range of behavioral settings. Second, it systematically predicts, with very few assumptions, the PTC effect that has long been observed in the laboratory settings of the newsvendor problem at the aggregate level. At the same time, it is also capable to interpret any none-PTC behavior on the individual level and a number of other phenomena observed in laboratory studies of the newsvendor problem. Lastly, the model is robust under a number of generic reference points, and produces more reasonable benchmark predictions than other candidate models in literature.

The idea of exploring the impact of weighting functions on operational decisions and how they depend on the decision paradigm is new to operations literature. We provide an experimental framework to correlate the paradigm of operational decisions with the shape of the weighting functions, and confirm its impact on operational decisions. Combining the theory and empirical evidence, our findings shed light on how risk and loss aversion interacts with weighting functions in influencing decisions.

Looking forward, our results raise several important questions that can be explored in future research. The central ones is how decision makers should be directed to comprehend the problem and its effect on the outcomes. Indeed, our analysis suggests that if decision makers ponder more on the contextual problem or experience more with the same, then PTC effects are more significant. This can be tested by broadening the experiments we have done to calibrate parameter values for the decision weights and reconciling them with decisions under risk. In the long run, this can help us understand how to enhance the decision support systems by factoring into inherent attitudes or traits of the decision makers, and how real life decisions should be framed to managers and how their experience is to be manipulated. These may eventually move towards the bigger question as to how humans and decision automation should make superior operational decisions together (Donohue et al. 2020).

On the technical level, comparison between this and other models in literature, especially under benchmark scenarios such as risk/loss neutrality and strong loss aversion, or their fitness to newsvendor-based experiments such as strategic customers (Zhang et al. 2019), would fuel meaningful future empirical studies. On the managerial level, both academics and practitioners are encouraged to review competitive strategies that rely upon human decisions to account for large numbers of uncertain future scenarios.

Acknowledgments

This research is partially supported by the Natural Sciences and Engineering Research Council of Canada [NSERC RGPIN-2011-402324] and the Concordia University Research Chair program.

References

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

Abdellaoui, M., O. L'Haridon, C. Paraschiv. 2011. Experienced vs. described uncertainty: Do we need two prospect theory specifications? *Management Science* **57**(10) 1879–1895.

Arrow, K. J. 1965. The theory of risk aversion. *Aspects of the Theory of Risk Bearing.* Yrjö Jahnssonin Säätiö.

Baillon, A., H. Bleichrodt, V. Spinu. 2020. Searching for the reference point. *Management Science* **66**(1) 93–112.

Baron, O., M. Hu, S. Najafi-Asadolahi, Q. Qian. 2015. Newsvendor selling to loss-averse consumers with stochastic reference points. *Manufacturing & Service Operations Management* **17**(4) 456469.

Barron, G., G. Ursino. 2013. Underweighting rare events in experience based decisions: Beyond sample error. *Journal of Economic Psychology* **39** 278–286.

Barron, G., E. Yechiam. 2009. The coexistence of overestimation and underweighting of rare events and the contingent recency effect. *Judgment and Decision Making* **4**(6) 447–460.

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 Operations Management* **10**(3) 519–538.

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

Camilleri, A. R., B. Newell. 2011. Description- and experience-based choice: Does equivalent information equal equivalent choice? *Acta Psychologica* **136**(3) 1276–1284.

Camilleri, A. R., B. Newell. 2013. The long and short of it: Closing the description-experience gap by taking the long-run view. *Cognition* **126** 54–71.

Chaudhry, S. J., M. Hand, H. Kunreuther. 2018. Extending the time horizon: Elevating concern for rare events by communicating losses over a longer period of time. *Working paper, The Wharton School University of Pennsylvania* 1–66.

Crawford, V., J. Meng. 2011. New york city cab drivers labor supply revisited: Reference-dependent preferences with rational ex- pectations targets for hours and income. *American Economic Review* **101**(5) 19121932.

Cui, Y., L. G. Chen, S. Gavirneni, Q. Wang. 2013. Chinese perspective on newsvendor bias: An exploratory note. *Journal of Operations Management* **31** 93–97.

Dai, Y., T. Pachur, T. J. Pleskac, R. Hertwig. 2019. What the future holds and when: A descriptionexperience gap in intertemporal choice. *Psychological Science* **30**(8) 12181233.

Danziger, S., L. Hadar, V. G. Morwitz. 2014. Retailer pricing strategy and consumer choice under price uncertainty. *Journal of Consumer Research* **41** 761–774.

De Giorgi, E. G., T. Post. 2011. Loss aversion with a state-dependent reference point. *Management Science* **57**(6) 1094–1110.

de Puy Kamp, M. 2020. Federal officials repeatedly warned that US hospitals lacked enough ventilators. CNN, March 27, 2020.

de Véricourt, F., K. Jain, J. N. Bearden, A. Filipowiczb. 2013. Sex, risk and the newsvendor. *Journal of Operations Management* **31** 86–92.

DeCambre, M. 2020. 'Black Swan' author says that if investors don't use a 'tail hedge,' he recommends 'not being in the market': 'We're facing a huge amount of uncertainty'. MarketWatch, June 27, 2020.

Donohue, K., Ö. Özer, Y. Zheng. 2020. Behavioral operations: Past, present, and future. *Manufacturing & Service Operations Management* **22**(1) 191–202.

Durbach, I., G. Montibeller. 2018. Predicting in shock: on the impact of negative, extreme, rare, and short lived events on judgmental forecasts. *EURO Journal on Decision Processes* **6**(1-2) 213–233.

Frey, R., R. Mata, R. Hertwig. 2015. The role of cognitive abilities in decisions from experience: Age differences emerge as a function of choice set size. *Cognition* **142** 60–80.

Gaertig, C., J. P. Simmons. 2018. Why (and when) are uncertain price promotions more effective than equivalent sure discounts? *Working paper, The Wharton School University of Pennsylvania* 1–45.

Goldstein, W. M., H. J. Einhorn. 1987. Expression theory and the preference reversal phenomena. *Psychological Review* **94** 236254.

Gonzalez, R., G. Wu. 1999. On the shape of the probability weighting function. *Cognitive Psychology* **38** 129–166.

Gopalakrishnan, S., M. Nagarajan, A. Shields, D. Griffin. 2020. Measuring individual differences in risk attitude with prospect theory. *Working Paper. University of British Columbia* .

Hertwig, R., G. Barron, E. U. Weber, I. Erev. 2004. Decisions from experience and the effect of rare events in risky choice. *Psychological Science* **15**(8) 534–539.

Hertwig, R., I. Erev. 2009. The description-experience gap in risky choice. *Trends in Cognitive Sciences* **13**(12) 517–523.

Hills, T. T., R. Hertwig. 2010. Information search and decisions from experience: do our patterns of sampling foreshadow our decisions? *Psychological Science* **21**(12) 1787–1792.

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

Köszegi, B., M. Rabin. 2006. A model of reference-dependent preferences. *The Quarterly Journal of Economics* **121**(4) 1133–1166.

Köszegi, B., M. Rabin. 2006. Reference-dependent risk attitudes. *American Economic Review* **97**(4) 1047–1073.

Kremer, M., S. Minner, L. N. VanWassenhove. 2010. Do random errors explain newsvendor behavior? *Manufacturing & Service Operations Management* **12**(4) 673–681.

Lariviere, M. A. 2006. A note on probability distributions with increasing generalized failure rates. *Operations Research* **54**(3) 602–604.

Lau, N., S. Hasija, J. N. Bearden. 2014. Newsvendor pull-to-center reconsidered. *Decision Supports Systems* **58** 68–73.

Lejarraga, T., J. Müller-Trede. 2017. When experience meets description: How dyads integrate experiential and descriptive information in risky decisions. *Management Science* **63**(6) 1953–1971.

Lejarraga, T., T. Pachur, R. Frey, R. Hertwig. 2016. Decisions from experience: From monetary to medical gambles. *Journal of Behavioral Decision Making* **29** 67–77.

Lien, J. W., J. Zheng. 2015. Deciding when to quit: Reference-dependence over slot machine outcomes. *American Economic Review* **105**(5) 366370.

Long, X., J. Nasiry. 2015. Prospect theory explains newsvendor behavior: The role of reference points. *Management Science* **61**(12) 3009–3012.

Nagarajan, M., S. Shechter. 2014. Prospect theory and the newsvendor problem. *Management Science* **60**(4) 1057–1062.

Ockenfels, A., R. Selten. 2014. Impulse balance in the newsvendor game. *Game and Economic Behavior* **86** 237–247.

Palma, A. d., M. Abdellaoui, G. Attanasi, M. Ben-Akiva, I. Erev, H. Fehr-Duda, D. Fok, C. R. Fox, R. Hertwig, N. Picard, P. P. Wakker, J. L. Walker, M. Weber. 2014. Beware of black swans: Taking stock of the description experience gap in decision under uncertainty. *Marketing Letters* **25** 269280.

Pratt, J. W. 1964. Risk aversion in the small and in the large. *Econometrica* **32**(1-2) 122–136.

Prelec, D. 1998. The probability weighting function. *Econometrica* **66**(3) 497–527.

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

Ren, Y., R. Crosen, R. T. Crosen. 2017. The overconfident newsvendor. *Journal of the Operational Research Society* **68** 496?506.

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

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

Su, X. 2008. Consumer returns policies and supply chain performance. *Manufacturing & Service Operations Management* **11**(4) 595–612.

Sugden, R. 2003. Reference-dependent subjective expected utility. *Journal of Economic Theory* **11** 172–191.

Tanaka, T., C. F. Camerer, Q. Nguyen. 2010. Risk and time preferences: Linking experimental and household survey data from vietnam. *American Economic Review* **100**(1) 557–571.

Toubia, O., E. Johnson, T. Evgeniou, P. Delquié. 2013. Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters. *Management Science* **59**(3) 613–640.

Tversky, A., D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5**(4) 297–323.

Tversky, A., P. Wakker. 1995. Risk attitudes and decision weights. *Econometrica* **63**(6) 1255–1280.

Ungemach, C., N. Chater, N. Stewart. 2009. Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science* **20**(4) 473–479.

Uppari, B. S., S. Hasija. 2019. Modeling newsvendor behavior: A prospect theory approach. *Manufacturing & Service Operations Management* **21**(3) 479–711.

van de Kuilen, G., P. P. Wakker. 2011. The midweight method to measure attitudes toward risk and ambiguity. *Management Science* **57**(3) 582–598.

Wulff, D. U., M. Mergenthaler-Canseco, R. Hertwig. 2018. A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin* **144**(2) 140–176.

Zhang, Y., B. Mantin, Y. Wu. 2019. Inventory decisions in the presence of strategic customers: Theory and behavioral evidence. *Production and Operations Management* **28**(2) 374392.