An Experimental Investigation of Pull Contracts in Supply Chains

Andrew M. Davis
The Samuel Curtis Johnson Graduate School of Management, Cornell University, 401J Sage Hall, Ithaca, New York 14853, USA
adavis@cornell.edu

In this study, I investigate supply chain contracts in a setting where a supplier uses its inventory to directly satisfy a retailer’s demand. These “pull” contracts have increased in popularity in practice but have not been studied experimentally. In a controlled laboratory setting, I evaluate a wholesale price contract and two coordinating contracts. The data suggest that the benefit of the two coordinating contracts over the wholesale price contract is less than the standard theory predicts, and that retailers, in the two coordinating contracts, exhibit a systematic bias of setting the coordinating parameter too low, and the wholesale price too high, relative to the normative benchmarks. In an effort to explain this deviation, I explore three behavioral models and find that loss aversion and reference dependence fit the data well. I empirically test this result in a follow-up experiment, which directly controls for loss aversion and reference dependence, and observe that retailers make significantly better decisions. Lastly, I administer a number of experiments which reduce the complexity of the problem, curtail the amount of risk, and increase the level of decision support, and find that none improve decisions relative to the treatment that controls for loss aversion and reference dependence.

Key words: behavioral operations management; pull contracts; supply chain management; loss aversion and reference dependence

History: Received: May 2012; Accepted: January 2014 by Karen Donohue, after 2 revisions.

1. Introduction

Today many suppliers are willing to carry inventory themselves and ship product directly to end customers when a retailer has realized demand. This fulfillment structure, which allows the retailer to avoid the cost of unsold inventory, has grown to a point that some companies even specialize in providing services for this setting. Consider CommerceHub. Their systems allow retailers to integrate with any one of 6000 suppliers, all prepared to ship product when a retailer has realized demand. CommerceHub’s customers include, among others, Walmart, General Electric, Walgreens, Costco, Staples, Dell, Toys R Us, and Kohls (CommerceHub 2013). As such, retailers must design and administer contracts for each one of the potentially thousands of suppliers with which they do business.

Contracts used in this supply chain setting, where the supplier holds the inventory and incurs the cost of any unsold product, are often referred to as “pull” contracts. Specialty items in stores are frequently ordered using pull contracts (Klein 2009), and e-commerce retailers utilize pull contracts extensively. In particular, Randall et al. (2006) report that between 23% and 33% of Internet retailers use pull contracts exclusively, where the US Census estimates that retailer e-commerce sales totaled $194 billion in 2011, up 16.4% compared to 2010 (US Census Bureau 2013). Pull contracts are not only common in retailer settings, but in industrial contexts as well, between manufacturers and suppliers. For example, just-in-time (JIT) fulfillment, where the supplier delivers in small batches, is a form of a pull structure, as the supplier effectively becomes responsible for holding and managing the inventory, as opposed to the manufacturer.

The prevalence of pull contracts in practice is for good reason: research suggests that a simple wholesale price pull contract generates more supply chain profit than a wholesale price push contract both theoretically (Cachon 2004) and experimentally (Davis et al. 2014). However, no study to date has investigated both a wholesale price contract and additional, coordinating pull contracts together, from an experimental standpoint. This study is a first attempt at addressing this. Specifically, I aim to address three research questions: (i) which pull contracts perform best for retailers in a controlled laboratory environment? (ii) what sort of behavioral models can explain retailers’ contract decisions, and thereby identify the bias driving results? and (iii) can controlling for the aforementioned bias improve decisions and ultimately profits?

As mentioned previously, pull contracts have yielded interesting theoretical results, even when focusing solely on wholesale price contracts. Cachon...
(2004) illustrates that a wholesale price pull contract results in higher supply chain profit compared to a wholesale price push contract. This improvement is primarily due to the structure of how demand is satisfied in the two contexts. Specifically, under a push contract, a retailer makes an order from a supplier in advance of demand. In this push setting, with full information, a supplier can determine exactly what the retailer will order, produce that amount, and avoid any demand uncertainty. However, under a pull setting, the retailer pulls product from the supplier as demand is realized. Therefore, in a pull context, both parties incur the risk associated with random demand. This unique risk sharing feature is reason to believe that established results in the existing behavioral supply chain literature may not necessarily extend to pull contracts, and therefore an impetus for conducting this study (see section 2 for a summary of relevant literature).

My initial experiment considers three pull contracts; a wholesale price contract and two coordinating contracts—a payback contract and a service-level agreement (SLA). A wholesale price contract states that a wholesale price be paid from the retailer to the supplier for each unit sold. A payback contract builds on the wholesale price contract by adding a per-unit payback amount that is paid from the retailer to the supplier for each unit that the supplier overstocks. The payback contract can perfectly coordinate the supply chain with an arbitrary split of profits. An SLA also includes a second parameter, a bonus, that is paid from the retailer to the supplier when the supplier satisfies a predetermined fraction of the retailer’s demand. A recent study shows that 91% of organizations utilize SLAs to manage suppliers and external customers (Oblicore 2007). Overall, this set of contracts includes one which is simple (wholesale price), one with ideal theoretical characteristics (payback), and one that is frequently observed in practice (SLA).

In each treatment of the experiment, the retailer begins by proposing contract terms to a supplier, such as a wholesale price per unit (followed by the supplier setting a stocking quantity). This most closely matches the theoretical literature on pull contracts (Cachon 2004), and also coincides with how pull contracts are often established in practice. In particular, under pull contracts, retailers act as intermediaries for suppliers to sell product. These intermediaries, such as Amazon, Etsy, and Clickbank, provide a fixed pricing menu, which discloses the commission fees that the retailers will charge the supplier, for each unit that the supplier sells through the retailer’s storefront (this may apply to brick-and-mortar intermediaries or sales representatives as well). When a product is sold, the retailer collects the selling price, keeps the commission fee, and pays the wholesale price to the supplier.

Therefore, in my experiment, a retailer proposing a wholesale price and contract terms to its supplier provides a direct test of existing models on pull contracts and also matches industry practice.

The first primary result from my initial experiment is that the payback contract and SLA outperform the wholesale price contract, but that their theoretical benefit does not translate into practice. In theory, the two coordinating contracts should result in a 41% increase in retailer profit over the wholesale price contract, but instead the payback contract increases profit by roughly 5%, and the SLA by 8%.

The second result is that retailers, in the coordinating contracts, systematically deviate and set the coordinating parameter (the payback amount or bonus) too low, and the wholesale price too high, relative to the normative benchmark. In an effort to formally explain this behavior, I explore three alternative models well known in the supply chain literature: risk aversion (Eeckhoudt et al. 1995), loss aversion (Ho and Zhang 2008, Tversky and Kahneman 1991), and reference dependence (Ho et al. 2010, Tversky and Kahneman 1991). I apply these models to the two coordinating pull contracts and structurally estimate their parameters. The results indicate that the loss aversion and reference dependence formulations fit well, specifically, that retailers suffer an extra psychological cost from observing the payment of the coordinating parameter, and how far the supplier’s stocking quantity is from realized demand.

To confirm that loss aversion and reference dependence are affecting decisions, I conduct an additional experiment that controls for both biases. In this treatment, I restrict information pertaining to whether the payback amount or bonus was paid out, and also reference dependence (Su 2008). However, unlike the treatment that controls for loss aversion and reference dependence, since retailers do not see whether the payback amount was paid out, and also reference dependence, since retailers do not see how far the supplier’s stocking quantity is from realized demand. Ultimately, retailers’ contract decisions greatly improve in this treatment, leading to substantially higher profits.

Lastly, I investigate a number of other competing interventions through additional experimental treatments. These treatments control for biases known to exist in other supply chain contexts, such as complexity (Kalkanci et al. 2011), variability and risk aversion (Engelbrecht-Wiggans and Katok 2009), and random errors (Su 2008). However, unlike the treatment that controls for loss aversion and reference dependence, in none of these interventions do retailer profits significantly change compared to the original baseline treatments.

In the next section, I summarize the relevant literature and how my study contributes to it. In section 3, I
Behavioral operations management has emerged as a stream of research that explores human decision making in various operational contexts. Some of the topics investigated in this area include the bullwhip effect (Croson and Donohue 2006), social preferences (Loch and Wu 2008), learning (Bostian et al. 2008), inventory management (Schweitzer and Cachon 2000), forecasting (Özer et al. 2011), and procurement (Wan et al. 2012). This study falls under the umbrella of behavioral supply chain contracting, specifically relating to those studies which include human-subject experiments and behavioral modeling. Within this stream, there are a few select works which are most closely related to one conducted here. Below, I summarize these papers and highlight how mine contributes to this existing literature.

Lim and Ho (2007) investigate one-, two-, and three-block tariffs between a manufacturer and retailer in a push setting. They find that, contrary to the standard theory, profits increase when moving from a two-block to three-block tariff. They proceed by developing a behavioral model for responders (retailers) and determine how the party initially proposing contract parameters (manufacturers) should take this behavior into account. In a similar study, Ho and Zhang (2008) consider how to frame a fixed fee in a two-part tariff contract, and find that framing the fixed fee as a quantity discount improves results. As with Lim and Ho’s (2007) work, they too consider a behavioral model for responders. My study differs from theirs in a number of ways. First, with respect to the settings, they both investigate a push context with linear demand and all human participants, whereas mine explores a pull context with random demand and one automated role. Second, because they incorporate all human participants, their behavioral model focuses on the responding party. In my work, this role is automated, thereby allowing my behavioral model to determine how proposers set contract parameters. Third, I consider two contracts different from theirs, a payback contract and SLA. Lastly, their results differ in that the behavioral models they develop are partially based on non-pecuniary payoffs, such as random errors, whereas the models I consider are exclusively preference-based.

Katok and Wu (2009) conduct an experimental study of wholesale price, buyback, and revenue sharing contracts between a supplier and retailer in a push setting. In their work, they evaluate contract decisions and stocking quantities, with particular emphasis on stocking quantities and common newsvendor biases, such as the pull-to-center effect. My work differs from Katok and Wu’s (2009) in three aspects. First, as already noted, I consider pull contracts, which differ theoretically from push contracts (see section 3 for details). Second, I devote all of the analysis and modeling to understand how contract parameters are set, as opposed to stocking quantities. Third, while their high-level experimental findings are not entirely unlike mine, my study further investigates alternative models and a number of competing interventions. Therefore, my results extend theirs by identifying which behavioral biases are at play, and how to mitigate them empirically.

Another related work on experimental supply chain contracting is Kalkanci et al. (2011). In their paper, they build on Lim and Ho’s (2007) and Ho and Zhang’s (2008) studies by evaluating two-price and three-price contracts in a push setting with asymmetric information; the retailer has more detailed information about the demand distribution than the supplier. In this context, they find that a two-price contract marginally increases supplier profits, but three-price contracts do no better than a one-price contract. They develop a simplified version of the experience-weighted attraction model (Camerer and Ho 1999) to explain contract decisions. The setting Kalkanci et al. (2011) explore differs from my study in considering push contracts with asymmetric information. Also, their results differ in that their behavioral model is largely based on learning and dynamic behavior, most likely due to the asymmetric information aspect of their study, whereas my behavioral models are static in nature.

Recently, there have been two other experimental contracting studies that are related to the one conducted here. First, Donohue et al. (2013) perform a human-subject experiment on buyback and revenue sharing contracts, where the decision maker is allowed to choose among contracts. This study is novel to the operations management literature, as it is the first (to my knowledge), which investigates why people have certain preferences for one contract over
another. However, their objective is quite different from mine, and, as with the other works mentioned, their study focuses on a push context. The second related work is Davis et al. (2014). In their paper, they directly compare the performance of a push wholesale price, pull wholesale price, and advance-purchase discount wholesale price contract, in an experimental setting with all human participants. Ultimately, they find that a wholesale price pull contract outperforms a wholesale price push contract in terms of supply chain efficiency. However, despite the favorable performance of the wholesale price pull contract, they neglect to consider any coordinating pull contracts. Similar to other studies, they too explore behavioral models for the party setting the stocking quantity, but not the party setting the contract parameters. My work extends theirs by looking at two pull coordinating contracts and how they compare to their wholesale price counterpart. Moreover, my study considers behavioral models with respect to the party setting contract parameters.

All of these studies have greatly advanced the behavioral supply chain contracting literature. My work attempts to follow this trend, by (a) experimentally investigating multiple contracts, including the first to evaluate SLAs, in a pull setting, as opposed to a push context, (b) exploring a number of behavioral models aimed at understanding how people set contract parameters, vs. models that focus on responders, and (c) directly testing the best fitting models empirically, along with other competing interventions, as opposed to relying on structural results.

3. Pull Model Overview

The setting I explore involves a single retailer and single supplier. In each period, a retailer R begins by offering contract terms to its supplier S. The supplier, upon receipt of those terms, sets a stocking quantity \( q \) for a single product, for a single period. The supplier sets its stocking quantity based on its production cost per unit \( c \), and wholesale price per unit \( w \) (and potentially other contract parameters outlined below). After both decisions are made, demand is realized. The retailer purchases product instantaneously from the supplier at the wholesale price per unit \( w \), receives revenue \( r \), for each unit sold, incurs no holding cost, and loses sales if demand is greater than the supplier stocking quantity \( D > q \). Let \( D \) represent a random variable for demand with cumulative distribution \( F \) and density \( f \). The demand distribution has the increasing generalized failure rate (IGFR) property, which is the case for many probability distributions, such as the Normal and the Uniform, and is utilized frequently in supply chain studies (Lariviere and Porteus 2001).

The supplier’s goal is to maximize its expected profit, \( \pi_s \), with respect to the stocking quantity, and the retailer’s goal is to maximize its expected profit, \( \pi_r \), with respect to the contract terms. This setup is a Stackelberg game where the retailer is the leader and the supplier is the follower. A combination of the supplier’s optimal stocking strategy, which is the best response to the retailer, and the retailer’s optimal contract parameters, given this behavior by the supplier, constitutes the subgame perfect Nash equilibrium of the game. I assume no fixed ordering costs, cost parameters are common knowledge, and suppliers and retailers are risk-neutral expected-profit maximizers. Lastly, let efficiency be defined as the ratio between the decentralized supply chain profit and the centralized supply chain profit.

3.1. Wholesale Price Model

Under a wholesale price (WP) pull contract, a retailer establishes a wholesale price \( w \). The supplier then sets a stocking quantity \( q \), for a given \( w \), that maximizes its expected profit \( \pi_s(q) \),

\[
\pi_s(q) = w\mathbb{E}[\min(q, D)] - cq.
\]

Let \( q^* \) be the quantity that maximizes the supplier’s expected profit \( q^* = \arg\max \pi_s(q) \), where \( q^* \) is implicitly defined by \( w \), such that \( q^*(w) \). It is well known that \( q^*(w) \) must satisfy

\[
F(q^*(w)) = \frac{w - c}{w}.
\]

The retailer’s decision under a WP contract is given by \( w \). Let \( w^* = \arg\max \pi_r(w) \), where \( \pi_r(w) \), the retailer’s expected profit function, is

\[
\pi_r(w) = (r - w)\mathbb{E}[\min(q^*(w), D)].
\]

Because the demand distribution has the IGFR property, there is a one-to-one relationship between \( w \) and \( q \) (Cachon 2004). Therefore, Equation (1) can be rearranged to solve for \( w \) in terms of \( q \)

\[
w(q) = \frac{c}{1 - F(q)}.
\]

One can then plug Equation (3) into (2) such that the retailer’s expected profit is a function of \( q \). Let \( \hat{q} = \arg\max \pi_R(q) \), therefore, a retailer under a pull WP contract will set \( w = \hat{w}(\hat{q}) \).

A few comments are in order regarding a WP contract in a traditional push context, and a WP contract in a pull setting. In a traditional push setting, the roles are reversed in that a supplier sets a wholesale price and the retailer sets the stocking quantity. The key difference, however, is that in the push setting, the expected profit function of the party setting the
contract terms, the supplier is deterministic, \( \pi_S = (w - c)q \), whereas in a pull setting, the retailer sets the wholesale price, and their expected profit function depends on random demand, noted above in Equation (2), \( \pi_R = (r - w)\mathbb{E}[\min(q, D)] \). In a sense, the WP pull contract allows the two parties to share the demand risk, compared to a WP push contract. The interested reader is referred to Cachon (2004) for additional details regarding the WP contract in a push vs. pull setting.

### 3.2. Payback Model

A payback (PB) contract builds on a WP pull contract in that the retailer now proposes, in addition to a wholesale price per unit, a payback amount per unit \( x \), which is paid from the retailer to the supplier for each unit that the supplier overstocks. Under a PB contract, a supplier sets a stocking quantity for a given \( w \) and \( x \), which maximizes its expected profit \( \pi_S(q) \), where \( \pi_S(q) \) under a PB contract is given by

\[
\pi_S(q) = w\mathbb{E}[\min(q, D)] - cq + \int_0^q x(q - x)dF(x).
\]

The supplier’s optimal stocking quantity, \( q^* = \arg \max \pi_S(q) \), is implicitly defined by \( w \) and \( x \). The supplier’s optimal stocking quantity under a PB contract must satisfy

\[
F(q^*(w, x)) = \frac{w - c}{w - x}.
\]

Let \( (w^*, x^*) = \arg \max \pi_R(w, x) \), where the retailer’s expected profit, \( \pi_R(w, x) \), is given by

\[
\pi_R(w, x) = (r - w)\mathbb{E}[\min(q^*(w, x), D)] - \int_0^{q^*(w, x)} x(q^*(w, x) - x)dF(x).
\]

Özer and Wei (2006) show that a payback contract can perfectly coordinate the channel with an arbitrary split of profits if

\[
x = (1 - \lambda)c,
\]

where \( \lambda \) is the supplier’s share of the total expected supply chain profit, \( \lambda = \frac{w - c}{w - x} \), and \( (1 - \lambda) \) is the retailer’s portion, \( (1 - \lambda) = 1 - \frac{w - c}{w - x} \). In this setting, since the retailer moves first and attempts to maximize her profits, she sets \( w \) in a way that extracts nearly all of the channel profit and sets \( x \) according to Equation (4) to ensure 100% efficiency.

### 3.3. Service-Level Agreement Model

Under a service-level agreement a retailer sets not only a wholesale price, but also a bonus \( \beta \), which is paid to the supplier when the realized fill rate is equal to or greater than a set target fill rate \( \tau \). The realized fill rate is defined as the fraction of satisfied demand for the retailer with the supplier stocking quantity, \( \min(q, D) \), and the probability of the supplier achieving the target fill rate and earning the bonus is \( F(\frac{q}{\tau}) \).

The supplier’s expected profit \( \pi_S(q) \), under an SLA is given by

\[
\pi_S(q) = w\mathbb{E}[\min(q, D)] - cq + \beta F\left(\frac{q}{\tau}\right),
\]

where the last term, differentiating an SLA from a WP contract, represents the expected value of receiving the bonus. The supplier’s optimal stocking quantity, \( q^* = \arg \max \pi_S(q) \), is implicitly determined by \( w \) and \( \beta \). Under an SLA, the supplier’s optimal stocking quantity must satisfy

\[
F(q^*(w, \beta)) = \frac{w - c}{w + \beta} + \beta F\left(\frac{q^*(w, \beta)}{\tau}\right).
\]

Notice that Equation (5) is the WP solution with an additional term relating to the bonus and the target fill rate where, for a fixed \( \tau > 0 \), any optimal solution \( q^*(w, \beta) \) is non-decreasing in \( \beta \).

Let \( (w^*, \beta^*) = \arg \max \pi_R(w, \beta) \), where the retailer’s expected profit, \( \pi_R(w, \beta) \), is

\[
\pi_R(w, \beta) = (r - w)\mathbb{E}[\min(q^*(w, \beta), D)] - \beta F\left(\frac{q^*(w, \beta)}{\tau}\right).
\]

As with the supplier’s expected profit, the retailer’s expected profit under an SLA is the same as a WP contract except for an extra term representing the expected value of paying out the bonus.

An SLA, in many but not all cases, can perfectly coordinate the supply chain with an arbitrary split of profits. In the case of demand uniformly distributed between 0 and \( Z \), the same demand distribution used in the experiment of this study, an SLA can achieve 100% efficiency for any positive target fill rate (please see the Appendix for the proof).

**Proposition 1.** Under an SLA contract, if demand is uniformly distributed between 0 and \( Z \), then, for any wholesale price \( w \) and positive fill rate \( \tau > 0 \), a bonus \( \beta \) can be found that perfectly coordinates the supply chain, where \( \beta \) must satisfy the following condition:

\[
\beta = c(1 - \lambda)Z.
\]

where \( \lambda \) is the supplier’s share of the total expected supply chain profit \( \lambda = \frac{w}{\tau} \), and \( (1 - \lambda) \) is the retailer’s portion \( (1 - \lambda) = \frac{w}{\tau} \).

As with the PB contract, since the retailer moves first and wants to maximize her profits, she sets \( w \) in a way that extracts nearly all of the channel profit. And
then sets $\beta$ according to Equation (6) to ensure 100% efficiency.

4. Experimental Design and Results

4.1. Experimental Design

The procedure for participants in each session was as follows. Upon entering the laboratory, participants were seated at a private computer terminal and given a few minutes to read over the instructions themselves. Following this, I read the instructions aloud and answered any questions in private. Then in each round of each treatment a participant, playing the role of a retailer, proposed contract terms to a single, automated supplier, where participants were aware that the supplier was automated. The supplier, upon receiving the retailer’s proposed terms, either set a quantity of 0 (which occurred when the supplier’s best response could not achieve a positive expected profit), or set a positive stocking quantity that maximized its own expected profit. Following both of these decisions demand was realized, and the participants acting as retailers viewed the results with respect to the decisions made and their own, private profit. This procedure was repeated for 60 rounds.

I evaluated each of the three contracts outlined in section 3, a WP contract, a PB contract, and an SLA, in separate treatments to avoid order effects. In the WP treatment, the retailer proposed a wholesale price per unit, for each unit sold. In the PB treatment, the retailer proposed a wholesale price, and a payback amount, which was paid for each unit of the supplier’s stocking quantity that exceeded demand. Lastly, in the SLA treatment, the retailer proposed a wholesale price, and a bonus that was paid to the supplier whenever the supplier’s stocking quantity satisfied 100% of demand.

My $3 \times 1$ between-subjects design totaled three treatments and used the same revenue and cost parameters, $r = 20$ and $c = 4$, which were common knowledge. In all treatments, demand was set to be a uniform integer between 0 and 100 that was independent of each period. Table 1 summarizes the design of experiment and the number of participants in each treatment, where the individual participant is the main unit of statistical analysis.

The standard theory outlined in section 3 assumes that all decision makers have the ability to make optimal choices. Therefore, to give the standard theory its best opportunity of being confirmed, and to mitigate random errors as a driver of results, I provided participants with a decision support tool where they could move scroll bar(s) that corresponded to their decision variable(s) and test different values. For example in the SLA treatment, a participant could move two scroll bars, one for the wholesale price and one for the bonus, and each time the participant adjusted one of these scroll bars, the screen displayed the supplier’s optimal stocking quantity for the set of test values, along with a graph illustrating the retailer profit for every value of demand, assuming this supplier stocking quantity (participants were told that the supplier would stock this quantity if they chose those values). Please see the online Appendix S1 for sample instructions which provide a screenshot of the experimental software.

Using the theory outlined in section 3 and the experimental parameters, it is straightforward to identify the predicted decisions and corresponding expected profits, net endowments. Those values are provided in Table 2. Note that the predicted supplier profit for the PB contract and SLA is 2 rather than 1. This is because of rounding restrictions in the experiment. In the PB contract, retailers were allowed to enter up to two decimal places for $w$ and $z$. For the SLA, participants could enter up to two decimal places for $w$, and integers for $\beta$.

In Table 2, we observe that the PB contract and SLA should outperform the WP contract in terms of retailer profit and supply chain efficiency. More specifically, as highlighted in section 3, the two coordinating contracts should result in 100% supply chain efficiency.

All of the experimental sessions were run at a large northeast US university in the fall of 2009 and summer of 2010. Participants in all treatments were students, mostly undergraduates from a variety of majors, who were recruited using an online system, where cash was the only incentive offered. Participants were paid a $5 show-up fee plus an additional amount that was based on their personal performance. Average compensation for the participants across all treatments, including the show-up fee, was $22. Each session lasted approximately 45 minutes to 1.5 hours, and the software was programmed using the zTree system (Fischbacher 2007).

<table>
<thead>
<tr>
<th>Table 2 Predicted Values for Each Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Retailer profit</td>
</tr>
<tr>
<td>Supplier profit</td>
</tr>
<tr>
<td>Efficiency (%)</td>
</tr>
<tr>
<td>Quantity</td>
</tr>
<tr>
<td>Wholesale price</td>
</tr>
<tr>
<td>Payback</td>
</tr>
<tr>
<td>Bonus</td>
</tr>
</tbody>
</table>

Please Cite this article in press as: Davis, A. M. An Experimental Investigation of Pull Contracts in Supply Chains. Production and Operations Management (2014), doi 10.1111/poms.12224
4.2. Results
To present the initial experimental results, I first report summary statistics for profits and decisions. Following this, I will focus on any experience effects and investigate individual-level behavior. For all results, I calculate expected profit of retailers’ decisions, net endowments, and report it as observed profit. Lastly, unless otherwise stated, the individual is used as the primary unit of statistical analysis, generating 20 independent observations in each treatment.

Figure 1 depicts the predicted and observed retailer profits for all three treatments. The vertical axis is scaled to illustrate the relative performance of all three contracts to each other and theory. Comparing the observed profit of each contract to one another, there is a significant difference in retailer profit between the WP contract and the two coordinating contracts; the PB contract and SLA (\( p = 0.032 \) and \( p = 0.003 \), two-sided t-tests, which will be the default statistical test unless otherwise noted).

Also in Figure 1, the observed retailer profits are slightly below predicted levels for the WP contract, and far below predicted levels for the PB contract and SLA (all \( p < 0.01 \)). Therefore, the predicted benefit of the coordinating contracts over the WP contract is far below what it should be in theory.

Table 3 delineates retailers’ average contract decisions. Starting with the left-most columns of this table, one can see that retailers in the WP contract set wholesale prices slightly higher than optimal. While the difference between predicted and observed wholesale prices is statistically significant in the WP contract, it appears as though retailers set wholesale prices rather well. In the PB contract and SLA, retailers made decisions poorly relative to predicted values. In both of these contracts, retailers set the wholesale price too high and the coordinating parameter, too low. This behavior drives the supply chain efficiency far below 100%, to values of 91.6% in the PB contract and 89.6% in the SLA.

The original predictions outlined in Table 3 indicate that, in theory, \( w^* \) under a WP contract should be higher than \( w^* \) under a PB contract, which should be higher than \( w^* \) under an SLA (\( w^*_{WP} > w^*_{PB} > w^*_{SLA} \)). Despite retailers setting \( w \) too high within each contract, this comparative static exists across all three contracts. This suggests that participants generally responded correctly to each contract when setting \( w \), but not as strongly as they should (the differences are statistically significant for all comparisons of \( w \) except between the WP and PB contracts).

Past work has suggested that over time, decision makers in supply chains may learn with experience for their given task (Bostian et al. 2008). Figure 2 illustrates the observed retailer profit over time, aggregated every three decision periods. As one can see in Figure 2, it seems that retailers learned to increase profits over the first half (roughly) of their decisions. To formally confirm this, I ran a piecewise regression of retailer profits on the decision period with a knot at period 30. The results are shown in Table 4. Looking at the coefficients in Table 4, retailers clearly learned to increase retailer profit prior to period 30, as evidenced by the positive and significant coefficients on \( Per \leq 30 \), with the strongest learning taking place in the SLA. Because of the learning in early periods, I compared the retailer profit of all three contracts to each other only considering the second half of decisions. After doing this, I find the same conclusions as earlier; the PB contract and SLA outperform the WP contract, but the benefit of the coordinating contracts over the WP is smaller than predicted. In particular, the retailer profit is 440 for the WP contract, 464 for the PB contract, and 473 for the SLA, so that the two coordinating contracts are still far from their theoretical prediction of 638. In addition, I continue to observe that in both coordinating contracts, retailers set the wholesale price too high and the coordinating parameter too low, relative to the normative benchmark.

Analyzing individual behavior leads to similar results as those reported previously. To illustrate this, I ran 60 regressions, one for each participant, with retailer profit as the dependant variable, and the decision period shifted by 60, as the independent variable. Shifting the decision period by 60 allows one to account for any experience effects. Specifically, if the 95% confidence interval for the intercept does not include the normative expected profit for a particular retailer, then one can conservatively say that that retailer made decisions which were suboptimal. I ran all regressions with robust standard errors to account...
For any heteroskedasticity within an individual subject (White 1980). Using this approach, in the WP contract, only seven out of 20 retailers made decisions that were suboptimal, agreeing with the earlier results that retailers made relatively good decisions in the WP contract. However, 19 out of 20 retailers in the PB contract, and all 20 in the SLA, made decisions that were suboptimal.

In summary, on both an aggregate and individual level, retailers in the WP contract set contract parameters fairly well, but in the two coordinating contracts, they neglect to set contract parameters that coincide with the normative predictions.

5. Behavioral Models

The previous results suggest that in the WP contract, retailers set wholesale prices reasonably well. However, in the PB contract and SLA, they fail to identify the optimal contract parameters. In particular, in both the PB contract and SLA, retailers set the wholesale price too high and the coordinating parameter too low, leading to reduced profits. Given these results, in this section, I explore a number of different behavioral models and apply them to the PB contract and SLA. After this, I structurally estimate the parameters of each of these models through MLE and determine whether any of the models evaluated can organize the data better than the normative theory.

There have been a number of behavioral models applied to laboratory supply chain studies. Despite an attempt to design my experiment in a way to mitigate these (such as random errors or bounded rationality, through providing retailers with a decision support system, and social preferences, through automating suppliers), there are undoubtedly a number of models that may accurately describe the data. Therefore, I consider three alternative models well known in the literature: risk aversion, loss aversion, and reference dependence. For all of these models, and to most closely match the setting used in the experiment, I assume that a retailer is proposing terms to a risk-neutral expected profit maximizing supplier. Specifically, suppliers set stocking quantities according to the standard theory outlined in section 3, which correspond to

\[
F\left(\frac{q}{C}; \alpha\right) = \frac{w - c}{w} + \frac{b}{w}f\left(\frac{q}{\alpha}\right)\]

in the PB contract, and

\[
F\left(q^*(w, \beta)\right) = \frac{w - c}{w} + \frac{b}{w}f\left(\frac{q^*(w, \beta)}{\tau}\right)\]

in the SLA. To simplify the notation, I will refer to \(q^*(\cdot)\) as \(q^*\).

5.1. Risk Aversion

One possible model for explaining retailer decisions is risk aversion. This stems from the idea that retailers may interpret the coordinating parameter, \(\alpha\) or \(\beta\), as being riskier compared to the wholesale price. Specifically, retailers may assume that the coordinating parameter is a more variable term, as it may or may not change in a predictable manner.
Given that retailers tend to set the coordinating parameter too low, another plausible explanation for the observed data is that retailers perceive this extra payment as an unnecessary loss. Loss aversion (Tversky and Kahneman 1991) is one model that coincides with this behavior, and has recently been observed in related contexts. For instance, Ho and Zhang (2008) explore a two-part tariff, which includes a fixed fee, and theorize that the fixed fee may be interpreted as a loss to the decision maker. They base part of their model on the value function of prospect theory (Kahneman and Tversky 1979) and mental accounting (Thaler 1999). These state that a decision maker generates two mental accounts, one for the variable retailer profit which is a gain, and one for the fixed fee which is a loss, where one dollar in each account may not be interpreted as equal. Their results suggest that this is indeed true; they find that the decision maker weighs the cost of the fixed fee more than its true value.

In a pull context, a retailer establishing contract parameters in the PB contract and SLA may interpret the regular expected profit function in the realm of gains, and any paid out payback amount or bonus in the realm of losses. Following the approach of Ho and Zhang (2008, p. 694, equation (1)) leads to

\[
PB : \quad u(w, x) = \pi_R(w, x) - \int_0^q \gamma(x(q^* - x))x\ dF(x) \\
SLA_{\tau=1} : \quad u(w, \beta) = \pi_R(w, \beta) - \int_0^q \gamma\beta x\ dF(x),
\]

where \(u(\cdot)\) represents the expected-utility function of the retailer, and \(\gamma\) represents the loss aversion parameter, or the amount a subject may overweight the cost of paying the coordinating parameter amount. For example, if \(\gamma > 0\) in the SLA, then a decision maker suffers some extra psychological cost of paying out the bonus.

The loss aversion formulation depicted above extends Ho and Zhang (2008) by including an additional parameter \(k > 0\), which defines the shape of the disutility in the coordinating parameter amount. In general, for \(k = 0.5\) and 1, I find that a loss averse retailer, for both the PB contract and SLA, should set contract parameters in line with the standard theory if \(\gamma\) is below a given threshold. However, once \(\gamma\) exceeds this threshold, and the retailer becomes sufficiently averse to paying out the coordinating parameter, the predictions starkly change such that a retailer should set the wholesale price equal to 8.00 and the coordinating parameter roughly equal to 0. This does not coincide with the observed data, and will cause identification issues in the estimation. However, a loss aversion model, where the disutility is increasing and convex in the coordinating parameter, yields interior solutions which are in line with the experimental data. Therefore, for the structural estimation results, I assume losses are quadratic in the coordinating amount, \(k = 2.3\).

5.3. Reference Dependence

Another model that has been shown to fit experimental data well is reference dependence (Ho et al. 2006, Kahneman and Tversky 1979, Thaler 1985). Recently, in a push newsvendor setting, Ho et al. (2010) show that a newsvendor’s utility function, under reference dependence, is comprised of the regular expected-profit function, plus two disutility terms: one term associated with the psychological cost of having leftover units or overages, and a second term associated with the psychological cost of stockouts or underages. Ho et al. argue that decision makers use observed demand as the reference point, and is one of the most salient pieces of information provided to newsven- dors after setting a stocking quantity. Specifically, page 1894 of Ho et al. (2010, equation (1)), displays the expected-utility function of the retailer setting a quantity, in a push setting as

\[
u(q) = \pi_R(q) - \int_0^q \delta_x(q - x)\ dF(x) - \int_0^\infty \delta_x(q - x)\ dF(x),
\]

where \(\pi_R(q)\) is the retailer’s expected profit function, \(\delta_x \geq 0\) represents the psychological per-unit cost of an overage, and \(\delta_u \geq 0\) is the psychological per-unit cost of an underage. Again, this is for a party setting quantities in a push setting, which differs from the setting explored in this study. Consequently, a direct application of this formulation to my setting yields a poor overall fit relative to the observed data. However, one can extend the reference dependence formulation of Ho et al. (2010) to retailers in a pull...
setting. First, a natural reference point is the level of demand in which the coordinating parameter is paid out, as it triggers any additional payments made to the supplier. In the case of the experiment I explore, this coincides with realized demand. Second, one can compound the same two disutility terms in Equation (7), by the coordinating contract parameters. In the case of having overages, one can premultiply the leftover units expression by the coordinating parameter, \( \alpha \) or \( \beta \). This essentially says that a retailer suffers disutility from setting contract terms in a way that induced the supplier to overstock, as the supplier’s optimal stocking quantity is still a function of both contract parameters, but that this disutility is exacerbated by setting the coordinating parameter unnecessarily high. In the case of underage disutility, the same logic is applied, such that the stockout term is divided by the coordinating parameter. This makes it so that a retailer incurs a psychological cost from not only setting contract parameters in a way that caused the supplier to understock, but that, by dividing, this cost is worsened if the coordinating parameter is set exceptionally low. The resulting expected-utility functions for the retailer setting contract parameters in the pull setting are then

\[
\begin{align*}
PB : u(w, x) &= \pi_R(w, x) - \int_0^q \delta_x (q^* - x) dF(x) \\
&\quad - \int_q^\infty \delta_x x^{-1} (x - q^*) dF(x) \\
SLA_{\tau=1} : u(w, \beta) &= \pi_R(w, \beta) - \int_0^q \delta_\beta (q^* - x) dF(x) \\
&\quad - \int_q^\infty \delta_\beta \beta^{-1} (x - q^*) dF(x),
\end{align*}
\]

where \( u(\cdot) \) represents the expected-utility function of the retailer. Maximizing these functions for the experimental setting used in this study suggests that as the proportion of disutility associated with overages to underages increases, \( w^* \) increases while \( x^* \) or \( \beta^* \) decreases. These predictions qualitatively agree with retailers’ decisions and suggest that the reference dependence model will also result in a favorable fit to the data.

5.4. Structural Estimation

I now turn to MLE to structurally estimate the models outlined in the previous section and see which models organize the data best. For simplicity, I outline the parameter estimation process used for the risk aversion model in the PB contract.

Let \( i \) denote the index for a individual set of contract parameters, \( i = 1, \ldots, N \) where \( N \) is the total number of decisions estimated. I assume the errors are distributed normally with mean 0 and variance \( \sigma^2_\text{w} \) and \( \sigma^2_\text{x} \) with correlation \( \rho_{\text{wx}} \) where the individual decisions \( w_i \) and \( x_i \) follow the bivariate normal distribution:

\[
(w_i, x_i) \sim N\left( \left( \begin{array}{c} w^*(\eta) \\ x^*(\eta) \end{array} \right), \left( \begin{array}{cc} \sigma^2_\text{w} & \rho_{\text{wx}} \sigma_\text{w} \sigma_\text{x} \\ \rho_{\text{wx}} \sigma_\text{w} \sigma_\text{x} & \sigma^2_\text{x} \end{array} \right) \right)
\]

where \( w^*(\eta) \) and \( x^*(\eta) \), respectively, are the wholesale price and payback amount that maximize the retailer’s expected-utility function under the risk aversion model for a particular selection of \( \eta \). The log-likelihood (LL) function, in the case of the PB contract for the risk aversion model, which is maximized over four parameters, is given by

\[
LL(\eta, \sigma_\text{w}, \sigma_\text{x}, \rho_{\text{wx}}) = \sum_{i=1}^N \left( -\ln(2\pi) - \frac{1}{2} \ln |\Omega| \right. \\
- \frac{1}{2} \left( \frac{w_i - w^*(\eta)}{\sigma_\text{w}} \right)^T \Omega^{-1} \left( \frac{w_i - w^*(\eta)}{\sigma_\text{w}} \right) \\
- \frac{1}{2} \left( \frac{x_i - x^*(\eta)}{\sigma_\text{x}} \right)^T \Omega^{-1} \left( \frac{x_i - x^*(\eta)}{\sigma_\text{x}} \right)
\]

where \( \Omega \) is the covariance matrix.

Recall that there is evidence of learning in the first half of decisions. Therefore, to avoid any estimation issues associated with serial correlation, I remove the first half of the decisions and fit the remaining data to each of the models (30 decisions for each participant, constituting 600 decisions in each treatment).

In the case of the risk aversion and loss aversion models, I estimate \( \eta \) or \( \gamma \), plus the covariance matrix, and report the results directly. For the reference dependence model, I estimate five parameters, \( \delta_\text{w}, \delta_\text{x}, \sigma_\text{w}, \sigma_\text{x} \), and the covariance matrix. To provide a straightforward interpretation of the estimates in the reference dependence model, after I estimate \( \delta_\text{w} \) and \( \delta_\text{x} \), I calculate the value of each disutility term and report the ratio of the two. If this ratio exceeds 1, then retailers tend to incur more disutility from overages and setting the coordinating parameter too high. Conversely, if this ratio is less than 1, then retailers incur more disutility from underages and setting the coordinating parameter too low. Lastly, I generated the standard errors for all estimates through bootstrapping the data.

Table 5 delineates the LL values, likelihood-ratio (LR) test results comparing the three behavioral models to the standard theory, and maximum-likelihood estimates for the standard theory, risk aversion, loss aversion, and reference dependence models. Focusing first on the individual estimates, with respect to risk
aversion, the observed levels of $\eta$ are somewhat expected. For example, $\eta = 0.0027$ implies that a retailer would be indifferent between a 50–50 chance of $0$ or $100$, or a certainty equivalence of $46.64$. In short, retailers did not exhibit high levels of risk aversion, which may be due in part to the natural risk-sharing property of pull contracts (recall that under a push contract, a supplier completely avoids demand risk, whereas under a pull contract, both parties face demand risk). For the loss aversion model, $\gamma$ is positive and significant in both the PB contract and SLA, suggesting that retailers weigh the cost of paying out the coordinating amount more than its true value. Lastly, in the reference dependence model, it appears that across both contracts, the disutility from overages and setting high coordinating parameters tends to outweigh the disutility of underages and low coordinating parameters, as the ratio of the two disutility terms are both greater than 1.6

Turning now to the overall $LL$ values, moving from left to right in Table 5, a series of likelihood-ratio tests shows that all three alternative models fit the data better than the standard theory. To determine which of the three behavioral models fits the data best for each contract, one can conduct a series of Vuong tests (Vuong 1989), which allows for comparisons among non-nested models. For both the PB contract and SLA, the loss aversion and reference dependence models fit the data better than risk aversion. Between loss aversion and reference dependence, it appears that loss aversion generates a slightly better fit in the PB contract, and that reference dependence fits best for the SLA. Table 6 depicts the predicted contract parameters for the PB contract and SLA, given the structural estimates from Table 5 for the two best fitting models, loss aversion and reference dependence. Note that there is still some room for improvement in the PB contract, but both models are a significant improvement over the standard theory overall.7

5.5. Managerial Intervention
While the loss aversion and reference dependence models provide a better fit of the data in a statistical sense, this does not necessarily mean that they are the

<table>
<thead>
<tr>
<th>Table 5 Structural Estimation Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard theory</strong></td>
</tr>
<tr>
<td><strong>PB</strong></td>
</tr>
<tr>
<td>Combined $LL$</td>
</tr>
<tr>
<td>Contract $LL$</td>
</tr>
<tr>
<td>LR test</td>
</tr>
<tr>
<td>$\eta$</td>
</tr>
<tr>
<td>$\gamma$</td>
</tr>
<tr>
<td>Overage/Underage</td>
</tr>
<tr>
<td>$\sigma_{\eta}$</td>
</tr>
<tr>
<td>$\sigma_{\kappa, \beta}$</td>
</tr>
<tr>
<td>$\rho$</td>
</tr>
</tbody>
</table>
| Standard errors are reported in square brackets.

<table>
<thead>
<tr>
<th>Table 6 A Comparison of the Predicted Contract Values Using the Loss Aversion and Reference Dependence Models to the Observed Values from Periods 31–60</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PB</strong></td>
</tr>
<tr>
<td>Loss aversion</td>
</tr>
<tr>
<td>Wholesale price</td>
</tr>
<tr>
<td>Payback</td>
</tr>
<tr>
<td>Bonus</td>
</tr>
</tbody>
</table>
| Standard errors are reported in square brackets.

Please cite this article in press as: Davis, A. M. An Experimental Investigation of Pull Contracts in Supply Chains. Production and Operations Management (2014), doi 10.1111/poms.12224
cause of retailers’ suboptimal behavior. Therefore, I ran an additional experimental treatment that controls for these effects. Specifically, I did not show participants whether the bonus or total payback amount was paid out, nor the reference point, realized demand. Note that these biases are closely intertwined, making it difficult to control for one at a time: if a retailer knows the reference point, then they automatically know whether the coordinating parameter was paid out, conversely, if a retailer knows that the coordinating parameter was paid out, then they have information about demand (in fact, in the PB contract, because the coordinating amount is a function of demand, they can perfectly infer true demand). Retailers did, however, continue to observe realized profit and the suppliers’ stocking quantities. In these treatments, I used the same experimental parameters and procedures as those outlined in section 4. The PB contract had 13 participants and the SLA treatment had 12 participants.

As with the original experiment, there was considerable learning in the early periods of these new treatments. Therefore, for comparing the results to the original experiment, the profits for the second half of decisions are presented. Figure 3 depicts the observed retailer profit for the original “Baseline” treatments and new loss and reference dependence, “LRD” treatments, for periods 31–60.

Retailer profits in both coordinating contracts are statistically higher than the Baseline treatments ($p < 0.05$ for the PB contract and $p < 0.01$ for the SLA). However, it should be pointed out that while the LRD manipulation results in improved profits, it cannot fully explain the deviation from theoretical predictions of 638.

Table 7 depicts the average contract proposals by retailers in the Baseline and LRD treatments. Following from Figure 3, the LRD treatments are a significant improvement over the Baseline treatments for both the PB contract and SLA.

In short, the previous analysis demonstrated that the loss aversion and reference dependence formulations outlined in section 5 fit the data well in a statistical sense and are therefore useful from a modeling perspective. However, the results from these additional treatments provide clear empirical evidence that these biases exist in pull contracts, and that controlling for them can improve decisions and profits.

6. Competing Interventions

It is plausible that a variety of other interventions could generate an improvement in behavior. For example, decisions aggregated over multiple periods might control for risk aversion. Similarly, a plethora of literature has supported the notion that decision makers are susceptible to random errors (Su 2008), and that reducing the complexity of a problem might result in better performance (Kalkanci et al. 2011).

To try and tease out these explanations, I conducted three additional “Intervention” manipulations. In all three of these, I used the same experimental parameters and procedures as in all other sessions. For each of the six treatments (2 contracts × 3 manipulations), there were 10 participants. Also, each of the interventions I evaluated was mutually exclusive (no two interventions were used at the same time). That way if there are any significant differences in the outcomes, then I can attribute them directly to that specific manipulation. The interventions I administered were as follows:

1. Single Decision (SingDec): Past studies have shown that an increase in contract complexity does not always lead to an increase in the performance of contracts (Kalkanci et al.
Therefore, in the first intervention, I altered the coordinating contracts so that the decision space was now a single variable, the coordinating parameter. For example, in the SLA treatment participants set a bonus, and for each bonus they set, the corresponding optimal wholesale price, conditioned on that bonus, was automatically used (it was also presented to them in their decision support tool). This allows all three contracts, WP, PB, and SLA, to have the same number of decision variables. Therefore, this treatment provides a means of determining whether the level of contract complexity in the two coordinating contracts is driving their profit down from theoretical predictions.

2. 10 Periods (10P): In this intervention, retailers’ earnings were based on the average of 10 demand realizations, rather than one. For example, when a subject entered a decision, 10 realizations of demand were drawn (all 10 draws were shown to the retailers in the results screen), where participants were paid the average profit from each set of 10 draws. Therefore, in this treatment, a retailer’s 60 decisions allowed them to observe 600 demand realizations. In other contexts, such as auctions, this manipulation has been shown to determine whether risk aversion is at play (Engelbrecht-Wiggans and Katok 2009). As such, the results from this treatment allow me to confirm whether or not risk aversion is prevalent in pull contracts.

3. Expected Profit Support (ExpProf): In the third intervention retailers were provided with an additional piece of decision support, expected profit. This allowed retailers to test different values and see, in addition to their previous decision support, their expected profit. Past newsvendor studies have illustrated that participants may be limited in their thought processes and exhibit bounded rationality (Su 2008). This intervention allows me to determine if subjects simply could not compute the optimal values due to limited cognitive ability, or rule out random errors as an explanation for behavior (if they continue to set suboptimal contract parameters).

Figure 4 depicts the observed retailer profit for the Baseline treatments, three Intervention treatments, and LRD treatments for periods 31–60. Looking at the three Intervention treatments (SingDec, 10P, and ExpProf) in Figure 4, in both the PB contract and SLA, there is only a marginal improvement in retailer profits over the Baseline treatments.

In particular, there is weak significance comparing the Baseline treatment to the SingDec treatment in the PB contract. In addition, none are above the LRD treatment results.

Table 8 shows the average contract proposals by retailers from all treatments. Following from Figure 4, there is weak significance between the SingDec treatment and Baseline treatment for the PB contract. The only contract proposals that are significantly different from the Baseline treatments are the LRD decisions. Lastly, the LRD treatment is the only manipulation where the wholesale price was set below the supplier’s cost, agreeing qualitatively with the theoretical prediction.

While outside the scope of this study, one question is why we do not see more improvement in decisions for the competing intervention treatments. One informal hypothesis is that two of the interventions outlined above, SingDec and ExpProf, are based on past studies which structurally estimate behavioral models, as opposed to direct empirical tests. For example, while contract complexity appears to be a factor in other studies from a model fitting sense, no study has actually conducted an experimental treatment, in a supply chain contracting setting, that dramatically increases the level of decision support aimed at empirically confirming or rejecting this theory. A second hypothesis is that the ExpProf and 10P interventions are motivated from past newsvendor and procurement studies, two very different contexts than a pull contracting setting. In fact, one example of an intervention not improving performance in an alternative setting is Davis et al. (2011). They provide the same decision support as used in the ExpProf treatment here, for people setting reserve prices in auctions, and find virtually no change in behavior.
relative to their original treatments, where no support was provided. When one considers these aspects, combined with the fact that the pull context has different characteristics than a push context (outlined earlier), it is not unexpected that these theories do not automatically extend to the setting explored here.

7. Conclusion

There is clear evidence that pull contracts are used extensively in practice. While there also exists a considerable body of research on behavioral supply chain contracting, this stream of work has generally overlooked pull contracts. My study is the first to address this gap in the literature. I first evaluate three pull contracts in a controlled laboratory setting: a wholesale price contract which is simple for companies to administer, a payback contract which perfectly coordinates the supply chain with an arbitrary split of profits, and a SLA which is often seen in practice. I find that the performance of the two coordinating contracts is better than that of the wholesale price contract, but that they fall far short of their standard theoretical predictions.

After examining a number of behavioral models, I find that two closely related models, loss aversion and reference dependence, fit the data well. I then proceed to test whether controlling for these biases can actually improve decisions empirically. Through a supplemental experimental treatment that hides whether the coordinating parameter is paid out along with realized demand, I find that decisions substantially improve, thus providing further evidence that these biases are at play in pull contracts. In particular, this simple manipulation results in a 12.9% improvement in retailer profit for the payback contract, and a 12.7% improvement in retailer profit for the SLA.

To determine whether other managerial interventions might provide similar improvements, I conducted a series of additional experiments. After administering these sessions, I find little improvement in performance when reducing the complexity involved in setting contract terms, curtailing the variance in the mean profits, or providing additional decision support. One potential reason for this lack of improvement is that many of the interventions are based on either a different supply chain context or the structural fitting of a behavioral model, as opposed to an empirical test that controls for the relevant bias. For example, in this study, the risk aversion model provided a favorable structural fit to the data, but the results of a treatment controlling for risk aversion did not lead to improved decisions. As such, I believe that directly testing for behavioral biases, through controlled laboratory interventions, is an important opportunity for future research.

Given the favorable performance of the loss aversion and reference dependence models, it is important to recognize the relationship between the two. First, the reference dependence formulation I employ incorporates some degree of loss aversion, in that the disutility term from overstocking is increasing in the coordinating parameter. Second, the loss aversion model naturally includes some reference dependence, in that the reference point is the retailer’s expected profit, and the coordinating parameter is perceived as an extra loss. In short, loss aversion and reference dependence are closely related, therefore it is not entirely surprising that both perform well in organizing retailer decisions.

One limitation of this work is that it explores a setting where a retailer dictates contract terms to the supplier. As previously mentioned, this was done to not only provide a direct test of the theoretical models on pull contracts, but also because this practice is used in industry. However, pull contracts may also be used in settings where the contract terms are negotiated between the two parties. While my results may qualitatively inform how pull contracts perform under a more dynamic bargaining structure, they may not perfectly extend to these alternative bargaining environments. Studies looking at more back-and-forth bargaining settings have recently begun to appear in the operations management literature and

Table 8  Average Retailer Decisions for the Baseline, Intervention, and LRD Treatments for Periods 31–60

<table>
<thead>
<tr>
<th></th>
<th>PB</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>SLA</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>SingDec</td>
<td>10P</td>
<td>ExpProf</td>
<td>LRD</td>
<td>Baseline</td>
<td>SingDec</td>
<td>10P</td>
<td>ExpProf</td>
<td>LRD</td>
</tr>
<tr>
<td>Wholesale price</td>
<td>7.92</td>
<td>6.89*</td>
<td>7.09</td>
<td>6.44</td>
<td>6.08***</td>
<td>4.45</td>
<td>5.61</td>
<td>4.53</td>
<td>5.14</td>
<td>3.01***</td>
</tr>
<tr>
<td>[0.35]</td>
<td>[0.15]</td>
<td>[0.63]</td>
<td>[0.52]</td>
<td>[0.43]</td>
<td>[0.55]</td>
<td>[0.33]</td>
<td>[0.73]</td>
<td>[0.92]</td>
<td>[0.33]</td>
<td></td>
</tr>
<tr>
<td>Payback</td>
<td>1.47</td>
<td>1.63*</td>
<td>2.20</td>
<td>2.21</td>
<td>2.86***</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>[0.26]</td>
<td>[0.20]</td>
<td>[0.42]</td>
<td>[0.42]</td>
<td>[0.33]</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Bonus</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>230.8</td>
<td>140.3</td>
<td>203.1</td>
<td>202.7</td>
<td>319.1***</td>
</tr>
<tr>
<td>[20.8]</td>
<td>[18.4]</td>
<td>[34.3]</td>
<td>[47.9]</td>
<td>[21.9]</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

Standard errors are reported in square brackets. Significance of Hotelling T-square test, comparing the Baseline to each Intervention treatment, given by \( *p < 0.10, **p < 0.05, \) and \( ***p < 0.01. \)
provide an opportunity for future research on pull contracts (e.g., Davis and Leider 2014, Leider and Lovejoy 2013).

Lastly, my work provides a number of managerial implications. First, if a retailer is considering drop-shipping, JIT, vendor-managed inventory, or any other type of pull structure in practice, they should opt for a payback contract or SLA. Second, when implementing one of these contracts, they should decouple the party that establishes the contract terms from the party processing payments. For many retailers, this is not uncommon. For instance, in many companies top-level management is typically responsible for setting long-term contracts with their suppliers, whereas other employees and departments help process the individual payments to suppliers. In addition, large retailers may work with multiple suppliers. These two effects provide a natural mechanism for mitigating loss aversion and reference dependence; the executives who set the contract terms do not observe individual payments and realized demand levels, instead, they observe their company’s financial statements at a level where payments and demand are aggregated across suppliers and time. This is one plausible scenario for why we observe the prevalence of contracts such as SLAs in practice, and also supports the findings of this study, which suggest that controlling for loss aversion and reference dependence, in pull coordinating contracts, results in substantially higher retailer profits.

Acknowledgments

Financial support was provided by the National Science Foundation, grant 0962179, and the Penn State Smeal College of Business Dissertation Summer Stipend. I would like to thank Elena Katok, Tony Kwasnica, Gary Bolton, Doug Thomas, and Natalia Santamaria for their comments. I would also like to thank seminar participants at the Smeal College of Business at the Pennsylvania State University, the Harvard Business School, and the Behavioral Operations Conference at the Darden School of Business at the University of Virginia. Lastly, this study was greatly improved by comments from the entire review team: Karen Donohue, an anonymous senior editor, and three anonymous referees.

Appendix

Proof of Proposition 1. Using the condition from Equation (6), and the fact that \( w = \lambda r \), one can verify that the expected profit for a supplier from an SLA can be expressed as a portion \( \lambda \) of the total expected supply chain profit \( \Pi \), and that the expected profit for a retailer from an SLA can be expressed as the remaining portion \((1 - \lambda)\).

\[
\begin{align*}
\pi_S(q) &= (w - c)E[D] - (w - c)E[D - q]^+ \\
&= (\lambda r - c)E[D] - (\lambda r - c)E[D - q]^+ \\
&- cE[D - q]^+ + (1 - \lambda)q \\
&= \lambda \Pi \text{ if } \lambda \leq 0.05 \\
&= (1 - \lambda)\Pi \text{ if } \lambda > 0.05
\end{align*}
\]

Notes

1To keep the task in the domain of gains rather than losses, I provided each party with an endowment of 400 laboratory dollars for each round of all three treatments.

2Even though suppliers were automated in the experiment, it is worth noting that the observed profit for the suppliers, given the retailers’ decisions, were 124, 128, and 105, respectively, in the WP contract, PB contract, and SLA, with no statistical differences between them. The total supply chain profit from the initial treatments generates the same conclusions as the retailer profit; the WP contract is significantly lower than the PB contract and SLA, but the expected gain of the coordinating contracts is smaller than what theory predicts. Instead, the observed supply chain efficiencies are 87.3% for the WP contract, 91.6% for the PB contract, and 89.6% for the SLA.

3\( k \in \{1.5, 2, 2.5, 3\} \) all yield similar fits to the data.

4Note that for the SLA, in the general case, this reference point would simply be realized demand multiplied by the target fill rate.

5By dividing, this incorporates a non-linear component. Recall that the standard theory predicts contract values that are close to the boundaries (i.e., \( w^* = 0.05 \) and
β = 0.079 [0.001], d = 5690 [100.5].

One informal explanation in the PB contract is that the retailer’s expected-utility function is somewhat flat for a particular range of wholesale prices and payback amounts. If one plugs the estimates from Table 6 using both the pre-

References


Schweitzer, M. E., G. P. Cachon. 2000. Decision bias in the news-


**Supporting Information**

Additional Supporting Information may be found in the online version of this article:

Appendix S1: Instructions