



Invited research paper

## Fairness in supply chain contracts: A laboratory study

Elena Katok<sup>a,\*</sup>, Valery Pavlov<sup>b</sup><sup>a</sup> Jindal School of Management, The University of Texas at Dallas, Dallas, TX 75080, United States<sup>b</sup> Business School, The University of Auckland, Auckland, New Zealand

### ARTICLE INFO

#### Article history:

Received 11 January 2013

Accepted 16 January 2013

Available online 30 January 2013

#### Keywords:

Supply chain contracts

Fairness

Bounded rationality

Behavioral operations management

### ABSTRACT

Various contracts can be designed to coordinate a simple supplier–retailer channel, yet the contracts proposed in prior research and tested in a laboratory setting do not perform as standard theory predicts. The supplier, endowed with all bargaining power, can neither fully coordinate the channel nor extract all of the channel profit. We report on a sequence of laboratory experiments designed to separate possible causes of channel inefficiency. The three causes we consider are inequality aversion, bounded rationality, and incomplete information. It turns out that all three affect human behavior. Inequality aversion has by far the most explanatory power regarding retailers' behavior. Incomplete information about the retailer's degree of inequality aversion has the most explanatory power in regards to the suppliers' behavior. Bounded rationality affects both players, but is of secondary importance.

© 2013 Elsevier B.V. All rights reserved.

### 1. Introduction

The field of Supply Chain Management (SCM) draws upon a number of disciplines, such as sourcing, logistics, operations, marketing, information systems, and management (Chen and Paulraj, 2004). While each of these disciplines focuses on a separate aspect of SCM, they mostly agree in that the essence of SCM is coordination among separate independent firms. Coordination efforts focus on and derive benefits from "... information sharing, goal congruence, decision synchronization, incentive alignment, resource sharing, collaborative communication, and joint knowledge creation." (Cao and Zhang, 2011, p.61). It is well-known that contracts that fail to align incentives of independent, self-interested firms comprising a supply chain, are one of the biggest causes of suboptimal performance (Narayanan and Raman, 2004).

There is a good deal of analytical modeling literature in operations, starting with Spengler (1950), that deals with designing contracts to align incentives and coordinate channels (Cachon, 2003). The empirical evidence obtained in the laboratory tests of coordinating contracts, however, demonstrates that coordinating contracts usually fail to coordinate channels (see Katok, 2011 for a review). These studies report that participants who propose contracts (suppliers) tend to make efficient offers, but participants who respond to those offers (retailers) often reject them. It is those negotiation breakdowns that are the main cause of inefficiency in the laboratory. In fact, negotiation breakdowns are also observed in the

real world. In a well-cited example, Fisher et al. (2011) describe a negotiation for natural gas between the US and the Mexican governments that ended in the Mexicans burning off the gas rather than accepting a low-ball offer. Because negotiation breakdowns (rejections) are such a major cause of the inability of coordinating contracts to align incentives in practice, understanding their cause is an important step toward designing better-performing contracts.

We use laboratory experiments to investigate the cause of rejections in the laboratory. Our work is part of the Behavioral Operations Management (BOM) literature (Loch and Wu, 2008; Bendoly et al., 2006, 2010; Gino and Pisano, 2008). This literature has its roots in cognitive psychology (Thurstone, 1927; Simon, 1955, 1957; Kahneman and Tversky, 1979) and experimental economics (Kagel and Roth, 1995; Camerer, 2003; Bardsley et al., 2010).

Our main hypothesis is that preferences for fairness (also referred to as *inequality aversion*) are the main cause of rejections. Liu et al. (2012) identify four dimensions of fairness (or justice) relevant in supplier–buyer relationships: *distributional*, *procedural*, *interpersonal*, and *informational*. Our study focuses on the distributional aspect of fairness. Fairness has been long recognized as one of the most important factors guiding human interactions in everyday life (Adams, 1965 as well as in business Kahneman et al., 1986; Griffith et al., 2006; Kumar et al., 1998; Scheer et al., 2003). It is closely related to other-regarding preferences, such as status, altruism, reciprocity, so common in the everyday life of individuals, which also play an important role in the corporate environment. In project management, requests to share a resource tend to be accommodated even when this is counter-productive both for the person sharing the resource and the overall firm performance (Bendoly and Swink, 2007). Workers, when paid at a different rate from their peers, tend to adjust their outputs quality/quantity in

\* Corresponding author.

E-mail addresses: [ekatok@utdallas.edu](mailto:ekatok@utdallas.edu) (E. Katok), [v.pavlov@auckland.ac.nz](mailto:v.pavlov@auckland.ac.nz) (V. Pavlov).

a way that mitigates inequity between their pay and that of other workers (Goodman and Friedman, 1971). In the automobile industry, punitive behavior is not uncommon toward a supply chain partner whose actions are perceived to be unfair (Kumar et al., 1998).

Two streams of BOM literature are most closely related to our work. The first stream investigates the role of other-regarding preferences, such as fairness, on the performance of the wholesale-price contract. Cui et al. (2007) develop a model in which both parties care about fairness in a bilateral monopoly setting with a supplier and a retailer, and characterize conditions under which the supplier can coordinate the channel using wholesale pricing. Loch and Wu (2008) report on a set of laboratory experiments that show that wholesale price contract fails to coordinate the channel even after participants have been primed for cooperation. Katok et al. (2012) extend the Cui et al. (2007) model to include incomplete information. Ho et al. (2012) extend the model to a setting with multiple retailers and add peer-induced fairness.

The second literature stream investigates the performance of coordinating contracts. Ho and Zhang (2008) compare two mathematically equivalent contracts—two-part-tariff (TPT) and quantity discount contracts—and show that rejections are significantly higher under TPT. They show that a model that includes loss aversion can account for the reported treatment effect. Haruvy et al. (2012) investigate the TPT contract under different bargaining protocols, and find that a richer bargaining environment improves efficiency. Lim and Ho (2007) study 2- and 3- block tariffs and find that 3-block tariffs perform better in the lab even though in theory they should not. They attribute the treatment effect to counterfactual payoffs.

Neither of the two streams of the BOM literature we mentioned above, however, directly investigates the cause of rejections. Pavlov and Katok (2011) develop a model of coordinating contracts with fairness preferences, and their major finding is that rejections result from incomplete information about fairness preferences. Intuitively, if the supplier knows the extent to which the retailer dislikes inequality, she can offer the retailer a contract that this retailer would (just barely) accept. However, if the supplier does not know the specific retailer's preferences, some (highly inequality averse) retailers will reject the optimal contract.

The research question we address in this paper is to what extent inequality aversion, incomplete information about inequality aversion of other players, and errors (caused by factors other than fairness and incomplete information about it), exist in a laboratory contracting setting, and how they affect contract performance. Specifically, we measure the relative importance of these three factors. The main challenge we face is that the extent to which people dislike inequality and are prone to errors is, in fact, their own private information; it is part of their personality. And while there may be ways to measure some of these individual attributes (with survey instruments and hypothetical experiments, for example) these measures may well be confounded when combined with having participants play the contracting game. Therefore, we take a radically different approach, and design a unique and innovative experiment to directly get at the issue of incomplete information and error-making.

The essence of our design is to start with a treatment with two human players, use the retailers' decisions in this treatment to model their inequality aversion and propensity to make errors, and then conduct a sequence of additional treatments with automated retailers programmed to behave like their human counterparts. In these additional treatments we manipulate the extent to which retailers are prone to make random errors and, most importantly, the amount of information the supplier has about the specific retailer with whom she is matched. Neither of these experimental manipulations is possible with human retailers, thus our design

provides a clean test that we use to separate and measure the effect of behavioral factors on contract performance.

In Section 2 we present the key aspects of the basic model and formulate the research hypotheses. Section 3 details our experimental design and protocol. We present our results in Section 4, and conclude the paper with a summary and discussion in Section 5.

## 2. Model and hypotheses

### 2.1. The basic setting with full rationality

We are studying a distribution channel with a single supplier who produces units at a constant production cost of  $c$  per unit, and a single retailer. The retailer faces a linear market demand  $q = A - p$ , where  $p$  is the retail price and  $A$  is a constant. The supplier proposes a contract to the retailer, and the retailer either rejects the offer, in which case both parties earn zero profit, or places an order for  $q$  units. Since the retailer faces deterministic demand and the product has no salvage value, we assume that the retailer's order will match the amount sold, given the retail price.

We say that the channel is centralized if the outcome in terms of units produced is the same as the outcome that would have resulted from a single decision maker maximizing the entire channel profit. The channel profit to be maximized in the centralized channel is

$$\pi_C = (p - c)q = ((A - q) - c)q. \quad (1)$$

The order quantity that maximizes this channel profit is  $q^* = (A - c)/2$ , yielding the optimal (first-best) channel profit of  $(A - c)^2/4$ .

If the channel is not centralized—the two firms optimize separately and independently—we consider a wholesale-price contract in which the retailer pays the supplier  $w$  per unit and the retailer determines the order amount  $q$ . The retailer maximizes his own profit by ordering  $q_{WP}^* = (A - w)/2$  which is lower than the first-best order quantity  $q^*$  whenever  $w > c$ . The supplier must set the wholesale price so as to maximize his profit

$$\pi_S = (w - c)q_{WP}^* = \left( \frac{(w - c)(A - w)}{4} \right) \quad (2)$$

resulting in the profit-maximizing wholesale price  $w_{WP}^* = (A + c)/2$ . This optimal wholesale-price contract (with profit-maximizing players) results in the supplier's profit of  $\pi_S^{WP} = (A - c)^2/8$ , the retailer's profit  $\pi_R^{WP} = (A - c)^2/16$ , and the total channel profit  $\pi_C^{WP} = 3(A - c)^2/16$ , representing the efficiency of only 75% relative to the first-best channel profit. This inefficiency of the wholesale price contract relative to the integrated system is known as *double marginalization*.

A variety of different contracts can solve the double marginalization problem. They all, in one way or another, induce the retailer to place the first best order. The supplier then extracts some of the channel profit from the retailer. The contract on which we focus is the *minimum-order-quantity* (MOQ) contract, in which the supplier proposes a per-unit wholesale price  $w$  and a minimum order quantity  $q_{\min}$ , and the retailer either rejects the contract or orders  $q \geq q_{\min}$ . If we assume that both parties only care about maximizing their profits, the supplier should coordinate the channel by setting  $q_{\min}$  to the first-best order quantity  $q_{\min}^* = (A - c)/2$ , and then setting the wholesale price so as to extract the entire channel profit:  $w_{moq}^* = (A + c)/2$ .

There is a significant amount of laboratory evidence that contracts designed to solve the double marginalization problem do not solve it successfully. It is worth pointing out that the analysis we summarized above, which we refer to as the standard theory, critically depends on three assumptions:

1. Players care only about their profits.
2. Players are able to optimize and make no optimization errors.
3. All information is available to all players.

While it is well-understood that these three assumptions fail to hold, to what extent this failure invalidates the standard theory is not well understood. We design a set of experiments that allow us to cleanly separate and measure the effect of the failure of the three assumptions on contract performance. The three primary ways in which standard theory assumptions fail to hold that we explore are (1) fairness, (2) propensity to make random errors, and (3) incomplete information about fairness.

## 2.2. Fairness

People design and negotiate contracts, and many people care about things other than merely maximizing profits. One salient non-pecuniary motivation is the desire to be treated *fairly*, and possibly the desire to treat others fairly. The basic idea behind fairness models in behavioral economics, also termed *inequality aversion*, is that players care not only about their own profit, but also about how profits are distributed among all the players in the game. We refer the reader to two seminal papers (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) in behavioral economics that develop models based on inequality aversion, as well as an earlier paper by Bolton (1991) that was the first to note that players are primarily driven by the desire to earn at least their fair share (Cooper and Kagel, forthcoming, provide a review of inequality aversion experimental literature). Interestingly, in the channel coordination context, when fairness concerns are sufficiently strong the wholesale price contract can coordinate the channel (Cui et al., 2007).

Following the Cui et al. (2007) model, let  $\pi_R$  and  $\pi_S$  denote the retailer's and the supplier's profit, respectively, resulting from the retailer's acceptance or rejection of some contract. The retailer's utility from a contract that allocates  $\pi_R$  to the retailer and  $\pi_S$  to the supplier can be written as

$$U(\pi_R, \pi_S | \alpha, \beta) = U_R = \pi_R - \alpha [\max(\gamma\pi_S - \pi_R, 0)] - \beta [\max(\pi_R - \gamma\pi_S, 0)], \quad (3)$$

where  $\alpha \geq 0$  measures the retailer's disutility of earning less than the supplier (disadvantageous inequality),  $\beta \geq 0$  measures the retailer's disutility of earning more than the supplier (advantageous inequality) and  $\gamma$  defines what is considered to be a fair outcome. The supplier's utility is analogous to (3). Cui et al. (2007) assume full information, meaning that  $\alpha$ ,  $\beta$  and  $\gamma$  are known to both players (more on this aspect of the model in Section 2.4).

## 2.3. Errors

An alternative hypothesis that can explain retailer's rejections is that retailers do not maximize perfectly, but instead make random errors. This hypothesis follows the model in Su (2008). Options that result in higher utility are more likely to be chosen than options that result in lower utility, but are generally not chosen with certainty. If the retailer derives the utility of  $U_R$  from accepting some offer, and a utility of zero from rejecting it, then one way to model the probability that the retailer accepts the offer is

$$Pr(\text{reject}) = \frac{1}{1 + e^{\tau U_R}} \quad (4)$$

The parameter  $\tau$ , called the *precision parameter*, is used to represent the degree to which the decision maker departs from full rationality. The higher the  $\tau$ , the higher is the probability that the option with the highest utility is selected. At one extreme,  $\tau \rightarrow \infty$ , this model converges to the perfectly rational choice—the option

with the highest utility is made with certainty. At another extreme,  $\tau \rightarrow 0$ , the retailer rejects any contract 50% of the time.

## 2.4. Hypotheses separating fairness and errors

The first goal of our design is to distinguish between fairness preferences and random errors as the cause of behavioral deviations from the standard analytical model predictions. To this end we manipulate the effect of fairness concerns as follows:

1. In the MOQ treatment, the retailer and the supplier are human subjects. The supplier proposes the contract that consists of  $w$  and  $q_{\min}$ , and the retailer can either place an order  $q \geq q_{\min}$ , or reject the contract. In the event of rejection, both players earn zero.
2. The treatment we label MOQ-D, is identical to MOQ, with the single exception that if the retailer rejects the offer, the retailer earns zero, but the supplier earns  $\max[0, (w - c)q^a]$  where  $q^a$  is determined automatically as a quantity that would maximize the retailer's profit:  $q^a = \max(q_{\min}, (A - w)/2)$  as long as  $q_{\min}(A - q_{\min} - c) \geq 0$ , and  $q^a = 0$  otherwise.
3. In the treatment we label MOQ-A, the retailer is automated (the supplier knows this) and is programmed to order  $q^a$  (as defined above).

In all our experiments  $A = 100$ ,  $B = 1$ , and  $c = 20$ . We begin by formulating the null hypothesis based on the standard theory to provide a benchmark.

**Hypothesis 0.** (The Standard Theory benchmarks): In the MOQ, MOQ-D and MOQ-A treatments, a contract with  $q_{\min} = 40$  and  $w = 60$  should be offered by suppliers and always accepted by retailers, for a total channel profit of 1600 (100% efficiency). The entire profit should go to the supplier ( $\pi_S = 1600$ ,  $\pi_R = 0$ ).

The alternative hypotheses H1 and H2 state the implications of fairness and errors. We formulate the hypotheses so that rejecting a given hypotheses would allow us to rule out a particular explanation. If we fail to reject a hypothesis, we can conclude that observed behavior is consistent with a particular explanation. Of course as with any empirical results, failing to reject a hypothesis does not prove that it is true, because we cannot rule out all other potential explanations that we have not considered in this paper.

**Hypothesis 1.** (a) If retailers care about fairness, rejections will be higher in MOQ than in MOD-D. (b) If suppliers care about fairness, offers will be higher in MOQ-D than in MOQ-A.

H1(a) follows because a rejection in the MOQ treatment leads to both parties earning zero. According to (3), regardless of the values of  $\alpha$ ,  $\beta$  and  $\gamma$ , the retailer's utility from a rejection is exactly zero, which may well be higher than the retailer's utility from a very unfair offer. Thus, in the MOQ treatment, a rejection may lead to a fairer outcome. In the MOQ-D treatment, however, a rejection does not affect  $\pi_S$ , while decreasing  $\pi_R$  to zero, and this can only lead to less fair outcomes. H1(b) partially depends on an auxiliary hypothesis that when a player is automated, this automated player's profits do not enter the human player's utility function—people do not wish to treat computerized players fairly. Therefore, human suppliers who wish to be fair to a human retailer, nevertheless still prefer to extract all the profit from automated retailers. Because the only difference between the MOQ-D and MOQ-A treatments is that the retailer is human in the MOQ-D treatment, suppliers with fairness concerns will make higher offers to retailers in the MOQ-D treatment.

**Hypothesis 2.** If retailer rejections are caused by errors, rejections should be the same in the MOQ and MOQ-D treatments.

H2 is the precise counter-hypothesis to H1(a) (meaning that exactly one of them can be rejected by the data). It follows because retailers have the ability to reject offers in both MOQ and MOQ-D treatments. The difference between the two treatments is that a rejection in the MOQ treatment results in a fairer outcome, while a rejection in the MOQ-D treatment does not. Therefore, if rejections are caused only by errors, the fact that rejections do not punish suppliers in the MOQ-D treatment should not decrease them (H2). But if rejections are caused by fairness concerns (H1(a)), it will.

H2 provides a blunt test of retailers' rationality, because the MOQ-D treatment essentially asks retailers to choose between zero profit and positive profit. But the error-making explanation is in fact more nuanced, because the retailers' binary choice is between the utility of zero and the utility of  $U_R$ , defined by (3). While in the MOQ-D treatment rejections transparently lead to a lower utility, in the MOQ treatment they do not. Retailers may not be able to perfectly evaluate  $U_R$ . For example, they may be unsure of their own  $\alpha$  or of  $\gamma$  when attempting to evaluate their utility from an unfair offer, or make an error in combining those behavioral parameters to evaluate whether the utility from accepting the unfair offer is positive or not.

In the laboratory it may be possible to induce a certain  $\gamma$ , for example by manipulating parties' initial investments (Cui and Mallucci, 2012), but it is not possible to directly induce or manipulate  $\alpha$  (or  $\beta$ ) because how strongly a person feels about fairness is part of their personality. Nevertheless, a bit of introspection tells us that it may well be plausible that people experience some difficulty when faced with a choice such as comparing  $U_R$  as defined by (3), with the utility of zero. In other words, fairness concerns and making errors are not mutually exclusive. On the contrary, people may well be fully rational in a setting without fairness concerns (such as MOQ-D) and make errors in a setting with fairness concerns (MOQ).

Suppliers, faced with retailers who may care about fairness and may also make mistakes, may make more generous offers because, all else constant, a retailer who makes mistakes is more likely to accept a more generous offer.

Another issue to note has to do with whether suppliers themselves make mistakes. Suppliers may make errors in their offers, resulting in rejections even under full information and even with fully rational retailers.

2.5. Automated retailer treatments: information and errors

Analytical papers that deal with fairness assume full information (see Cui et al., 2007 model, as well as the Fehr and Schmidt, 1999 model that it extends). This assumption literally means that when making the offer to the retailer, the supplier knows this retailer's  $\alpha$  and  $\beta$  parameters. The  $\gamma$  parameter is also assumed to be common knowledge. If the retailer makes errors, the assumption is that the supplier knows the retailer's  $\tau$ . The full information assumption is likely to be wrong on the face of it in the single shot game. Without the possibility that retailers make errors, if a supplier knows her retailer's fairness parameters, she would not deliberately make an offer knowing the retailer would reject it. Therefore, in a setting with full information and no errors, there would not be any rejections. With incomplete information and the potential for errors, however, the situation is not so straightforward. So the next question we ask is to what extent is suppliers' behavior driven by retailers' fairness concerns, versus the fact that suppliers have only incomplete information about the retailers' behavioral parameters ( $\alpha$  and  $\tau$ )?

In three additional MOQ-A ("A" for automated retailers) treatments, the retailer is automated (suppliers are always human subjects; in treatments with automated retailers, they know that the retailer is automated and how it is programmed to behave). The automated retailers in the three treatments are programmed to

behave differently, and the human suppliers have varying levels of information about the precise behavior of the automated retailer in that given round. This design distinguishes the effects of incomplete information and the possibility that retailers might make errors.

In the three new MOQ treatments with automated retailers, we programmed the retailers to imitate the behavior of human retailers in the MOQ treatments. We did this by first estimating the rejection behavior of each human retailer when faced with a contract  $(w, q_{min})$  by fitting a logistic regression for each individual:  $REJECT_t = (1 + \exp(b_0 + b_1 w_t + b_2 q_{min,t}))^{-1}$ , where  $REJECT_t = 1$  if a particular contract  $(w_t, q_{min,t})$  is rejected in period  $t$  and 0 otherwise. We programmed each automated retailer to imitate a specific individual in the MOQ treatment, with the probability of rejecting any given offer  $(w, q_{min})$  equal to  $(1 + \exp(b_0 + b_1 w + b_2 q_{min}))^{-1}$ , where we estimate  $b_0, b_1$ , and  $b_2$  separately for each individual. The MOQ treatments with automated retailers programmed to be fair differ according to whether there is any noise associated with retailer decisions, and whether suppliers know the preferences of the specific retailer.

We first have a treatment with full information and no retailer errors. In that treatment, automated retailers are programmed to accept an offer  $(w, q_{min})$  whenever the probability of rejection  $(1 + \exp(b_0 + b_1 w + b_2 q_{min,t}))^{-1} < 0.5$  and reject it otherwise. We implemented full information by means of the suppliers' calculators that show them whether any specific  $(w, q_{min})$  offer will be accepted or rejected. We label this treatment MOQ-A-Full (for full information).

We next add retailer errors. We implemented the errors in this treatment by programming the retailers to reject an offer  $(w, q_{min})$  with probability  $(1 + \exp(b_0 + b_1 w + b_2 q_{min}))^{-1}$ . We implemented full information in this treatment by showing suppliers the exact probability that the computerized retailer will reject an offer  $(w, q_{min})$ . The software then rejects this offer with the given probability. We label this MOQ-A-Full-E (E for errors).

And finally, we take away full information about preferences. In the third treatment suppliers see  $(1 + \exp(b_0 + b_1 w + b_2 q_{min}))^{-1}$  for the entire group of automated retailers in the session instead of the specific retailer with whom she is matched this round. Thus, suppliers now have incomplete information about the  $\alpha$  and  $\tau$  parameters of the specific retailer with whom they are matched; we label this treatment MOQ-A-E (no "Full" indicates incomplete information).

Fig. 1 summarizes the experimental design and sample sizes. Table 1 summarizes the theoretical benchmarks for supplier decisions ( $w$  and  $q_{min}$ ), retailer decisions ( $q$ ), and contract outcomes for the four automated treatments. We compute these benchmarks based on the actual retailer types implemented in the automated treatments. These computations also assume that suppliers never make errors.

Going from left to right in Fig. 1 and Table 1, we add one behavioral feature at a time. This allows us to measure the effect of this behavioral feature. So going from MOQ-A to treatment MOQ-A-Full, we add fairness to the retailer's response, while changing nothing

Information	Full	Full	Full	Private
Fairness	No	Yes	Yes	Yes
Errors	No	No	Yes	Yes
Automated Retailer	MOQ-A <i>n</i> = 16	MOQ-A-Full <i>n</i> = 9	MOQ-A-Full-E <i>n</i> = 24	MOQ-A-E <i>n</i> = 18
Human Retailer	MOQ-D <i>n</i> = 18			MOQ <i>n</i> = 42

Fig. 1. Experimental design and sample sizes.

**Table 1**  
Theoretical benchmarks for automated treatments, assuming fully rational suppliers.

	MOQ-A	MOQ-A-Full	MOQ-A-Full-E	MOQ-A-E
$w$	60.00	52.28	50.32	47.20
$q_{\min}$	40.00	40.00	40.00	40.00
$q$	40.00	40.00	38.32	32.22
Retailer share (%)	0.00	19.31	24.21	32.00
Supplier share (%)	100.00	80.69	75.79	68.00
Rejection rate (%)	0.00	0.00	4.18	19.44
Efficiency (%)	100.00	100.00	95.82	80.56

else. We see from Table 1 that the only effect of this change that we should observe is a lower wholesale price and a higher retailer share. Rejection rate should remain zero, and efficiency should remain at 100%. Adding errors (going from MOQ-A-Full to MOQ-A-Full-E) slightly lowers wholesale price, which increases retailer share, but the main effect is that we should now observe some rejections. These rejections decrease average efficiency and average order.

Taking away full information (going from MOQ-A-Full-E to MOQ-A-E) further decreases wholesale price and increases retailer share. But an even stronger effect is that rejections should increase dramatically, resulting in a substantial drop in efficiency.

Because Table 1 benchmarks assume fully rational suppliers, we formulate a formal hypothesis about supplier rationality:

**Hypothesis 3.** (Fully rational supplier): Fully rational suppliers' decisions and resulting contract outcomes in treatments with automated retailers should not be significantly different from the benchmarks in Table 1.

Note that H3 is a very strong hypothesis, because rejecting any part of it would imply evidence that suppliers are behaving in a way that is not consistent with full rationality. While failure to reject all parts of H3 does not establish that suppliers are fully rational, it merely fails to show that they are not.

We next summarize the effect of full information and retailer errors in automated treatments, and bring the design full circle to the all-human treatment.

**Hypothesis 4.** (a) If the lack of full information about the retailers' fairness preferences plays a role, the rejection rate should be higher (and efficiency lower) in the MOQ-A-E than in the MOQ-A-Full-E treatment. (b) If both, incomplete information and errors play a role (and the supplier is fully rational), then the performance (as measured by efficiency,  $w$ ,  $q_{\min}$ , and profit distributions) in the MOQ-A-E treatment should not be different from the performance in the MOQ (all human) treatment.

H4 brings our design full circle, linking automated treatments back to the all-human treatment. By comparing contract performance (in terms of rejections, efficiency, and retailer share) in the three automated treatments and in the all human (MOQ) treatment, we will be able to measure the relative effect of three causes we are considering: fairness concerns, errors, and incomplete information.

### 3. Experimental design and protocol

In total, 127 human subjects participated in six treatments of our study. In the MOQ and MOQ-D treatments, participants were grouped into cohorts of six (three suppliers and three retailers who were matched randomly each round within a cohort).

Our participants were students at a large state university in the Northeast United States, mostly undergraduates, from a variety of majors, and they therefore represent the larger university community. We recruited them using an on line recruitment system, with cash as the only incentive offered. Earnings included a \$5

participation fee, and the rest of the earnings proportional to profits earned in the session. Paying participants based on performance is the cornerstone of experimental economics (see the seminal paper by Smith, 1976). This method has also been accepted in behavioral operations management (see for example Schweitzer and Cachon, 2000; Schultz et al., 2003; Carter and Stevens, 2007; Wu and Katok, 2006). Katok (2011) provides a comprehensive review of methodological issues in conducting laboratory experiments in behavioral operations management, including using students as subjects (also see Cantor and Macdonald, 2009; Narasimhan et al., 2009 for examples of articles that use students in laboratory experiments; see Bolton et al., 2012 for a study comparing professional managers and students).

Average earnings, including a participation fee of \$5, were \$25, but differences in supplier and retailer earnings were substantial. In the MOQ-D treatment (by design), retailers earned not significantly more than \$5 and suppliers earned approximately \$45. All sessions took place at a dedicated experimental laboratory in the college of business during the fall semester of 2007, spring semester of 2008, and summer semester of 2009.

Upon their arrival at the laboratory, participants were seated in visually isolated cubicles and read written instructions (see the Appendix) describing the rules of the game. After all participants finished reading the instructions, we read the instructions to them aloud, to ensure their common knowledge about the rules of the game. We also answered any questions, prior to the start of the game.

We programmed the computer interface using the zTree system (Fischbacher, 2007). Suppliers had a calculator on their screens that computed and displayed, for any combination of  $q_{\min}$  and  $w$ , the retailer's profit-maximizing order and the resulting earnings for both players. In the treatments with automated retailers, the calculator also displayed, depending on the treatment, corresponding information about the likelihood of the specific contract being accepted. Suppliers could try multiple parameters before transmitting their offer to the retailer. Retailers in the MOQ treatment had access to a calculator that computed, for any  $q$  they entered, the resulting earnings for both players. Retailers in this treatment could try any number of  $q$ 's before settling on their final decision. They also had a "Reject" button that resulted in earnings of 0 for both players. Retailers in the MOQ-D treatment observed the  $q$  that the computer entered on their behalf and had "Accept" and "Reject" buttons. The "Reject" button did not affect the supplier's earnings but resulted in 0 earnings for the retailer.

## 4. Experimental results

### 4.1. Comparisons with theoretical benchmarks: Hypothesis 0

We summarize averages and standard deviations of the decisions and outcomes in Table 2.

The comparison of the contract decisions and outcomes with standard theoretical benchmarks (Hypothesis 0) in the MOQ, MOQ-D, and MOQ-A treatments produces the results indicated by \*\*\* in Table 2. The  $p$ -values we report here and elsewhere for one-sample tests are from the Wilcoxon signed-rank test, and those for two-sample tests are from the Wilcoxon rank-sum (Mann-Whitney) test.<sup>1</sup> The unit of analysis is the average for an individual subject.

Contract outcomes in the MOQ-A and MOQ-D treatments are generally very close to theoretical predictions (supporting H0). Outcomes of the MOQ treatment, however, are significantly different

<sup>1</sup> We compute average values for each individual and use the individual subject as the unit of analysis.

**Table 2**  
Mean values of decisions and outcomes (standard deviations in parenthesis).

Decision	MOQ	MOQ-D	MOQ-A	MOQ-A-Full	MOQ-A-Full-E	MOQ-A-E
$w$	50.49** (6.64)	58.13 (5.27)	57.75 (4.87)	52.05 (3.63)	51.64 (3.46)	50.19 (5.23)
$q_{min}$	37.03** (8.65)	39.58 (2.27)	40.41 (3.87)	39.65 (1.84)	38.65 (3.33)	37.61 (8.28)
$q$	24.35** (7.71)	39.24 (2.49)	40.27 (3.60)	39.55 (1.84)	35.22 (3.68)	27.14 (8.67)
Supplier share demanded (%)	72.44** (10.05)	93.92** (10.24)	95.77** (9.75)	79.35 (8.09)	76.79 (8.95)	72.37 (7.04)
Rejection Rate (%)	19.52**	0.56	0.47	0.28	9.17	29.44
Conditional on acceptance Efficiency (%)	92.30** (8.11)	98.52 (2.01)	97.84 (4.72)	98.43 (0.75)	98.53 (2.57)	96.07 (4.46)
$\pi_S$	1011.47** (176.55)	1480.47 (163.93)	1503.28 (189.43)	1250.06 (129.70)	1218.28 (155.40)	1091.68 (134.24)
Entire sample Efficiency (%)	74.87** (17.47)	98.00** (1.97)	97.00** (5.16)	98.00 (0.90)	90.00 (7.82)	68.00 (19.99)
$\pi_S$	818.17** (215.65)	1,479.47** (163.22)	1,496.57** (191.59)	1,246.90 (133.06)	1,108.64 (186.35)	771.39 (231.78)

\*\* Different from Hypothesis 0 benchmarks,  $p < 0.05$ .

in that rejections are higher, efficiency is lower, wholesale prices are lower, and supplier profit is lower.

Fig. 2 plots the average efficiency and supplier profit over time in the six treatments. It is clear from the figure that in the MOQ-D and MOQ-A, efficiency levels quickly reach their theoretical benchmarks after a few initial periods, whereas in the MOQ treatment, efficiency and supplier profit remain low. If we measure efficiency and supplier profit at the end of the session, efficiency is not significantly different from 100% and supplier profit is not significantly different from 1600 in the MOQ-D and MOQ-A treatments, indicating that there is some learning that occurs in those treatments.<sup>2</sup> There is also some learning that occurs in the MOQ treatment, in that wholesale prices and rejections decrease over time, and efficiency increases over time. The supplier profit, however, shows no significant time trend.

Overall, the data from the MOQ-A and MOQ-D treatments mostly falls in line with H0. Suppliers in those treatments can figure out how to coordinate their channels and extract virtually all the profit. Where the standard theory completely fails is the MOQ treatment; even though suppliers understand how to coordinate the channel (they set  $q_{min}$  close to the optimal level of 40), rejections remain high, destroying efficiency, so the channel is not coordinated. In the end, suppliers extract only about 70% of the channel profit. This pattern is consistent with the explanation that retailers demand fairness for themselves, and this motive is most clear in the MOQ treatment. We will present additional analyses of the retailer motives in the next section.

#### 4.2. Fairness and errors: Hypotheses 1 and 2

Retailers can reject contracts in MOQ and MOQ-D treatments. But in the MOQ treatment rejections punish the supplier, while in the MOQ-D treatment they do not. In the MOQ treatment, retailers reject 19.52% of offers, while in the MOQ-D treatment only 0.56% of offers are rejected (in fact, only 2 non-zero offers were rejected). This data is line with H1(a), and allows us to rule out errors as the cause of rejections (reject H2).

Is there any evidence that suppliers try to be fair to retailers? To answer this question we compare retailer profits in the MOQ-A

and MOQ-D treatments. In the MOQ-D treatment retailers earn, on average, 95.60 tokens, while in the MOQ-A treatment they earn on average 62.82 tokens. These retailer profits are not significantly different ( $p = 0.201$ ). However, another way to look at this is to compare the frequency of positive offers in MOQ-A and MOQ-D treatments. In the MOQ-D treatment, around half of the offers result in the retailer earning a small but positive profit, whereas in the MOQ-A treatment, this figure is less than 8%. To show this difference more formally, we run a logit regression on the MOQ-A and MOQ-D treatment data, with the dependent variable equal to 1 when the offer is non-zero and 0 otherwise. The independent variable is the indicator variable for the MOQ-D treatment. The coefficient for the MOQ-D treatment is positive and significant ( $p = 0.014$ ), which indicates that suppliers are more likely to make non-zero offers to human than to automated retailers. This evidence is in line with H1(b), although suppliers' altruism is quite modest.

#### 4.3. Suppliers' rationality: Hypothesis 3

Next we examine suppliers' behavior. In the MOQ-A-Full treatment, suppliers know whether the automated retailer will accept any specific offer. Not surprisingly, rejections are almost non-existent in this treatment (0.28%), and efficiency is nearly 100% after learning in the initial rounds occurs. We can conclude that when suppliers have full information about the retailers' behavior, they are fully capable of maximizing their profit. This is in line with H3.

The automated treatments also allow us to check to what extent suppliers are rational by comparing average contract parameters ( $w$  and  $q_{min}$ ) and the resulting rejection rates in Table 2 to the corresponding benchmarks in Table 1. In the MOQ-A-Full treatment none of the average values is significantly different from their benchmarks. But when retailers make errors, the problem becomes more difficult for suppliers: in the MOQ-A-Full-E treatment the average  $q_{min}$  is slightly lower than 40 ( $p = 0.0049$ ), the average  $w$  is not different from 50.32 ( $p = 0.1300$ ) and the rejection rate is significantly higher than 4.18% ( $p = 0.0148$ ). The supplier's problem becomes still more difficult when they do not have full information: in the MOQ-A-E treatment the average  $w$  is significantly higher than 47.20 ( $p = 0.0347$ ) and the average rejection rate is weakly significantly higher than 19.44% ( $p = 0.0706$ ), although the average  $q_{min}$  is not significantly different from 40. Overall, there is some evidence that suppliers make errors (rejecting H4) because suppliers are more likely to deviate from fully rational benchmarks when their tasks are more difficult (MOQ-A-Full-E and MOQ-A-E treatments) than when their tasks are very straightforward (MOQ-A-Full and MOQ-A treatments).

<sup>2</sup> A simple way to measure the value of a variable  $Y_{it}$  at the end of the session (where  $i$  is an individual participant and  $t$  is period number) is to fit a linear regression model  $Y_{it} = \text{End} + \beta_T - (40 - \text{Periodit}) + \eta_i + \varepsilon_{it}$  (with random effects for individuals), and test whether the intercept term End is significantly different from its benchmark. We base our conclusions about the values of efficiency and supplier profit levels at the end of the session on this method.

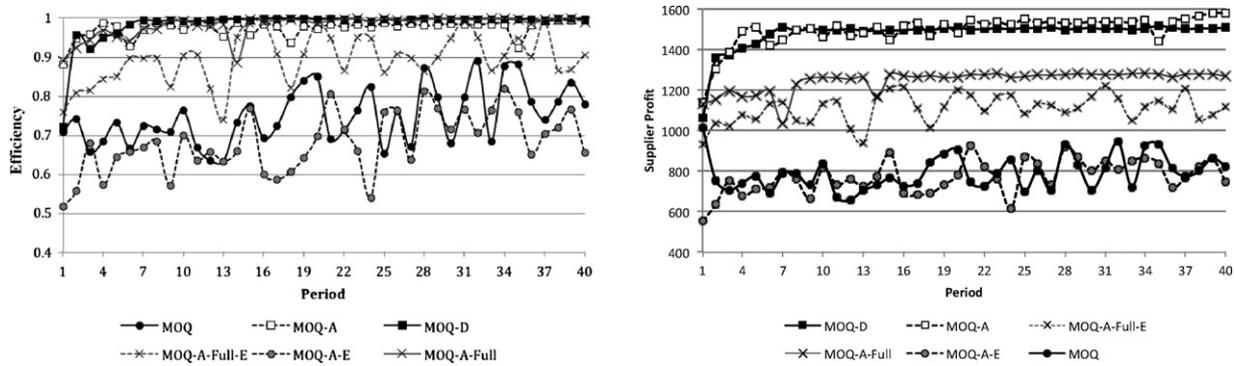


Fig. 2. Average efficiency and supplier profit over time.

4.4. Full information: Hypothesis 4

How do errors interact with the lack of full information? Treatments MOQ-A-E and MOQ-A-Full-E differ only in whether suppliers have access to full information about their retailer's preferences. Indeed, rejections in the MOQ-A-E treatment are 29.44%, which is significantly higher than the rejections of 9.17% in the MOQ-A-Full-E treatment ( $p = 0.0003$ ). This is consistent with H4(a). There are no significant differences in terms of average contract parameters, players' profits, or efficiency between the MOQ and MOQ-A-E treatment (all  $p$ -value are above 0.05), consistent with H4(b).

Our result is consistent with the explanation that incomplete information about fairness, combined with the retailer's propensity to make errors, accounts for the observed behavior in the MOQ treatment. Of course there may be other potential explanations that we did not consider in this paper. Failure to reject a hypothesis is not a proof that the hypothesis is true.

4.5. Discussion

In Fig. 3, we classify contracting arrangements in each treatment into five categories:

- The "50-50" category is when retailers earn at least 40% of the total channel profit.
- The "Retailer High" category is when retailers earn between 20% and 40% of the channel profit.
- The "Retailer Low" category is when retailers earn less than 20% of the profit.

- The "Retailer Zero" category is when the supplier offers exactly 0% of the channel profit to the retailer.
- The "Reject" category corresponds to retailer rejections.

In the first two categories (represented by light colors in the figure), the retailer does relatively well. In the last three categories (represented by dark colors), the retailer does poorly.

Fig. 3 thus reveals three key observations. First, in the treatments in which the retailer can punish the supplier with a rejection and there is incomplete information, virtually all accepted offers occur in the first two categories. Thus, in MOQ and MOQ-A-E, we observe a large number of rejections, but accepted offers still fall mostly into the top two categories. In contrast, in the MOQ-A and MOQ-D treatments, where the retailer either is programmed not to reject any non-negative offer (MOQ-A) or retailer rejection only results in the retailer earning 0, the retailers' profit share is much smaller and, at the same time, the rejection rate for positive offers is at essentially zero.

Second, providing full information about fairness allows suppliers to extract more channel profit and offer retailers the minimum amount they are willing to accept. This finding becomes manifest in the increase in small positive offers that are accepted ("Retailer Low" category). Some retailers are not very demanding, whereas others are. When suppliers do not know the type, they must make fairly generous offers across the board. But when suppliers know the type, they do not offer retailers much more than the minimum that retailers are willing to accept. Therefore, approximately 50% of accepted offers in the MOQ-A-Full and MOQ-A-Full-E treatments are quite small and not rejected, while the corresponding figure is less than 10% in the MOQ-A-E treatment.

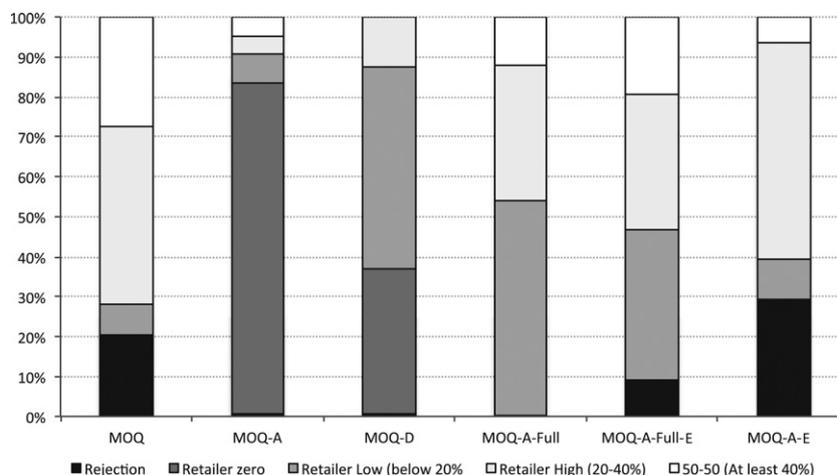


Fig. 3. Contract outcomes in all treatments.

Third, we can see graphically the extent of suppliers' altruism: even though in absolute terms retailer earnings are not significantly higher in the MOQ-D than in the MOQ-A treatment, suppliers are much more likely to offer small positive profits (as opposed to zero profits) to human retailers. More than 60% of offers in the MOQ-D treatment are positive, while this figure is less than 20% in the MOQ-A treatment. Supplier behavior in our MOQ-D treatment is roughly in line with the results of the Dictator Game (Bolton et al., 1998).

## 5. Conclusion

This research investigates the major factor contributing to poor performance of supply chain contracts observed in experimental studies, namely rejections. We report on a sequence of laboratory experiments designed to separate possible causes of channel inefficiency. The three causes we consider are inequality aversion, errors, and incomplete information. It turns out that all three affect human behavior. Inequality aversion has by far the most explanatory power regarding retailers' behavior. Incomplete information about the retailer's degree of inequality aversion has the most explanatory power in regards to the suppliers' behavior. Errors affect both players, but are of secondary importance.

### 5.1. Contributions to research

Channel coordination has been widely recognized as an important problem in supply chain management, and analytical literature has focused on deriving mechanisms to achieve coordination, namely coordinating contracts. Under the assumptions of standard analytical models, the more powerful firm should be able to use a coordinating contract to achieve 100% channel efficiency and extract all of the channel profit. However, when these contracts are tested in the laboratory, the results reveal several systematic deviations from the predictions of standard theory. First, contract efficiency is significantly lower than 100%. Second, a contract's poor performance is primarily due to rejections that the standard theory cannot explain. Third, parties tend to split the channel profit closer to 50–50 than to 100–0. These discrepancies call for further exploration of contracting in supply chains. Our study advances understanding of how fairness, errors, and incomplete information, affect contract performance.

Specifically, we report on a set of controlled laboratory experiments that we designed to cleanly separate and measure the effect of different behavioral factors. Our innovative design, in part, uses automated retailers in order to cleanly control and manipulate the information and the extent to which retailers make errors. We find that our data is consistent with the model that states that retailers care deeply about fairness, especially in terms of demanding their own fair share (see also Bolton, 1991). An important goal of our experimental design is to assess separately the impact of fairness and errors on contract outcomes; we find that as far as retailers go, fairness concerns have a qualitatively larger effect than do errors, though both are important for organizing the data. However, as far as suppliers go, we find minimal evidence of fairness concerns (which in the context of this game is altruism). The primary driver consistent with suppliers' behavior we observe is incomplete information about retailers' preferences for fairness. We also find evidence for the effect of errors, but qualitatively, the effect is relatively small.

### 5.2. Practical implications

In the modern global economy competition takes place between complex supply chains that sometimes include thousands of

companies spread across all over the world. For many of them successful channel coordination is not a question of academic interest but rather a prerequisite of their survival. The importance of our study for practitioners is two-fold. First, it uncovers that incomplete information about preferences is an extremely influential factor diminishing competitiveness of a supply chain. Second, it identifies information about preferences for fairness as potentially the most important for the efficiency improvement. There seem to be different ways in how practitioners can make use of our findings, from taking the fact of incomplete information into account when designing contracts to, perhaps, trying to eliminate incompletes of information about preferences for fairness by enforcing procedural fairness.

### 5.3. Study limitations

Although we believe that our results inform both theory and practice helping better understand the fundamental factors underlying the problem of supply chain coordination, by no means should they be considered absolute. The laboratory experiments, as a methodology, features high internal validity because of the strong control over the environment (Roth, 1988; Smith, 1994). However, the same strong control almost inevitably eliminates many features of the real-life situation that motivated the research and may limit the extent to which findings generalize (Harrison and List, 2004). One clear reason why our findings need not apply to the real-life supply chains is that in our experiment we used undergraduate students from a variety of majors whereas in real-life supply chains contracts are typically negotiated by experienced procurement/sales managers. Although there is evidence that decisions made in simple games by professional managers are qualitatively the same as decisions made by undergraduate students (Bolton et al., 2012) the extent to which laboratory experiments that deal with contracting problems generalize should be a subject of future research.

### 5.4. Future research

The main implication of our work is that contracts that are coordinating in theory may not actually coordinate the channel. The reason for this failure is the presence of incomplete information in the game. To the extent that the bargaining process can help lessen incomplete information, the way theorists model and experimentalists implement bargaining may have a strong effect on contract performance. Supply chain coordination literature usually does not include models of bargaining, but the existing evidence suggests that bargaining process can make substantial impact on performance (Radner and Schotter, 1989), possibly by establishing procedural fairness (Jambulingam et al., 2009). Therefore, one fruitful direction for future research will be to start developing a better understanding of how to incorporate it. Haruvy et al. (2012) is one study that reports on some initial steps in that regard.

## Acknowledgements

Katok gratefully acknowledges the support of the National Science Foundation (SES 0849054) and thanks Axel Ockenfels and the Deutsche Forschungsgemeinschaft for financial support.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jom.2013.01.001>.

## References

- Adams, J., 1965. Inequity in social exchange. *Advances in experimental social psychology* 2, 267–299.
- Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., Sugden, R., 2010. *Experimental Economics: Rethinking the Rules*. Princeton University Press, Princeton, NJ.
- Bendoly, E., Croson, R., Goncalves, P., Schultz, K., 2010. Bodies of knowledge for research in behavioral operations. *Production and Operations Management* 19 (4), 434–453.
- Bendoly, E., Donohue, K., Schultz, K.L., 2006. Behavior in operations management: assessing recent findings and revisiting old assumptions. *Journal of Operations Management* 24 (6), 737–752.
- Bendoly, E., Swink, M., 2007. Moderating effects of information access on project management behavior, performance and perceptions. *Journal of Operations Management* 25 (3), 604–622.
- Bolton, G.E., 1991. A comparative model of bargaining: theory and evidence. *The American Economic Review* 81 (5), 1096–1136.
- Bolton, G.E., Katok, E., Zwick, R., 1998. Dictator game giving: rules of fairness versus acts of kindness. *International Journal of Game Theory* 27 (2), 269–299.
- Bolton, G.E., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity, and competition. *The American Economic Review* 90 (1), 166–193.
- Bolton, G.E., Ockenfels, A., Thonemann, U.W., 2012. Managers and students as newsvendors. In: *Management Science*, Available at: <http://mansci.journal.informs.org/content/early/2012/07/18/mnsc.1120.1550> (accessed 7.11.12).
- Cachon, G.P., 2003. Supply chain coordination with contracts. In: *Supply Chain Management: Design Coordination and Operation*. Elsevier, Available at: <http://www.sciencedirect.com/science/article/pii/S0927050703110067> (accessed 16.05.12), pp. 227–339.
- Cantor, D.E., Macdonald, J.R., 2009. Decision-making in the supply chain: examining problem solving approaches and information availability. *Journal of Operations Management* 27 (3), 220–232.
- Cao, M., Zhang, Q., 2011. Supply chain collaboration: impact on collaborative advantage and firm performance. *Journal of Operations Management* 29 (3), 163–180.
- Camerer, C.F., 2003. *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton University Press, Princeton, NJ.
- Carter, C.R., Stevens, C.K., 2007. Electronic reverse auction configuration and its impact on buyer price and supplier perceptions of opportunism: a laboratory experiment. *Journal of Operations Management* 25 (5), 1035–1054.
- Chen, I.J., Paulraj, A., 2004. Towards a theory of supply chain management: the constructs and measurements. *Journal of Operations Management* 22 (2), 119–150.
- Cooper, D.J., Kagel, J.H., forthcoming. Other-regarding preferences: a selective survey of experimental results. In: Kagel, J.H., Roth, A.E. (Eds.), *The Handbook of Experimental Economics*. Princeton University Press, Available at: [http://www.econ.ohio-state.edu/kagel/Other\\_RegardingPreferences\\_survey.pdf](http://www.econ.ohio-state.edu/kagel/Other_RegardingPreferences_survey.pdf)
- Cui, T.H., Mallucci, P., 2012. Fairness Ideals in Distribution Channels. Working Paper. Available at SSRN: <http://ssrn.com/abstract=2021248> or <http://dx.doi.org/10.2139/ssrn.2021248>
- Cui, T.H., Raju, J.S., Zhang, Z.J., 2007. Fairness and channel coordination. *Management Science* 53 (8), 1303–1314.
- Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics* 114 (3), 817–868.
- Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2), 171–178.
- Fisher, R., Ury, W., Patton, B., 2011. *Getting to Yes: Negotiating Agreement Without Giving In*. Penguin Books, New York, NY.
- Gino, F., Pisano, G., 2008. Toward a theory of behavioral operations. *Manufacturing Service Operations Management* 10 (4), 676–691.
- Goodman, P.S., Friedman, A., 1971. An examination of Adams' theory of inequity. *Administrative Science Quarterly* 16 (3), 271–288.
- Griffith, D.A., Harvey, M.G., Lusch, R.F., 2006. Social exchange in supply chain relationships: the resulting benefits of procedural and distributive justice. *Journal of Operations Management* 24 (2), 85–98.
- Harrison, G.W., List, J.A., 2004. Field experiments. *Journal of Economic Literature* 42 (4), 1009–1055.
- Haruy, E., Katok, E., Pavlov, V., 2012. Can Coordinating Contracts Improve Channel Efficiency? Working paper.
- Ho, T.H., Su, X., Wu, Y., 2012. Distributional and peer-induced fairness in supply chain contract design. Working Paper.
- Ho, T.-H., Zhang, J., 2008. Designing pricing contracts for boundedly rational customers: does the framing of the fixed fee matter? *Management Science* 54 (4), 686–700.
- Jambulingam, T., Kathuria, R., Nevin, J.R., 2009. How fairness garners loyalty in the pharmaceutical supply chain: role of trust in the wholesaler-pharmacy relationship. *International Journal of Pharmaceutical and Healthcare Marketing* 3 (4), 305–322.
- Kagel, J.H., Roth, A.E. (Eds.), 1995. *The Handbook of Experimental Economics*. Princeton University Press, Princeton NJ.
- Kahneman, D., Knetsch, J.L., Thaler, R., 1986. Fairness as a constraint on profit seeking: entitlements in the market. *The American Economic Review* 76 (4), 728–741.
- Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47 (2), 263–292.
- Katok, E., 2011. Laboratory experiments in operations management. In: Geunes, J. (Ed.), *Tutorials in Operations Research*. INFORMS, Hanover, MD, pp. 15–35, Available at: <http://www.informs.org/Pubs/Tutorials-in-OR/2011-Tutorials-in-Operations-Research-ONLINE/Chapter-2> (accessed 6.11.12).
- Katok, E., Olsen, T., Pavlov, V., 2012. Wholesale pricing under mild and privately known concerns for fairness. *Production and Operations Management*, <http://dx.doi.org/10.1111/j.1937-5956.2012.01388.x>.
- Kumar, N., Scheer, L.K., Steenkamp, J.-B.E.M., 1998. Interdependence, punitive capability, and the reciprocation of punitive actions in channel relationships. *Journal of Marketing Research* 35 (2), 225–235.
- Lim, N., Ho, T.H., 2007. Designing price contracts for boundedly rational customers: does the number of blocks matter? *Marketing Science* 26 (3), 312–326.
- Liu, Y., Huang, Y., Luo, Y., Zhao, Y., 2012. How does justice matter in achieving buyer–supplier relationship performance? *Journal of Operations Management* 30 (5), 355–367.
- Loch, C., Wu, Y., 2008. Social preferences and supply chain performance: an experimental study. *Management Science* 54 (11), 1835–1849, <http://dx.doi.org/10.1287/mnsc.1080.0910>.
- Narasimhan, R., Nair, A., Griffith, D.A., Arlbjørn, J.S., Bendoly, E., 2009. Lock-in situations in supply chains: a social exchange theoretic study of sourcing arrangements in buyer–supplier relationships. *Journal of Operations Management* 27 (5), 374–389.
- Narayanan, V.G., Raman, A., 2004. Aligning incentives in supply chains. *Harvard Business Review* 82 (11), 94–102.
- Pavlov, V., Katok, E., 2011. Fairness and coordination failures in supply chain contracts. Working Paper.
- Radner, R., Schotter, A., 1989. The sealed-bid mechanism: an experimental study. *Journal of Economic Theory* 48 (1), 179–220.
- Roth, A.E., 1988. Laboratory experimentation in economics: a methodological overview. *The Economic Journal* 98 (393), 974–1031.
- Scheer, L.K., Kumar, N., Steenkamp, J.-B.E.M., 2003. Reactions to perceived inequity in U.S. and Dutch interorganizational relationships. *The Academy of Management Journal* 46 (3), 303–316.
- Schultz, K.L., McClain, J.O., Thomas, L.J., 2003. Overcoming the dark side of worker flexibility. *Journal of Operations Management* 21 (1), 81–92.
- Schweitzer, M.E., Cachon, G.P., 2000. Decision bias in the newsvendor problem with a known demand distribution: experimental evidence. *Management Science* 46 (3), 404–420.
- Simon, H.A., 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics* 69 (1), 99–118.
- Simon, H.A., 1957. *Models of Man: Social and Rational; Mathematical Essays on Rational Human Behavior in Society Setting*. Wiley, New York.
- Smith, V.L., 1976. Experimental economics: induced value theory. *American Economic Review* 66 (2), 274–279.
- Smith, V.L., 1994. Economics in the Laboratory. *The Journal of Economic Perspectives* 8 (1), 113–131.
- Spengler, J.J., 1950. Vertical integration and antitrust policy. *The Journal of Political Economy* 58 (4), 347–352.
- Su, X., 2008. Bounded rationality in newsvendor models. *Manufacturing & Service Operations Management* 10 (4), 566–589.
- Thurstone, L.L., 1927. A law of comparative judgment. *Psychological Review* 34 (4), 273–286.
- Wu, D.Y., Katok, E., 2006. Learning, communication, and the bullwhip effect. *Journal of Operations Management* 24 (6), 839–850.