POMS PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT POMS

Vol. 23, No. 2, February 2014, pp. 269–284
ISSN 1059-1478 | EISSN 1937-5956 | 14 | 2302 | 0269

DOI 10.1111/poms.12067

Complexity as a Contract Design Factor: A Human-to-Human Experimental Study

Basak Kalkancı
Scheller College of Business, Georgia Institute of Technology, Atlanta, Georgia 30308, USA, basak.kalkanci@scheller.gatech.edu

Kay-Yut Chen
HP Laboratories, Hewlett-Packard Company, Palo Alto, California 94304, USA, kychen@hpl.hp.com

Feryal Erhun
Department of Management Science and Engineering, Stanford University, Stanford, California 94305-4026, USA, ferhun@stanford.edu

Despite being theoretically suboptimal, simpler contracts (such as price-only contracts and quantity discount contracts with limited number of price blocks) are commonly preferred in practice. Thus, exploring the tension between theory and practice regarding complexity and performance in contract design is especially relevant. Using human subject experiments, Kalkancı et al. (2011) showed that such simpler contracts perform effectively for a supplier interacting with a computerized buyer under asymmetric demand information. We use a similar set of experiments with the modification that a human supplier interacts with a human buyer. We show that human interactions strengthen the supplier's preference for simpler contracts. We find that suppliers have fairness concerns even when they interact with computerized buyers. These fairness concerns tend to be even stronger when suppliers interact with human buyers, particularly when the complexity of the contract is low. We also find that suppliers are more prone to random decision errors (i.e., bounded rationality) when interacting with human buyers. In the absence of social preferences, Kalkancı et al. identified reinforcement and bounded rationality as key biases that impact suppliers' decisions. In human-to-human experiments, we find evidence for social preference effects. However, these effects may be secondary to bounded rationality.

Key words: human-to-human interactions; all-unit quantity discount contracts; price-only contracts; supply chain efficiency; contract performance
History: Received: February 2011, Accepted: December 2012 by Elena Katok, after 2 revisions.

1. Introduction

Recent research has been updating the traditional contract design's paradigm of assuming self-interested, expected-profit-maximizing individuals to reflect the reality of human behavior. For example, Becker-Peth et al. (2013) calibrate parameters of an optimal buyback contract to reflect the behavior of agents; they extrapolate contract terms with experimental responses to the contract. Lim and Ho (2007) observe that increasing the number of price blocks beyond two in quantity discount contracts (while keeping the number of parameters determined by subjects the same by fixing some of the decisions to their optimal values) continue to increase supply chain efficiency in a complete information setting. This is contrary to theory, which predicts that all benefits should accrue when the number of price blocks increases from one to two and that no additional benefits should be observed from increasing the number of blocks further. Kalkancı et al. (2011) study complexity as a design dimension with human subject experiments in a setting of a supplier–buyer supply chain under asymmetric demand information and conclude that the notion that complex contracts can optimize the supplier's profit is flawed. These studies, among others, argue that even though human subjects' decisions do not perfectly mimic self-interested, expected-profit-maximizing theory, their decisions can be explained and predicted, and contracts can be designed accordingly.

We follow Kalkancı et al.'s (2011) approach to the contract design problem, which differs from most other behavioral studies. Instead of designing optimal contracts for "realistic" agents, the authors study principals while keeping the agent rational. Their premise is that the "mechanism designer" is only an adviser to the "true" principal.[1] Theoretically, the mechanism designer is the principal. In practice, however, contract structure and contract parameters are determined by different individuals. Furthermore, the involvement of human decision makers is often unavoidable because crucial information, such as reactions of downstream channel partners on the receiving end of the contract, may not be available anywhere other than from the knowledge and judgment of managers. As such, understanding decision biases of these human principals is important.

In a setting with a single supplier and a single buyer, Kalkancı et al. characterize the impact of contract complexity and asymmetric information on performance. In their experiments, the computerized buyer (who maximizes her expected profits) faces a newsvendor setting and has better information on end-consumer demand than the human supplier. The supplier offers either a quantity discount contract (with two or three price blocks) or a price-only contract: contracts that are commonplace in practice (e.g., in pharmaceuticals, Cui et al. 2007; components for airplanes, Boeing 2008; furniture industry, Guo et al. 2010; and electronics, Kayış et al. 2012), yet different in complexity. The authors experimentally determine the appropriate level of complexity, measured in the number of price blocks. They show that, contrary to what traditional contract design theory predicts, simpler contracts, either a price-only contract or a quantity discount contract with a low number of price blocks, perform very well under asymmetric demand information.

With their human-supplier, computerized-buyer setting, Kalkancı et al. minimize the effects of social preferences, such as fairness, and focus on the complexity-induced decision biases of the supplier. In this study, we extend their research by considering the interaction of a human supplier with a human buyer to recapture the effects of social preferences. We observe that the main conclusions of human-to-computer experiments remain valid in this setting. Moreover, we demonstrate that human-to-human interactions strengthen the preference for simpler contracts; even the price-only contract performs effectively under human-to-human interactions. We also observe that profit splits between suppliers and buyers are not more equitable in human-to-human experiments. Even though we find evidence for social preference effects in our experiments, this result suggests that such effects may play a secondary role to complexity.

In addition to providing validation for Kalkancı et al.'s human-to-computer experiments, human-to-human experiments further our understanding of when automating some players in an experiment makes a difference and when it does not. First, we find that suppliers have fairness concerns even when they interact with computerized buyers. These fairness concerns tend to be even stronger when suppliers interact with human buyers, particularly when the complexity of the contract is low. We also find that suppliers are more prone to random decision errors (i.e., bounded rationality) when interacting with human buyers.

The rest of this study is organized as follows. In section 2, we summarize the related theoretical and behavioral literature. We provide the model definition in section 3.1, the theoretical predictions in section 3.2, and the experimental design in section 3.3. Section 4 details our methodology and displays the consequent results. We discuss our behavioral observations in section 5 and conclude the study in section 6.

2. Literature Review

There are two streams of literature relevant to this research, which we briefly review below. For a review of the theoretical literature on all-unit quantity discounts, we refer readers to Benton and Park (1996) and Munson and Rosenblatt (1998), as well as to Altintas et al. (2008) and the references therein.

2.1. Behavioral Literature on Contracts

Keser and Paleologo (2004) study the price-only contract in the interaction of a newsvendor buyer with a supplier. The authors conclude that even though the inefficiency of the price-only contract is not significantly different from theoretical predictions, the behaviors of the players are. Katok and Wu (2009) further extend this line of research by comparing the performance of a simple price-only contract with more complex buyback and revenue-sharing contracts. The authors automated one of the decision makers to eliminate the effect of social preferences. Their experiments show that complex contracts improve the supply chain performance significantly compared with a price-only contract; however, the improvement is less than what is theoretically predicted.

The behavioral literature on quantity discount contracts is still developing. Lim and Ho (2007) study the optimal number of price blocks under complete information where demand is deterministic and characterized by a linear-inverse function. In this case, theory predicts that all benefits should accrue when the number of price blocks increases from one to two and that no additional benefits should be observed from increasing the number of blocks further. However, the authors observe that benefits continue to increase in the number of blocks, as subjects cannot coordinate the supply chain with only two price blocks. Even though Lim and Ho study quantity discount contracts, their focus is not on the issue of complexity. Their subjects make exactly the same number of decisions under different numbers of price blocks, as the authors fix some of the decisions to their optimal values. In addition, Lim and Ho study a situation with complete information even though almost all contracting situations in practice exhibit asymmetric information. Indeed our results show that commonly observed results in the behavioral literature under complete information may no longer hold under asymmetric information, thereby validating our focus.

Ho and Zhang (2008) question the role of framing by comparing a two-part tariff with an all-unit quantity discount in a setting similar to that of Lim and Ho. The authors study contracts with different levels of complexity, including the single-price contract, the two-part tariff, and the quantity discount contract. In their setting, there is one contract decision in the single-price contract and two decisions in both the two-part tariff and the quantity discount contract. The focus of Ho and Zhang is the role of the fixed fee (two-part tariff) and its framing (quantity discount). They show that the fixed fee fails to increase the efficiency of the channel when framed as a two-part tariff, but achieves a higher efficiency when framed as a quantity discount. This is contrary to theory, which predicts that these two mechanisms should be equal. Thus, the authors conclude that framing matters when designing supply contracts. Even though Ho and Zhang examine the issue of contract complexity, there are substantial differences between their setting and ours. There is no demand uncertainty in Ho and Zhang, and the authors do not include asymmetric information in their setting. When the supplier has complete information and there is no demand uncertainty, he can simply design a quantity discount contract with two prices (which essentially acts as a two-part tariff) and can extract the entire supply chain profit. Thus, the supplier does not have any reason to offer a contract more complex than a quantity discount contract with two prices. However, in the presence of information asymmetry and demand uncertainty, this is no longer the case. The supplier can effectively utilize a quantity discount contract with three prices (by taking into account the buyer's quantity choice under different private demand conditions she might have) and even then he has to leave some profit to the buyer. Given the substantial differences in the basic setting and the philosophy of treatment design between Ho and Zhang and this study, these two studies complement each other.

2.2. Literature of Task Complexity

The notion of complexity is not new and has been a subject of studies in many fields. With his seminal work, Simon suggests that "when faced with complexity beyond his ken" a human subject "finds ways of action that are sufficient unto the day" (Simon 1978); that is, he relies on simpler decision rules (Simon 1955). For example, unless there is a compelling reason to do so, humans tend not to change their established behavior (Samuelson and Zeckhauser 1988, Simon 1978). Similarly, humans have a tendency to weigh recent decisions more heavily than earlier decisions, especially when the complexity of decisions is high (Hogarth and Einhorn 1992). Such behavioral biases clearly indicate that when faced with complex decisions, humans act boundedly rational. A boundedly rational decision maker lacks the propensity to optimize; although he may choose better alternatives more often, his decision making is subject to decision noise and random decision errors (McKelvey and Palfrey 1995, Su 2008).

Unfortunately, there is no consensus of how complexity should be defined, particularly with regard to decisions faced by human decision makers. Kauffman, an evolutionary biologist, and his colleagues (Kauffman 1993) proposed the *NK* model, originally to examine how proteins and biological organisms evolve. The *NK* model has been adapted to study complexities in many other domains. In particular, Rivkin (2000) employed the *NK* model to study the complexity of management strategies. The model has two parameters governing the complexity of a firm's decision problem; N is the number of decisions the firm faces and K ($0 \leq K \leq N-1$) is the degree of interaction between these decisions. Our definition of complexity is the most similar to this definition. The *NK* model is *a* way to characterize one prominent type of complexity in our setting, but in no way captures every notion of complexity in many different literatures; we further discuss this point in section 3.1.

Complexity, with respect to tasks carried out by humans, is also well studied in the psychology literature. Campbell (1988) provides a comprehensive review of the area. In this literature, the notion of complexity, often, is examined together with the human psychological experience. The literature is divided into three areas. The first is framing complexity primarily as a psychological experience. One example is the level of challenge experienced by an individual. A second literature examines complexity as a task–person interaction. Examples are difficulty levels perceived by an individual, his amount of experience of the particular task, and whether he is familiar with the task. The last is the study of complexity as a function of the objective characteristics of the task. This approach is similar, in spirit, to the *NK* modeling framework and to the study presented in this article. The literature includes studies of characteristics, such as information load and diversity, beyond those captured in the *NK* model and our study. In our study, the scope is limited to two objective characteristics of the decision task (i.e., the number of decisions as a measure of complexity). As our study focuses on operations management issues, we do not explicitly analyze psychological factors or the interaction of psychological experiences and the task (such as the subjects' prior knowledge of supply chain management). However, every subject in our experiments only participated in one experiment, and he or she would have no prior experience (which can be interpreted as a rudimentary control) of the specific setting of our experiments.

3. Model and Experimental Design

3.1. Model Definition

We study a setting identical to that of Kalkancı et al. (2011) with a single supplier (he) and a single buyer (she). The buyer procures a product from the supplier and sells it to the end consumer at p dollars per unit. The supplier has ample capacity and produces the buyer's orders at a cost of k dollars per unit, where $k < p$. Products are ordered before demand is realized. Any unmet demand is lost without a stockout penalty. There is no salvage value or disposal cost for leftover products. The supplier determines the *pricing scheme* of the component and the buyer selects an *order quantity* that maximizes her expected profit.

Our buyer is a newsvendor who faces a random demand, D, which is uniformly distributed between $(\mu - v)$ and $(\mu + v)$, where μ is the mean demand and v defines the range. The mean demand is the buyer's private information. The supplier only knows that μ is one of three types: high (μ_H), medium (μ_M), or low (μ_L), each with equal probability. We assume that types are equally spaced; that is, $\mu_M = \mu_L + \delta$ and $\mu_H = \mu_M + \delta$, where δ is the degree of separation between types. We also assume that the lowest possible demand is zero; that is, $v = \mu_L$. Note that our theoretical development coincides with our experimental setting. In our experimental setting, we assume that the lowest possible demand is zero, as this assumption increases the incentive to use more complex contracts by sharpening the comparison between simple and more complex contracts. Theoretically, this assumption simplifies the comparisons by removing the guaranteed payoff of the buyer.

We study different contracts between the supplier and the buyer. In a price-only contract (one-price contract), the supplier sets a single wholesale price w_1 and the buyer procures the quantity she chooses. We also study two all-unit quantity discount contracts. In the quantity discount contract with two prices (two-price contract), the supplier quotes two prices ($w_1 \geq w_2$) and a single price break Q_1. If the buyer orders less than Q_1, she pays w_1 per unit; otherwise, she pays a cheaper unit price w_2. In the quantity discount contract with three prices (three-price contract), the supplier quotes three prices ($w_1 \geq w_2 \geq w_3$) and two price breaks ($Q_1 \leq Q_2$). If the buyer orders less than Q_1, she pays w_1 per unit. If she orders more than Q_1 but less than Q_2, she pays a cheaper unit price w_2. Otherwise, she pays an even cheaper unit price of w_3. Note that the number of the supplier's decision variables depends on the contract employed, whereas the buyer's only decision is the order quantity.

Theoretically, increasingly complex contracts would benefit the supplier. In our setting, as the mean demand is one of three types, a quantity discount contract with three price blocks and a minimum quantity commitment Q_0 for the highest price is an optimal mechanism, as it has the capability to separate the types completely when needed; more price blocks would have no additional benefit for the supplier or the supply chain. As the gap between the optimal mechanism and three-price contract is only 7% (in terms of the supplier's profits) in our experimental setting, we conclude that the three-price contract models the optimal contract effectively, and we choose not to include the optimal mechanism as an additional treatment in our experiments.

We note that our definition of complexity is the most similar to the one in Kauffman's (1993) *NK* model. In particular, for the supplier's contract design problem, we study three decision scenarios (one-price, two-price, and three-price) where the subjects made one, three, and five decisions. Thus, $N = 1$, 3, and 5, respectively, in the three scenarios. In our case, $K = 2$ and 4 in the two-price and three-price cases as decisions interact. Whereas our definition of complexity can be framed in the *NK* model, our results are not explicitly derived from the analysis of the N and K parameters. We only use the framework to rank order the levels of complexities in our scenarios. That is, the three-price contract's (N,K) values are (5,4) compared with that of the price-only contract of (1,0). In this context, the three-price contract is more complex. According to the *NK* model, the complexity of the buyer's problem under different treatments is comparable. When we observe the buyer's decisions, we find similarities in the way buyers make their decisions under different treatments; in particular, they anchor to the mean, as we discuss in section 5.3.

3.2. Expected Profit Maximizing Theory of Price-Only and Quantity Discount Contracts

To derive the expected profit maximizing theory of price-only and quantity discount contracts, we first present the computerized buyer's optimal behavior. Let

$$S_j(\mu) = F^{-1}\left(\frac{p - w_j}{p}\right)^+ = \begin{cases} (\mu - v) + 2v\left(\frac{p-w_j}{p}\right) & p \geq w_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

be the order-up-to level for the wholesale price w_j (j = 1,2,3), where $F(\cdot)$ is the cumulative distribution function of demand distribution and $(x)^+ = \max(0,x)$. For a price-only contract, the buyer's optimal order quantity is $S_1(\mu)$; for quantity discount contracts, it is either at one of the order-up-to levels or at one of the price breaks (Jucker and Rosenblatt 1985).

Given the buyer's optimal response to a contract (Equation (1)), when types are equally spaced (i.e., $v = \mu_L$, $\mu_M = \mu_L + \delta$, and $\mu_H = \mu_M + \delta$, where δ is the degree of separation between types) and equally likely, the supplier's optimal wholesale price under a price-only contract is as follows:

$$w_1 = \min\left\{\frac{p}{4v}\left(\frac{\mu_L + \mu_M + \mu_H}{3}\right) + \frac{p}{4} + \frac{k}{2}, p\right\} = \frac{p}{2}\min\left\{1 + \frac{1}{2}\left(\frac{\delta}{\mu_L}\right) + \frac{k}{p}, 2\right\}. \quad (2)$$

Therefore, the buyer's and the supplier's optimal decisions under the price-only contract are fully characterized by Equations (1) and (2), respectively.

The supplier's pricing decisions under quantity discounts with asymmetric information are more involved. The supplier not only has more decisions to make, but also he utilizes prices and price breaks to separate the different types of the buyer. That is, the supplier must consider the buyer's incentive compatibility while designing contracts, which increases the complexity of his decision. However, as our goal is to study the role of contract complexity, we can restrict our attention to the parameter values where contract complexity pays off the most; that is, where the separation between types is high and the profit margin is relatively low. For these parameter values, the supplier completely separates the high type from the others, and the components of the pricing scheme can be characterized (Kalkancı et al. 2011). For a two-price contract, w_2 is set to make the high-type buyer indifferent between buying at w_1 and w_2, and

$$w_1 = \min\left\{\frac{3p}{10v}\left(\frac{\mu_L + \mu_M + \mu_H}{3}\right) + \frac{3p}{10} + \frac{2k}{5}, p\right\} = \frac{p}{5}\min\left\{3 + \frac{3}{2}\left(\frac{\delta}{\mu_L}\right) + 2\left(\frac{k}{p}\right), 5\right\},$$

$$Q_1 = (\mu_H - v) + 2v\left(\frac{p-k}{p}\right) = 2\mu_L\left(1 + \frac{\delta}{\mu_L} - \frac{k}{p}\right).$$

A similar analysis follows for a three-price contract. For a more general treatment of theory, we refer the reader to Kalkancı et al. (2011).

3.3. Experimental Design

We used a between-subject design where each subject was faced with one of three different treatments: the one-price, two-price, and three-price contracts defined above. Our human subjects were recruited from the Stanford student body and were provided monetary compensation according to their performance. Subjects were given Web-based instructions and a quiz before the experiment. (The detailed instructions and the quiz are provided in an online appendix.) We implemented our experiments in the HP Experimental Economics software platform and conducted them at the Stanford School of Medicine computer labs. To facilitate subjects' decision process, we provided them with decision support tools (Figure 1). Before a subject submitted a decision, he or she could use this "what-if" decision support tool to evaluate his/her decision. In this tool, a buyer could enter her potential order quantity based on her type. Then, the tool showed the subject her profit under different demand conditions. A supplier could enter his potential decision as well as his buyer's potential average demand and order quantity. Then, the tool showed the subject his potential profit and his buyer's profit under different demand conditions. Subjects were able to test several different decisions before submitting their order quantity/pricing scheme. They could also see a summary of results from past periods in a history table. The table included the supplier's prices/price breaks, the buyer's order quantity, the buyer's realized average demand, and the subject's revenue and profit from that period.

We performed three sets of experiments in the fall of 2009. In all experiments, an equal number of subjects were assigned either the role of supplier or buyer and each subject assumed the same role throughout the experiment. We randomly re-matched a supplier and a buyer in each period to prevent reputation-building behavior. In each period, each buyer was assigned one of three types (high, medium, or low) to guarantee that no single player determined the behavior of each type. We recruited 16 subjects each for one-price and two-price experiments and they played the game for 40 periods. In the three-price experimental session, 20 subjects were recruited and played the game for 33 periods. Overall, 52 subjects participated in our human-to-human study. In all experiments, we set the parameter values as follows: the supplier's unit cost was k = 40, the unit selling price was p = 200, and the mean demand for the low, medium, and high types were $\mu_L = 50$, $\mu_M = 90$, and $\mu_H = 130$, respectively. The range on the demand distribution was

Figure 1 A Snapshot of Buyers' and Suppliers' Decision Support Tools

(a)

My Potential Order	Wholesale Price
20	100

Demand	My Sales	My Profit	Probability of Observing a Smaller or Equal Demand
0	0	-2,000	0.01
10	10	0	0.11
20	20	2,000	0.21
30	20	2,000	0.31
40	20	2,000	0.41
50	20	2,000	0.50
60	20	2,000	0.60
70	20	2,000	0.70
80	20	2,000	0.80
90	20	2,000	0.90
100	20	2,000	1.00
0	0	-2,000	0.01

(b)

	Buyer's Potential Information	Buyer's Potential Order	My Potential Wholesale Price
Other possible values are 90 and 130	50	90	100

Buyer's Demand	Buyer's Sales	Buyer's Profit	My Profit	Probability of Observing a Smaller or Equal Demand
0	0	-9,000	5,400	0.01
10	10	-7,000	5,400	0.11
20	20	-5,000	5,400	0.21
30	30	-3,000	5,400	0.31
40	40	-1,000	5,400	0.41
50	50	1,000	5,400	0.50
60	60	3,000	5,400	0.60
70	70	5,000	5,400	0.70
80	80	7,000	5,400	0.80
90	90	9,000	5,400	0.90
100	90	9,000	5,400	1.00
0	0	-9,000	5,400	0.01

$2v = 100$. Note that for their experiments, Kalkancı et al. (2011) recruited 19 subjects for each treatment. These subjects were assigned the role of supplier using a one-price, two-price, or three-price contract, and they played against a computerized buyer for 40 periods with the same parameter values as ours.

On the basis of our parameter set, we display the optimal values of prices and price breaks in Table 1 along with the buyer's optimal procurement quantities under the one-price, two-price, and three-price contracts and the expected supplier, buyer, and total supply chain profits. In our setting, theory predicts a 20% (65%) increase (decrease) in the supplier's (buyer's) expected profits when the supplier moves from a one-price contract to a two-price contract, and an 11% (5%) increase (decrease) when the supplier moves from a two-price contract to a three-price contract. Hence, as the contract complexity increases, the supplier's expected profits increase, and the buyer's expected profits decrease. Overall, the supply chain efficiency increases 11% when a simple price-only contract is replaced with more complex and efficient quantity discount contracts.

3.4. Description of the Analysis

In our analysis, similar to Kalkancı et al., we combine regression analysis with linear hypothesis testing. Our regression model characterizes the impact of time, buyer's demand types, contract types, and variation among supplier–buyer pairs on supplier, buyer, and the total supply chain profits. All three regression equations are similar; we provide the one for the supplier's profit:

$$\begin{aligned}\Pi^{j}_{i,t} = {}& Intercept + \beta_t \times t + \beta_{QD2} \times QD2 + \beta_{QD3} \times QD3 \\ & + \beta_M \times M + \beta_H \times H + \beta_{QD2\times M} \times (QD2 \times M) \\ & + \beta_{QD2\times H} \times (QD2 \times H) + \beta_{QD3\times M} \times (QD3 \times M) \\ & + \beta_{QD3\times H} \times (QD3 \times H) + \beta_{QD2\times t} \times (QD2 \times t) \\ & + \beta_{QD3\times t} \times (QD3 \times t) + v_i + \epsilon_{i,t}.\end{aligned} \tag{3}$$

The dependent variable $\Pi^{j}_{i,t}$ is the supplier's profit under a j-price contract ($j = 1,2,3$), considering the effects of each individual supplier–buyer pair (i) and period (t). The independent variables $QD2$ and $QD3$ are dummy variables for the two-price and three-price contracts, followed by two dummy variables M and H for the medium- and high-demand types, respectively. The contract–demand type terms consider possible interactions in the relations of contract and demand types. We represent the effect of time in two ways: with an independent variable t to

Table 1 Optimal Contract Parameters (Supplier's Decisions), Procurement Quantities (Buyer's Decision), and Expected Supplier, Buyer, and Total Supply Chain Profits for Our Parameter Set under Different Contracts (Rounded up to the Next Integer)

	Supplier's optimal decisions						Buyer's optimal decisions				
	Prices			Price breaks			Procurement quantities				
	w_1	w_2	w_3	Q_1	Q_2	Profit	q_L	q_M	q_H	Profit	Total Profit
One-price	160					7200	20	60	100	2000	9200
Two-price	184	152		160		8640	8	48	160	704	9344
Three-price	200	177	147	84	160	9576	0	84	160	669	10,245

capture the direct effect and interactions terms with contract type to capture the incremental effect. We use a random-effects model (v_i) to control for the heterogeneity among supplier–buyer pairs in the subject pool. Both v_i and the error term $\epsilon_{i,t}$ are assumed to be normally distributed, with mean zero and a positive standard deviation. In our human-to-human experiments, we observe many instances where the buyer chooses not to buy, especially when faced with relatively high prices even if the expected-profit-maximizing decision would be to buy. This can be interpreted as a rejection of the contract offered by the supplier (buyers rejected 1.9%, 7.5%, and 3.8% of the contracts in the one-price, two-price, and three-price treatments, respectively; see section 5.4 for the analysis of this behavior). To have a fair profit comparison between the contracts, we restrict our attention to the instances where the buyer accepts the contract offer (see Lim and Ho 2007 for a similar approach). The regression estimates for Kalkancı et al.'s human-to-computer experiments and for our human-to-human experiments conditional on contract acceptance are provided in Table 2.

Although the regression analysis helps us devise some basic insights, a direct comparison of the average profits under different contracts does not follow immediately from this analysis because the incremental effect of quantity discounts on the profits obtained from different demand types as well as the time trend is represented by separate terms and is not averaged. Therefore, we make use of linear hypothesis testing to analyze the theoretical predictions (Freund et al. 2006). We first calculate the average profits, which corresponds to period $t^* = 20$ profits in the one-price and two-price experiments and to period $t^* = 17$ ($t^* = 20$) profits in the three-price experiment in the human-to-human (human-to-computer) setting. For example, the average profits of the supplier under one-price, two-price, and three-price treatments estimated from the regression are defined, respectively, as

$$\pi_s^1 := Intercept + \beta_t \times t^* + 1/3 \times \beta_M + 1/3 \times \beta_H,$$
$$\pi_s^2 := Intercept + \beta_t \times t^* + \beta_{QD2} + 1/3 \times \beta_M + 1/3 \times \beta_H + 1/3 \times \beta_{QD2\times M} + 1/3 \times \beta_{QD2\times H} + \beta_{QD2\times t} \times t^*,$$
$$\pi_s^3 := Intercept + \beta_t \times t^* + \beta_{QD3} + 1/3 \times \beta_M + 1/3 \times \beta_H + 1/3 \times \beta_{QD3\times M} + 1/3 \times \beta_{QD3\times H} + \beta_{QD3\times t} \times t^*.$$

We then use these average profits under different contract types to test our hypotheses.

Table 2 Regression Coefficients for Supplier, Buyer, and Total Supply Chain Profits in Kalkancı et al. (2011) Human-to-Computer (HTC) and Our Human-to-Human (HTH) Experiments

	Regression coefficient estimates and standard errors (in parentheses)					
	Supplier's profit		Buyer's profit		Total profit	
Variable	HTC	HTH	HTC	HTH	HTC	HTH
Intercept	2290.58***	2769.20***	1735.91***	1049.76	4026.7***	3765.83**
	(176.11)	(424.05)	(370.81)	(592.98)	(272.94)	(492.56)
t	15.26***	15.27	−39.39***	−29.71	−24.14***	−10.07
	(4.48)	(13.12)	(4.64)	(18.98)	(3.6)	(16.58)
$QD2$	−620.19*	155.34	−768.77	−804.92	−1390.18***	−609.62
	(249.46)	(605.70)	(524.62)	(847.69)	(386.16)	(705.86)
$QD3$	3.64	566.86	−781.23	−1186.62	−778.41*	−429.86
	(249.46)	(593.65)	(524.62)	(833.05)	(386.16)	(695.79)
M	4060.82***	4106.81***	2482.55***	3292.22***	6543.88***	7397.61***
	(126.52)	(376.60)	(131.24)	(544.75)	(101.92)	(475.42)
H	8282.34***	8520.36***	4525.1***	4840.95***	12,806.78***	13,300.87***
	(127.17)	(373.89)	(131.89)	(542.52)	(102.43)	(477.49)
$QD2 \times M$	1496.23***	−636.54	−271.43	−21.41	1227.08***	−664.99
	(179.19)	(539.78)	(185.88)	(782.24)	(144.35)	(686.14)
$QD2 \times H$	856.14***	−2096.15***	399.22*	2019.41*	1255.96***	78.48
	(179.93)	(541.00)	(186.61)	(784.76)	(144.92)	(690.38)
$QD3 \times M$	969.29***	733.13	295.54	−828.31	1267.75***	−185.45
	(179.18)	(534.79)	(185.86)	(770.71)	(144.34)	(668.56)
$QD3 \times H$	128.7	−1111.68*	1847.7***	2174.03**	1977.5***	999.83
	(179.93)	(532.36)	(186.61)	(768.78)	(144.92)	(670.51)
$QD2 \times t$	7.48	29.14	14.24*	15.89	21.72***	38.67
	(6.34)	(18.68)	(6.57)	(27.05)	(5.10)	(23.67)
$QD3 \times t$	−13.39*	−11.30	19.84**	49.64	6.44	31.54
	(6.34)	(20.49)	(6.58)	(29.60)	(5.11)	(25.80)
Standard deviation of v_i	532.91***	1589.62***	1511.36***	1905.36***	1103.34***	995.76***

Significance levels: ***$p = 0.001$, **$p = 0.01$, *$p = 0.05$.

4. Human Experiments vs. Theory

With human suppliers facing computerized buyers, Kalkancı et al. (2011) test the theoretical prediction that suppliers increase their profits by increasing the complexity of their contracts, that the supply chain efficiency of complex quantity discount contracts is higher than the efficiency of a price-only contract, and that buyers' profits decrease as the complexity of the contract increases. They do so by comparing supplier, buyer, and total supply chain profits in their experiments with theory. With human suppliers facing human buyers, we test similar predictions. Table 3 reports supplier, buyer, and total supply chain profits as well as their respective comparisons with theoretical predictions in human-to-computer and human-to-human experiments.

Our human-to-human experiments provide support for the conclusions of Kalkancı et al.'s human-to-computer experiments as well as extend their results to more general settings. When we compare different types of contracts, we observe that contract type has no significant effect on supplier, buyer, and total supply chain profits, with the exception that the two-price contract leads to higher profits for the supplier than a one-price contract in human-to-computer experiments.[2] That is, even when there is potential improvement to suppliers' profits from using quantity discounts, this improvement is captured with a low number of price blocks. Furthermore, when human-to-human interactions and strategic anticipations are considered, the simplest contract works as effectively as the more complex ones for suppliers.

In both human-to-computer and human-to-human experiments, suppliers earn significantly less (with the exception of one-price contract in human-to-computer experiments) and buyers earn significantly more than what theory predicts. That is, there is a more equitable distribution of profits between suppliers and buyers compared with what is theoretically predicted. Furthermore, total supply chain is significantly higher (with the exception of three-price contract in human-to-human experiments) than what theory predicts. Therefore, statistically speaking, human subjects do not achieve better coordination with more complex quantity discount contracts, but they do achieve better coordination than theory predicts.[3]

Note that for our profit comparisons we excluded rejected contract offers. When the rejections are included, we observe that one-price and three-price treatments are similar in terms of supplier and total supply chain profits (two-sided p-value: 0.61 and 0.28, respectively). The two-price treatment, on the other hand, leads to significantly lower supplier and total supply chain profits (p-values are 0.04 for the comparisons with one-price and three-price treatments). We observe no significant difference in buyer's profits under the three treatments (two-sided p-values are 0.62, 0.96, and 0.64 for one-price and two-price, two-price and three-price, and one-price and three-price treatment comparisons, respectively). All our observations regarding profit comparisons with theory remain the same.

5. Behavioral Results

In the absence of social preferences, Kalkancı et al. (2011) identify reinforcement and bounded rationality as key biases that impact human subjects' decisions. In this section, we consider the behavioral drivers of the results in our human-to-human experiments.

5.1. Suppliers Offer More Favorable Contracts to Human Buyers Compared with Computerized Buyers

In both human-to-computer and human-to-human experiments, suppliers chose lower prices compared with the theoretical predictions under all contracts

Table 3 Means and Standard Deviations (in Parentheses) for Supplier, Buyer, and Total Supply Chain Profits in Human-to-Human and Human-to-Computer Experiments (Rounded to the Nearest Integer) and Comparison of Respective Human-to-Human and Human-to-Computer Profits with Theoretical Predictions in Table 1

	Human-to-human			Human-to-human (including rejected contracts)			Human-to-computer		
	Supplier	Buyer	Total	Supplier	Buyer	Total	Supplier	Buyer	Total
One-price	7626	3310***	10,839***	7471	3247***	10,633***	**6722*****	3270***	9992***
	(1124)	(552)	(2876)	(4342)	(4124)	(6209.65)	(540)	(1314)	(888)
Two-price	**7538*****	3708***	11,312***	**6897*****	3412***	10,620**	**7065*****	2841***	9906*
	(2286)	(806)	(2973)	(4205)	(747)	(3217)	(480)	(1382)	(1094)
Three-price	**7726*****	3478***	10,736**	**7411*****	3366***	10,381*	**6840*****	3619***	10,459
	(1747)	(1302)	(4112)	(3942)	(1298)	(4141)	(727)	(1830)	(1320)

Note. For the significant values in bold (roman), experimental profits are lower (higher) than theoretical predictions.
Significance levels: $***p = 0.001$, $**p = 0.01$, $*p = 0.05$.

Figure 2 Suppliers' Price Decisions under All Contracts; (a) One-Price, (b) Two-Price, and (c) Three-Price

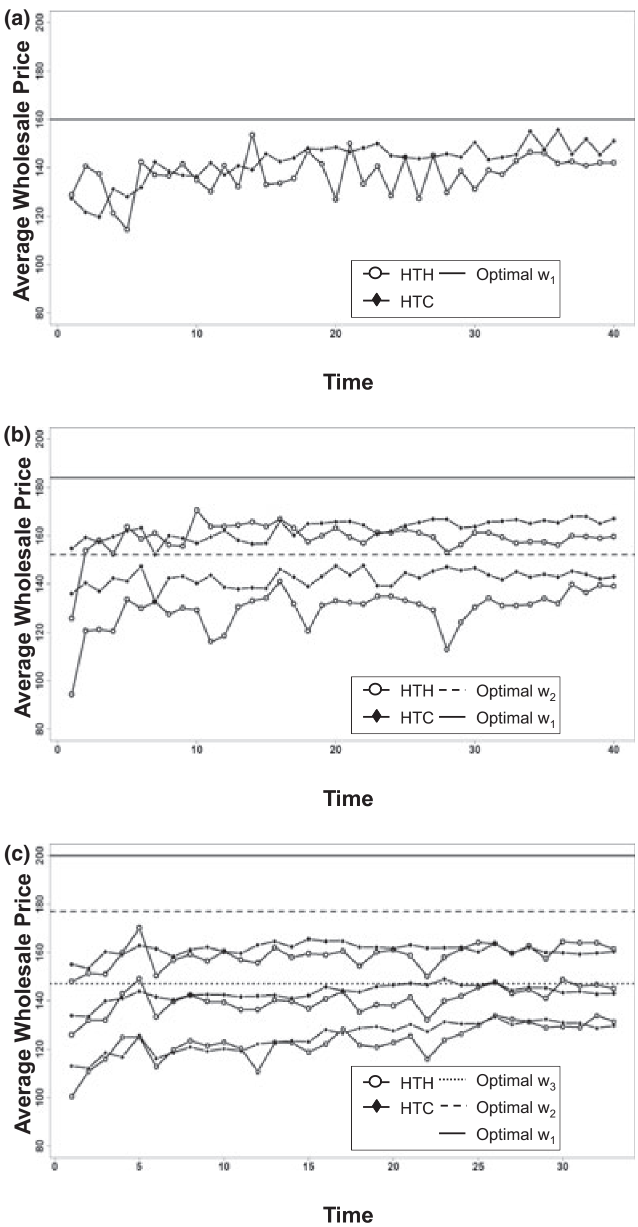
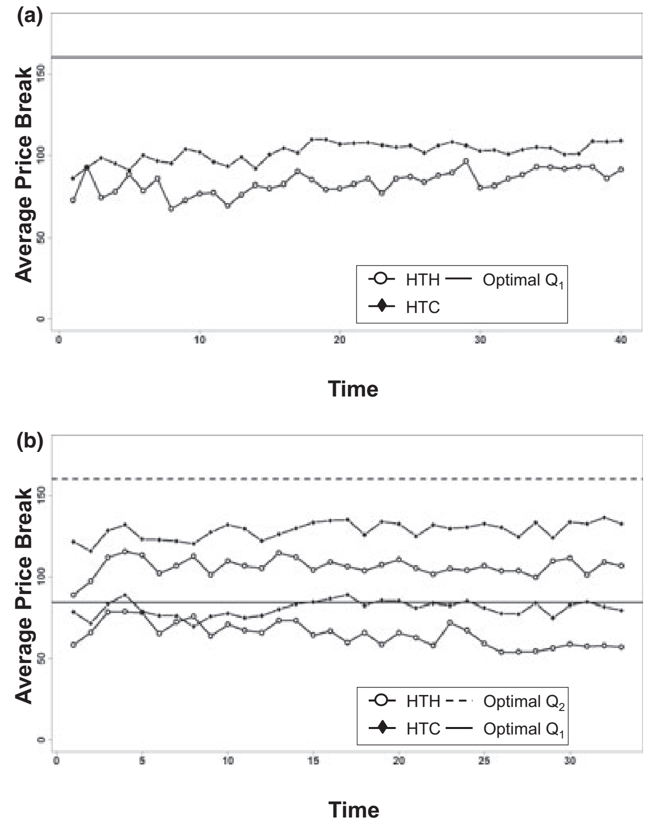


Figure 3 Suppliers' Price Break Decisions under All Contracts: (a) Two-Price and (b) Three-Price



(Figure 2); two-sided Wilcoxon tests comparing the median of the average prices with its theoretical prediction also support this conclusion at a significance level under 0.05 for all tests. Suppliers also had difficulty setting the correct price breaks (Figure 3). At optimality, the discount scheme should be set such that for medium and high types of the buyer the optimal procurement quantity should be at the price break in the three-price treatment (Table 1). That is not the case in the experiments: in the human-to-human setting, out of a total of 111 (112) high (medium)-demand type occurrences, a rational buyer's order quantity would be equal to the correct quantity break in only 43 (16) cases.[4] Furthermore, we observe that prices and price breaks set by human suppliers are closer to the midpoints of the decision ranges. This behavior is consistent with the midpoint bias under bounded rationality observed in newsvendor models for the buyer's decisions (Su 2008). Although this behavior may limit the supplier's ability to separate different types of the buyer and to extract profits, it actually reduces the inefficiency in the system and increases the total profits compared with theory.

When we compare the supplier's decisions in human-to-human and human-to-computer experiments using Wilcoxon tests, we see that prices are not significantly different from each other (two-sided p-values are 0.418 for w_1 comparison of the one-price treatment; 0.935 and 0.285 for w_1 and w_2 comparisons of the two-price treatment; and 0.86, 0.90, and 0.98 for w_1, w_2, and w_3 comparisons of the three-price treatment, respectively). However, all price breaks are statistically lower in human-to-human experiments (p-values are 0.051 for Q_1 comparison of the two-price treatment and 0.065 and 0.060 for Q_1 and Q_2 comparisons of the three-price treatment, respectively). These contracts are more favorable to buyers than those offered in human-to-computer experiments because buyers can achieve discounts at lower quantities. As the only difference between human-to-human and human-to-computer experiments is that humans, instead of computers, assume the role of the buyer, we argue that the most probable drivers for any differences between these two sets of experiments are social preference effects. Indeed, our results are consistent with this interpretation, as we discuss in more detail in section 5.4.

Interestingly, even though suppliers select more favorable price breaks under human-to-human experiments than under human-to-computer experiments,[5] when we compare the *initial* decisions between human-to-human and human-to-computer experiments, we do not observe any significant differences. Therefore, the differences between price breaks under these two experiments are caused by the interactions between suppliers and human buyers. While suppliers are boundedly rational with respect to break point decisions, they do seem to understand, at least conceptually, the impact of their decisions on buyers' order quantities over time, and they adjust their decisions due to social preferences. Therefore, the differences in the decisions over time can be attributed to buyers' behavior and its impact on the decisions of suppliers.

5.2. Buyers Cannot Fully Take Advantage of the More Favorable Contracts Offered by Suppliers

A comparison of supplier, buyer, and total profits between our experiments and human-to-computer experiments by Kalkancı et al. (2011) reveals that profits are not significantly different in most cases.[6] The only exception is the buyer's profit in the two-price treatment, which is significantly higher in human-to-human experiments (p-value 0.03). However, when we replace buyers with a hypothetical rational buyer who maximizes her expected profits (means and standard deviations (in brackets) for rational-buyer profits are 3900 [580], 4300 [442], and 4035 [1244] under one-price, two-price, and three-price treatments, respectively), the rational buyer would in fact make significantly higher profits under all contracts compared with human-to-human experiments (p-values are 0.004, 0.004, and 0.002 in Wilcoxon tests, respectively) and under the one-price and two-price contracts compared with human-to-computer experiments (p-values are 0.022 and 0.002 in Wilcoxon tests, respectively). That is, suppliers offer more favorable contracts to *human* buyers, although buyers may not be able to take advantage of them.

This observation also provides strong evidence that human buyers, just like human suppliers, are boundedly rational and that they deviate from the expected-profit-maximizing behavior. Indeed, we observe that under the one-price contract, buyers deviate from theory and significantly over-order, particularly when they have medium or high types (at a 0.01 significance level) (the details of the analysis are provided in section 5.3). Note that a higher order quantity leads to higher profits for suppliers as well as for the supply chain. Therefore, buyers' over-ordering behavior under the one-price contract can be tied to our results in section 4 and helps us explain why this contract performs effectively compared with other contracts in our experiments.

5.3. Buyers Exhibit Mean Demand Anchoring and Demand Chasing Behaviors

As shown in the previous section, human buyers act boundedly rational, and they deviate from the expected-profit-maximizing behavior. We use a regression model to characterize the difference between the buyer's actual order quantity and the expected-profit-maximizing order quantity considering the impact of the demand type, contract type, and variation among buyers:

$$\begin{aligned}\Delta^{j}_{i,t} = {} & Intercept + \beta_{QD2} \times QD2 + \beta_{QD3} \times QD3 + \beta_{M} \times M \\ & + \beta_{H} \times H + \beta_{QD2\times M} \times (QD2 \times M) \\ & + \beta_{QD2\times H} \times (QD2 \times H) + \beta_{QD3\times M} \times (QD3 \times M) \\ & + \beta_{QD3\times H} \times (QD3 \times H) + v_i + \epsilon_{i,t}.\end{aligned}$$

The dependent variable $\Delta^{j}_{i,t}$ is the difference between the buyer's order quantity and the best-response quantity under a j-price contract ($j = 1,2,3$), considering the effects of each individual buyer i and period t. (We do not consider the effect of time trend; however, when period is added to the regression model, it turns out not to be significant and the general conclusions remain valid.) The independent variables are defined similarly to the ones in Equation (3), with the exception that v_i represents the variability between buyers. Table 4 displays the regression estimates.

Table 4 Regression Coefficients and Standard Errors (in Parentheses) for the Difference between the Buyer's Order Quantity and the Best-Response Quantity Conditional on Contract Acceptance

Variable	Intercept	M	H	$QD2$	$QD3$	$QD2 \times M$	$QD2 \times H$	$QD3 \times M$	$QD3 \times H$	Std. dev. of v_i
Difference	3.59	5.10	5.41	1.41	4.00	-12.88^{**}	-8.99^{*}	-11.01^{*}	-8.69^{*}	8.22^{***}
	(3.70)	(3.07)	(3.11)	(5.31)	(5.04)	(4.45)	(4.46)	(4.32)	(4.35)	

Significance levels: $^{***}p = 0.001$, $^{**}p = 0.01$, $^{*}p = 0.05$

Table 5 Estimates and Standard Errors (in Parentheses) for the Difference between the Buyer's Order Quantity and the Best-Response Quantity Conditional on Contract Acceptance

	Low	Medium	High
One-price	3.59	8.69**	9.00**
	(3.70)	(3.54)	(3.57)
Two-price	5.00	−2.79	1.41
	(3.80)	(3.57)	(3.56)
Three-price	7.59*	1.68	4.31
	(3.41)	(3.32)	(3.32)

Significance levels: $**p = 0.01$, $*p = 0.05$

We then use linear hypothesis testing to estimate the differences between the actual and theoretical order quantities (as shown in Table 5). We find that under the one-price contract, buyers deviate from theory and significantly over-order, particularly when they have medium or high types (at a 0.01 significance level). However, we fail to see any systematic differences in actual order quantities and the best-response quantities in the two-price and three-price treatments.

Decision biases in the newsvendor setting as outlined by the behavioral operations management literature can be used to explain the differences in the buyer's behavior across different treatments. The existing behavioral operations management literature shows that under the newsvendor setting, human subjects make their decisions consistent with (i) mean demand anchoring and (ii) demand chasing. In our experiments, the buyer's problem in each period is equivalent to the newsvendor setting in the one-price treatment. Consistent with the existing results, we observe that buyers order more than the best response when the buyer's best response is below the mean demand in this treatment. As this is prevailing in the one-price treatment (i.e., in 284 of 311 instances in the one-price treatment, the buyer's best response is below the mean demand), mean demand anchoring leads to significant over-ordering in the one-price treatment.

Under the two-price and three-price treatments, suppliers price and price break decisions may lead to a best-response quantity above or below the mean demand: for example, we have 193 (209) instances where the mean demand is above the best response and 85 (100) instances where the mean demand is below the best response in the two-price (three-price) treatment. In the region where mean demand is above (below) the best response, we see significant over-ordering (under-ordering). These effects, however, tend to cancel out each other and therefore we cannot see any significant difference between the actual and best-response quantities when we consider the buyer's overall ordering behavior. Interestingly, the mean absolute deviations between the actual order quantities and the best responses are 15.76, 16.63, and 18.18 for one-price, two-price, and three-price treatments, respectively. These values are not significantly different from each other (p-values are 0.73, 0.51, and 0.31 for the comparisons of one-price and two-price, two-price and three-price, and one-price and three-price pairs of contracts, respectively). This observation is consistent with our interpretation of complexity with respect to the *NK* model, according to which the complexity for the buyer is the same across treatments.

To validate that these decision biases are significant in our experiments, we test mean demand anchoring and demand chasing models. The mean demand anchoring model is represented by

$$q_{i,t} = \alpha \times \mu_{i,t} + (1 - \alpha) \times br_{i,t} + v_i + \epsilon_{i,t},$$

where α is estimated in the model, $q_{i,t}$ is the quantity for subject i in period t, $\mu_{i,t}$ is the mean demand for subject i in period t, $br_{i,t}$ is the best-response quantity for subject i in period t, v_i is the random-effects term for subject heterogeneity, and $\epsilon_{i,t}$ is the error term for subject i.

To capture demand chasing, we utilize a partial linear adjustment model (as outlined by Bostian et al. 2008):

$$(q_{i,t} - \mu_{i,t}) = (q_{i,t-1} - \mu_{i,t-1}) + \beta \times (\hat{D}_{i,t-1} - \mu_{i,t-1}) + v_i + \epsilon_{i,t},$$

where β is estimated in the model and $\hat{D}_{i,t-1}$ is the realized demand for subject i in period $(t - 1)$.

Table 6 shows the results of the estimated parameters. We see traces of both of these behavior patterns in our experiments. The parameters of each model

Table 6 Model Coefficients for Buyers in One-Price, Two-Price, and Three-Price Treatments and Their Significance Levels

	Mean demand anchoring				Demand chasing			
Treatment	α	$\sigma(v_i)$	AIC	LLK	β	$\sigma(v_i)$	AIC	LLK
One-price	0.25***	7.6***	2726.3	−1360.15	0.37***	3.05	2848.23	−1421.12
Two-price	0.59***	9.03***	2584.34	−1289.17	0.39***	5.52*	2648.49	−1321.24
Three-price	0.63***	8.8***	2694.77	−1344.98	0.29***	0.001	2784.34	−1389.17

Note. LLK = log likelihood; AIC = Akaike information criteria.
Significance levels: $***p = 0.001$, $*p = 0.05$

under each treatment are significant. However, mean demand anchoring provides a better fit to our observation, resulting in better model selection criteria such as log likelihood (LLK) and Akaike information criteria (AIC).

On the basis of these observations, we conclude that even though quantity discounts are supposed to lead to higher buyer ordering in theory, over-ordering in the one-price treatment can help to close the gap between price-only contracts and quantity discount contracts in our experiments and further contributes to the limited benefit from complexity in supplier and total profits.

5.4. We Find Partial Support for Fairness Concerns in Buyers' Ordering Behavior

We observe many instances where buyers chose not to buy especially when faced with relatively high prices even if the expected-profit-maximizing decision would be to buy; that is, they rejected the contract offered by the supplier. Buyers rejected 1.9%, 7.5%, and 3.8% of the contracts in the one-price, two-price, and three-price treatments, respectively. This behavior is also consistent with social preference effects, as subjects rejected money-making offers when they considered them unfair. Our analysis shows that the rejected offers, on average, tend to have lower absolute and relative profits overall compared with other offers. This result is statistically significant at the 0.05 level for one of three players (who occasionally rejected the contract offers) in the one-price treatment, three of six in the two-price treatment, and two of four subjects in the three-price treatment. Note that we did not explicitly control for different types of social preferences, that is, fairness, altruism, and reciprocity.

To examine if the buyer's ordering behavior is mainly driven by fairness concerns, we estimate the following model in each of our treatments: For each possible order quantity q (including $q = 0$), we assume that the buyer's utility $U_B(\cdot)$ is a function of the buyer's and the supplier's expected profits ($\pi_B(q)$ and $\pi_S(q)$, respectively) and is given by

$$\begin{aligned} &U_B(\pi_B(q,D), \pi_S(q,D)|\alpha, \gamma) \\ &= \pi_B(q,D) - \alpha \max(\gamma \pi_S(q,D) \\ &\quad - \pi_B(q,D), 0). \end{aligned} \tag{4}$$

In this equation, α is the fairness parameter and γ is the coefficient of the supplier's share that is perceived as fair by the buyer. This utility structure closely follows Cui et al. (2007). Unlike their model, we do not include an additional term to capture the buyer's inequality aversion if she expects to earn more than her supplier, as past research has shown that "subjects suffer more from inequity that is to their monetary disadvantage" (Fehr and Schmidt 1999). We choose to do that so as to prevent over-parametrizing the model.

Given the buyer's utility function, we assume that the buyer makes a probabilistic choice from the set of possible order quantities. Due to the heterogeneity in the subject pool, we used a mixed-effects model together with the fairness and bounded rationality model in Equation (4). We assume that each of the fairness parameters, α and γ, follows a lognormal distribution (with means μ_α and μ_γ and standard deviations σ_α and σ_γ for α and γ, respectively). We use lognormal instead of normal distribution to capture the non-negativity and the heavy right tail of the parameters. Each participant i also has a bounded rationality parameter λ_i. We maximize the likelihood of the experimental observations over the participant pool, which can be calculated as

$$\max_{\mu_\alpha,\sigma_\alpha,\mu_\gamma,\sigma_\gamma,\lambda_1,\cdots,\lambda_n} \int_0^\infty \int_0^\infty \prod_i \prod_t Prob(x_{i,t}|\alpha = \alpha_l, \gamma = \gamma_k, \lambda_i) \times f(\alpha_l|\mu_\alpha, \sigma_\alpha) g(\gamma_k|\mu_\gamma, \sigma_\gamma) d\alpha_l d\gamma_k.$$

Here, i is the index for participants, t is the index for the period in the experiment, $x_{i,t}$ is the player's decision at period t, and $Prob(\cdot)$ is the probability function (which follows the logit form):

$$\begin{aligned} &Prob(x_{i,t}|\alpha = \alpha_l, \gamma = \gamma_k, \lambda_i) \\ &\quad = \frac{e^{\lambda_i E_D[U_B(\pi_B(x_{i,t},D),\pi_S(x_{i,t},D)|\alpha_l,\gamma_k)]}}{\sum_q e^{\lambda_i E_D[U_B(\pi_B(q,D),\pi_S(q,D)|\alpha_l,\gamma_k)]}}. \end{aligned}$$

Using this model, we obtain more stable parameters for α and λ than estimating these parameters for each participant separately. Table 7 shows the estimated model parameters.

We find some support that the design of the experiments affects the parameters. For example, one-price

Table 7. Model Coefficients for Buyers in One-price, Two-price, and Three-price Treatments.

	One-price	Two-price	Three-price
λ_1	0.11	0.43	0.28
λ_2	1.35	0.67	0.28
λ_3	0.78	0.55	0.34
λ_4	0.13	0.54	0.33
λ_5	0.36	1.62	0.39
λ_6	0.13	2.82	0.36
λ_7	0.93	0.35	0.75
λ_8	0.80	0.51	0.44
λ_9			0.85
λ_{10}			0.63
$\mu_\alpha(\sigma_\alpha)$	1.01 (0.15)	1.19 (0.78)	1.01 (0.15)
$\mu_\gamma(\sigma_\gamma)$	1.16 (0.68)	1.20 (0.80)	1.23 (0.89)

Note. All λ values are scaled by 1000 to improve exposition.

and two-price treatments are not significantly different from each other in terms of the buyer's bounded rationality parameter, whereas the bounded rationality parameter under the three-price treatment is significantly lower than that of the two-price treatment at the 0.05 level using a Wilcoxon test. This observation provides some support that complexity reduces the precision in decision making. Note that buyers were provided with a decision support tool and that the number of possible decisions remains constant across treatments. Thus, the difficulty of the decision tasks may have been equalized to a degree across treatments. Nevertheless, we do find some differences between two-price and three-price treatments.

We observe evidence for fairness concerns, as the means of α are at least 1 in all treatments. The actual fair point for participants can vary significantly, as the standard deviation of γ is quite large compared with its mean in all treatments. The estimated parameters complement our findings from the rejection rates in that the stronger evidence for fairness concerns is obtained under the two-price treatment while we also observe a higher variability for this treatment.

5.5. We Find Partial Support for Higher Supplier Beliefs on Fairness Concerns in Human Buyers' Ordering Behavior

We use a two-stage behavioral model to capture suppliers' decisions. In the second stage, buyers make probabilistic choice in the standard quantal response formulation, with a utility function given by Equation (4). There are three parameters for each buyer: the bounded rationality parameter λ_b, the fairness parameter α, and the coefficient of the supplier's share that is perceived to be fair γ. In the first stage, the supplier can calculate the expected utility associated with every decision, given the quantal response decision probabilities of the buyer in the second stage. The supplier then chooses a decision based on the same probabilistic choice framework and these expected utilities. The supplier has one parameter λ_s for bounded rationality. We estimate these four parameters only from suppliers' decisions. Thus, the parameter estimates of $(\lambda_b, \alpha, \gamma)$ are the beliefs of the supplier on how the buyer would react instead of the true parameters of the buyer.

$$\max_{\lambda_s,\lambda_b,\alpha,\gamma} \prod_t \frac{e^{\lambda_s E_\mu \left[\sum_q E_D[\pi_S(\mathbf{w_t},q,D)] Prob(q|\lambda_B,\mu,\alpha,\gamma,\mathbf{w_t})\right]}}{\sum_{\mathbf{w}} e^{\lambda_s E_\mu \left[\sum_q E_D[\pi_S(\mathbf{w},q,D)] Prob(q|\lambda_B,\mu,\alpha,\gamma,\mathbf{w})\right]}}, \text{ where}$$

$$Prob(q|\lambda_B,\mu,\alpha,\gamma,\mathbf{w_t}) = \frac{e^{\lambda_B E_D[U_B(\pi_B(\mathbf{w_t},q,D),\pi_S(\mathbf{w_t},q,D)|\mu,\alpha,\gamma)]}}{\sum_q e^{\lambda_B E_D\left[U_B\left(\pi_B(\mathbf{w_t},q,D),\pi_S(\mathbf{w_t},q,D)|\mu,\alpha,\gamma\right)\right]}}$$

In this problem, $\mathbf{w_t}$ describes the decision vector of the supplier at period t. The summation in the denominator of the objective function is over the set of possible decisions (where $\mathbf{w}$ denotes a decision vector in this set). Note that, because of the high dimensionality of decisions in our two-price and three-price treatments, we discretized each price decision in a step size of 10 ($k = 40 \le w_3 \le w_2 \le w_1 \le p = 200$) and each quantity break in a step size of 5 ($0 \le q_1 \le q_2 \le 180$). We rounded each actual decision by the supplier to the nearest decision in the subset decision space we are considering. Table 8 summarizes the parameter estimates.

We found that even when the suppliers interact with computerized buyers, they have fairness concerns, as α is significant for almost all players even in human-to-computer experiments. Moreover, α is significantly higher in human-to-human experiments compared with human-to-computer experiments for the one-price and two-price treatments at the 0.05 level with a Wilcoxon test. This is partial support that suppliers believe that human buyers have higher fairness concerns than computerized buyers, which is consistent with the intuition that human suppliers would attach more importance to social preferences when they face human (as opposed to computerized) buyers. The comparison, however, is not significant for the three-price setting. We speculate that in the three-price setting, complexity of the contract decisions overwhelms other considerations, and the suppliers no longer focus on the fairness concerns of the buyers.

All other comparisons between beliefs on buyers' behavioral parameters are insignificant. In addition, λ_s, the suppliers' bounded rationality parameter, is significantly higher in the one-price setting for human-to-computer experiments compared with human-to-human experiments. It is partial support that suppliers can make better decisions facing computers instead of another human. This difference is no longer significant when contracts become more complex in the two-price and three-price settings. Table 9 summarizes all comparison statistics.[7]

6. Concluding Remarks

Despite being theoretically suboptimal, simpler contracts (such as price-only contracts and quantity discount contracts with a limited number of price blocks) are commonly preferred in practice. Thus, exploring the tension between theory and practice regarding complexity and performance in contract design is especially relevant. Using human subject experiments, Kalkancı et al. (2011) showed that such simpler contracts perform effectively for a supplier interacting with a computerized buyer under asymmetric demand information. We use a similar set of experiments with the modification that human suppliers

Table 8 Model Coefficients for Buyers in One-Price, Two-Price, and Three-Price Treatments and Significance Levels of these Coefficients Compared with Their Rational Values (1000 for λ, 0 for α, and 1 for γ)

	One-price				Two-price				Three-price			
Player	λ_b	α	γ	λ_s	λ_b	α	γ	λ_s	λ_b	α	γ	λ_s
Human-to-computer experiments												
1	0.02	108.55	1.11*	4.63	0.43	0.10	1.78	5.65	0.45	41.76	0.46	19.09
2	3.12	4.38	0.33	3.26	0.51	1.40	0.59	7.02	0.79	0.06	1.18	2.72
3	1.72	0.71	1.64	10.64	0.13	4.06	0.70	48.73	0.09	0.98	2.23	5.13
4	67.30	0.03**	0.91	25.48	0.11	1.25	2.07**	6.13	0.39	0.87	1.02	8.34
5	16.81	0.74*	0.01*	4.52	1.37	0.12	1.00	2.80	0.00	1.59	98.24	7.89
6	29,470.47	0.26	0.33*	10.03	1.95	0.00**	244.75	1.53	0.60	0.20	1.00	4.37
7	0.28	128.66	1.00	1.18	0.55	0.00	3937.59	12.43				
8	0.11	1.20	2.72	3.67	0.03	19.52	0.33	2.23	0.30	14,056.00	0.60	11.42
9	0.06	16.44	0.45	6.77	0.01	0.54**	31.87**	2.64	0.92	0.02	1.00	2.85
10	196.96	0.25	0.33	1.83	0.23	2.49	0.78	7.00	2.07	3.95	0.07	1.62
11	1.15	1.48	0.65	36.58	0.66	0.00	1.00	0.69	0.17	1.31	2.26	10.89
12	652.75	0.00	79.01	9.24	0.71	0.00	2332.49	5.83	0.29	2.46	0.91	12.25
13	3.93	0.41	1.00	3.70	0.52	0.07	1.58	4.52	0.56	8.56	1.38*	11.88
14	6.62	56.11	0.14	5.62	0.46	0.33	4.82	5.38	2.12	8.54	0.68	9.73
15	2.10	0.32	0.99	3.43	0.41	0.36	2.87	14.90	14.01	0.30	2.60	4.19
16	11.42	9.06	0.07	1.15	0.98	0.00	1076.39	3.97	0.72	0.01	1.00	4.27
17	3.10	0.49	1.37	3.83					0.25	0.36	1.00	6.27
18	0.92	77.52	1.00	1.48	0.08	0.00	1.00	2.45	0.47	3.09	0.50	4.04
19	3.02	1.23	1.00	12.57	0.21	0.62	1.35	1.93	0.69	0.26	1.00	6.79
Human-to-human experiments												
1	3.06	25,839.47	0.33	5.47	0.15	4.30	3.00	3.95	6.92	9.11	0.86	2.56
2	0.03	24.20*	0.25	1.37	0.97	0.95	0.93	0.00	1.00	1.02	1.02	0
3	114.60	0.17	0.94	0.50	1.99	11.10	0.15	1.15	0.00	1.33	35.59	6.71
4	0.00	1101.49	0.33	0.97	0.83	115.41	0.14	1.41	1.08	0.00	836.40	2.71
5	7.15	14.38	23.11	0.00	0.01	1.04	26.68	4.46	0.13	2.80	0.49	14.93
6	0.55	1.05	0.88	2.29	1.54	0.55	0.49	3.68	0.21	3.60	1	29.06
7	1.48	40.07	0.33	2.25	0.20	1.26	0.78	4.54	26,236	2.14	0.52	4.62
8	0.34	196.71	31.00	5.83	0.79	0.03	6.05	5.76	0.17	2.90	0.52**	16.46
9									42.38	3.66	0.48**	1.25
10									0.91	2.29	0.45	0.85

Note. The significance levels of each coefficient are calculated using a log-likelihood ratio test. All λ values are scaled by 1000 to improve exposition. Significance Levels: The values that are not significant are in gray boxes. $**p = 0.01$, $*p = 0.05$. All other entries are significant at the 0.001 level.

Table 9 Comparison Statistics of Model Coefficients for Buyers in One-Price, Two-Price, and Three-Price Treatments in Human-to-Computer and Human-to-Human Experiments

	λ_b			α			γ			λ_s		
	One-price	Two-price	Three-price	One-price	Two-price	Three-price	One-price	Two-price	Three-price	One-price	Two-price	Three-price
HTC	3.12	0.83	0.51	1.23	0.95	1.15	1.00	1.01	0.93	4.52	2.95	4.66
HTH	1.02	0.81	1.00	32.13	1.15	2.55	0.60	0.86	0.63	1.81	3.82	3.18
p-value	0.24	0.32	0.47	**0.02**	**0.02**	0.41	0.42	0.16	0.42	**0.01**	0.13	0.83

Note. *p*-values in bold are one-sided and are significant at the 0.05 level; Other *p*-values are two-sided.

interact with human buyers. Like Kalkancı et al., we test the key theoretical predictions that increasing the complexity of contracts employed improves suppliers' profits and supply chain efficiency while decreasing buyers' profits. Similar to Kalkancı et al.'s results, our experimental findings contradict these predictions. First, we find that increasing the complexity of the contract does not increase suppliers' profits. Furthermore, the price-only contract achieves equal total supply chain profits compared with more complex quantity discount contracts. Finally, we observe a more equitable distribution of profits between suppliers and buyers. As such, our human-to-human experiments provide support for the conclusions of Kalkancı et al.'s human-to-computer experiments and extend their results to more general settings. Based on human-to-computer and human-to-human experiments, we conclude that there is a non-trivial tradeoff between complexity and inefficiency of all-unit quantity discount contracts: the notion that complex contracts can optimize the supplier's profit is flawed and requires deeper consideration.

In the absence of social preferences, Kalkancı et al. argue that reinforcement and bounded rationality are key biases that impact subjects' decisions. In human-to-human experiments, we find evidence that suppliers provide better contracts to human buyers, compared with a computerized one, by giving her discount breaks that would trigger the discount sooner. We also observe many instances where buyers choose not to buy even if the expected-profit-maximizing decision would be to do so; that is, they reject contracts that they consider to be unfair. These behaviors of suppliers and buyers can be interpreted as social preference effects. However, these effects may be secondary to bounded rationality, as profit splits between suppliers and buyers are not more equitable in human-to-human experiments, once again highlighting the importance of considering complexity as a contract design factor.

In addition to providing validation for Kalkancı et al.'s human-to-computer experiments, human-to-human experiments further our understanding of when automating some players in an experiment makes a difference and when it does not. First, we find that suppliers have fairness concerns even when they interact with computerized buyers. These fairness concerns tend to be even stronger when suppliers interact with human buyers, particularly when the complexity of the contract is low. We also find that suppliers are more prone to random decision errors (i.e., bounded rationality) when interacting with human buyers.

It is evident from our experiments that bounded rationality is an important factor in explaining the behavior of our subjects. Moreover, changing contract complexity plays a role in shaping the bounded rationality of players. This can easily be seen in the supplier's behavior as the number of decisions the supplier has to make increases with complexity. However, as the supplier's problem is not affected by the state of the environment, suppliers can anchor to their previous decisions, which can be explained well by an experience-weighted attraction learning model (Camerer and Ho 1999, Kalkancı et al. 2011). Thus, it would be interesting to consider the complexity-dependent bounded rationality of both parties theoretically. One approach is to model the interactions, under different contracts, with quantal response equilibrium, modified to capture social preferences.

Acknowledgments

This research was funded in part by NSF grants CAREER-0547021 and DMI-0400345. We would like to thank Elena Katok and our anonymous review team for helpful suggestions. We are also grateful to participants at the 2012 INFORMS Annual Meeting (Phoenix, AZ) and the 2012 POMS Annual Meeting (Chicago, IL) for their comments.

Notes

[1]In this case, mechanism design focuses on one dimension of design choice (i.e., the level of contract complexity) and should not be confused with the traditional principal–agent theory where the whole payment schedule is determined. We use the term "mechanism design" in a broader sense.

[2]Note that human-to-computer experiments show that subjects, albeit not perfectly, can understand and take advantage of the more complex two-price contract. Thus, the failure to do so in human-to-human experiments points to reasons beyond subjects not understanding the mechanics of the discount contract.

[3]Our analysis is based on the average profits, which correspond to period $t^* = 20$ profits in the one-price and two-price experiments and to period $t^* = 17$ profits in the three-price experiment in the human-to-human setting. To check for the robustness of our observations, we replicated the analysis using $t^* = 40$ in the one-price and two-price experiments and $t^* = 33$ in the three-price experiment. We observe that our conclusions for supplier and buyer profits remain the same. We see a reversal of our result in total supply chain profit comparisons under the one-price and the three-price treatments. Total supply chain profits under the three-price treatment are higher than total profits under the one-price treatment with the new t^*; the other observations on total supply chain profits remain the same. As our aim is to understand the supplier's contract design problem, we conclude that our main managerial insights do not change with the choice of t^*.

[4]Similar to Kalkancı et al.'s (2011) experimental design, subjects in our experiments are allowed to use price-only contracts in the two-price and three-price treatments even though this option is not publicized. Kalkancı et al. report that 1 of 19 subjects in their two-price treatment and 1 of 19 subjects in their three-price treatment consistently reduced the contracts that they offered to one-price and two-price, respectively. The former subject performed better than the average and the latter subject performed worse than the average. In our experiments, one supplier reduced a three-price contract to a two-price contract by setting a prohibitively high second price break in 5 of 33 instances. This observation shows that at least some subjects were aware of that option. However, in majority, they chose not to use it.

[5]We are unaware of any explanations in the literature as to why suppliers would choose to provide a more favorable price break, instead of more favorable prices, to buyers. In a related set of experiments where human suppliers interact with computerized buyers, Kalkancı et al. (2011) observe that suppliers are not able to separate out different types of the buyer by using the price breaks effectively, and the price breaks, if improved, would lead to the highest improvement in suppliers' profits. This can be interpreted as a difficulty on the suppliers' part to make price break decisions. In a human buyer to human

supplier setting, we speculate that the interaction of the difficulty of setting price breaks and the learning dynamics driven by the human buyer's social preferences may lead suppliers to make inferior price break decisions for themselves, which is favorable to their buyers.

[6]We use Wilcoxon tests to compare the mean profit per period among supplier–buyer pairs. Two-sided p-values are 0.24, 0.26, and 0.11 for the supplier's profit comparison under one-price, two-price, and three-price treatments, respectively. Similarly, p-values for total profits are 0.31, 0.3, and 0.74, respectively. The buyer's profits are not significantly different under one-price and three-price treatments, with p-values of 0.42 and 0.84.

[7]To check the robustness of the significance results for the α parameter in Table 9, we repeated our analysis by removing the outliers. In particular, we removed players 1, 7, 14, and 18 from one-price human-to-computer treatment and players 1, 4, and 8 from the one-price human-to-human treatment. The α value is still significantly higher under human-to-human treatment at the 0.07 level. Similarly, if player 4 is removed from the two-price human-to-human treatment and the α values are compared for the human-to-human and human-to-computer experiments, the α value in the human-to-human treatment is significantly higher at the 0.04 level.

References

Altintas, N., F. Erhun, S. Tayur. 2008. Quantity discounts under demand uncertainty. *Manage. Sci.* **54**(4): 777–792.

Becker-Peth, M., E. Katok, U. Thonemann. 2013. Designing buyback contracts for irrational but predictable newsvendors. *Manage. Sci.* Advance online publication. doi: 10.1287/mnsc.1120.1662

Benton, W. C., S. Park. 1996. A classification of literature on determining the lot size under quantity discounts. *Eur. J. Oper. Res.* **92**(2): 219–238.

Boeing. 2008. Boeing Anaheim supplier payment information. Avialable at http://www.boeing.com (accessed date March 17, 2008).

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

Camerer, C., T.-H. Ho. 1999. Experience-weighted attraction learning in normal-form games. *Econometrica* **67**(4): 837–874.

Campbell, D. J. 1988. Task complexity: A review and analysis. *Acad. Manag. Rev.* **13**(1): 40–52.

Cui, T. H., J. S. Raju, Z. J. Zhang. 2007. Fairness and channel coordination. *Manage. Sci.* **53**(8): 1303–1314.

Fehr, E., K. M. Schmidt. 1999. A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**(3): 817–868.

Freund, R. J., W. J. Wilson, P. Sa. 2006. *Regression Analysis*, 2nd edn. Academic Press, Oxford, UK.

Guo, P., J. S. Song, Y. Wang. 2010. Outsourcing structures and information flow in a three-tier supply chain. *Int. J. Prod. Econ.* **128**(1): 175–187.

Ho, T.-H., J.-J. Zhang. 2008. Designing price contracts for boundedly rational customers: Does the framing of the fixed fee matter? *Manage. Sci.* **54**(4): 686–700.

Hogarth, R. M., H. J. Einhorn. 1992. Order effects in belief updating: The belief-adjustment model. *Cogn. Psychol.* **24**(1): 1–55,

Jucker, J. V., M. J. Rosenblatt. 1985. Single-period inventory models with demand uncertainty and quantity discounts: Behavioral implications and a new solution procedure. *Naval Res. Logist.* **32**(4): 537–550.

Kalkancı, B., K.-Y. Chen, F. Erhun. 2011. Contract complexity and performance under asymmetric demand information: An experimental evaluation. *Manage. Sci.* **57**(4): 689–704.

Katok, E., D. Y. Wu. 2009. Contracting in supply chains: A laboratory investigation. *Manage. Sci.* **55**(12): 1953–1968.

Kauffman, S. A. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, Oxford, UK.

Kayış, E., F. Erhun, E. L. Plambeck. 2012. Delegation vs. control of component procurement under asymmetric information and simple contracts. *M&SOM* **15**(1): 45–56.

Keser, C., G. Paleologo. 2004. Experimental investigation of retailer-supplier contracts: The whole-sale price contract. Working paper, IBM T. J. Watson Research Center, Yorktown Heights, NY.

Lim, N., T.-H. Ho. 2007. Designing price contracts for boundedly rational customers Does the number of blocks matter? *Market. Sci.* **26**(3): 312–326.

McKelvey, R. D., T. R. Palfrey. 1995. Quantal response equilibria for normal form games. *Games Econ. Behav.* **10**(1): 6–38.

Munson, C. L., M. J. Rosenblatt. 1998. Theories and realities of quantity discounts: An exploratory study. *Prod. Oper. Manag.* **7**(4): 352–369.

Rivkin, J. W. 2000. Imitation of complex strategies. *Manage. Sci.* **46**(6): 824–844.

Samuelson, W., R. Zeckhauser. 1988. Status quo bias in decision making. *J. Risk Uncertainty* **1**(1): 7–59.

Simon, H. A. 1955. Behavioral model of rational choice. *Q. J. Econ.* **69**(1): 99–118.

Simon, H. A. 1978. Rational decision-making in business organizations. Nobel Memorial Lecture, December 8.

Su, X. 2008. Bounded rationality in newsvendor models. *M&SOM* **10**(4): 566–589.

Supporting Information

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

Appendix A: Instructions for the Human-to-Human Experiments