

Trade Contingencies in Procurement Interactions*

Matthew J. Walker[†]

September 02, 2020

[JOB MARKET PAPER]

[LATEST VERSION HERE]

Abstract

I analyse the effects of a contingent contract in procurement interactions. The contract is allocated at a procurement auction and the trade contingency is implemented via an arbitration mechanism. With no contractual contingency, a seller moral hazard problem arises, and trade is inefficient. It is shown that there exists a contingent contract, defined by the degree of price flexibility and the probability of arbitration, for which trade is efficient in the unique equilibrium outcome. This outcome benefits the buyer, who holds the balance of market power, suggesting that commonly held beliefs about the fairness benefits of arbitration may not carry over to imbalanced market settings. I test the main predictions of the model in a laboratory experiment. Consistent with the theory, bidding is more aggressive and high quality is incentive compatible with a contingent contract. Yet the arbitrator crowds out buyer reciprocity and the contingent contract yields no significant increase in trade efficiency relative to a non-contingent contract. A sizeable fraction of buyers exhibit reference-dependent fairness preferences. The findings have implications for the establishment of trust and procurement efficiencies driven by competition.

Keywords: Procurement Auction, Contingent Contract, Arbitration, Moral Hazard, Fairness

JEL classifications: D44, D86, J52, C92

* This paper would not have been possible without the guidance and support of my advisor, Jason Shachat, to whom I am very grateful. I thank Daniel Zizzo, Kenju Kamei, Kyle Hyndman, Timo Heinrich and Matthew Embrey for valuable feedback and discussions. Funding from the Economic and Social Research Council Award No. ES/R500963/1 is gratefully acknowledged.

[†] Durham University Business School, Millhill Ln, Durham DH1 3LB; email: matthew.j.walker@durham.ac.uk.

1. Introduction

When potential sellers compete to win a contract, they have an incentive to price as low as possible. Price competition has well-recognised benefits. The use of auctions to procure goods and services yields significant cost-savings and advantages in terms of transparency and curbing favouritism. Yet competitive bidding can also lead to cost reductions by sellers during trade, to preserve profit margins, undermining trust in the procurement relationship (Fugger et al. 2019).

As a consequence, procuring relation-specific (as opposed to standardized) goods poses challenges to the buyer. Procurement contracts are often incomplete and contain a renegotiable, i.e. “flexible”, component. A common source of payment adjustments in the construction industry are deductions (Bajari et al. 2014). Deductions reduce the seller’s final compensation. The buyer may deduct damages in response to seller negligence, which the seller may dispute. In many instances, recourse to the courts to resolve disputes is not practical in terms of time or money. Rent-seeking behaviours may encourage costly pre-trial settlements that reduce welfare (Cooter et al. 1982). An alternative and frequently invoked dispute resolution method is arbitration.

Arbitration is a type of contingent contract, in which the contingency is the arbitrator’s decision (Bazerman and Gillespie 1999). The arbitrator is an independent third party, with the power to impose binding settlements in the event that negotiations break down.¹ The arbitrator’s role is “to remember who sunk costs in the past, and to ensure that future compensation is paid in a way that creates the *correct* incentives to make such investments” (Crawford 1985, p 375).²

This paper examines the effects of a contingent contract, implemented via an arbitration mechanism, in procurement interactions. The model developed here is a stylized representation of the following scenario. A buyer seeks to procure a seller to deliver a contract (e.g. for construction or engineering works) via competitive bid process. After allocation of the contract, the seller receives some pre-agreed proportion of the winning bid up front and takes costly actions, before final payments – inclusive of any deductions – are made. Where disputes arise on buyer or seller performance, an arbitrator is invoked to provide binding resolution, either by

¹ The arbitration process is distinct to mediation (or other forms of amicable dispute resolution), in which an independent person facilitates the parties to negotiate an agreement and identify options but has no power to render a judgement.

² Arbitration finds varied applications for resolving contractual disputes. Perhaps the most well-known is for the settlement of wage disputes in labour bargaining. Mandatory consumer arbitration clauses are also prevalent. The eBay user agreement contains a requirement to submit dispute claims, e.g. due to an unpaid auction bid, to binding and final arbitration (<https://www.ebay.com/help/policies/member-behaviour-policies/user-agreement?id=4259>). Legal scholars posit one reason for the use of mandatory consumer arbitration agreements as a nefarious attempt by corporations to prevent collective actions (Eisenberg et al. 2008). This claim is debated (Drahozal and Ware 2010) and we do not consider it further here.

mutual agreement or a clause written into the contract.³ The arbitrator must verify the claims *ex post* and adjudicate on how best to compensate the trade parties.

The starting point for the model is a first-price procurement auction. This is the dominant bidding mechanism in procurement – second-price mechanisms are rarely used (Engelbrecht-Wiggans et al. 2007). At date 0, multiple potential sellers compete to deliver a relation-specific product to the buyer and the lowest bid wins. In a departure from standard first-price auction setups, the winning bid determines the bounds of the contract price range, rather than the final trading price. The degree of price flexibility is an exogenous parameter in the contract. If the buyer accepts the outcome of the bidding process, a bilateral trade relationship forms at date 1. The seller then determines the product quality level, high or low. High quality is associated with a higher seller delivery cost than low quality but also generates a larger trade surplus. After observing her value from trade, the buyer has discretion to propose a final trading price out of the date 0 contract price range.

The main contribution of this paper is to embed a contingency into the contract. This is achieved via an exogenous arbitration clause. If the buyer's proposed final price is below some reference notion of a fair outcome, then the seller has the option to dispute the final price and call upon an arbitrator to adjudicate. No restriction is placed on what the arbitrator deems a fair settlement: the arbitrator's preferences are a known function of the trade surplus. This mirrors the arbitrator's role in real-world dispute resolution. The arbitrator is not always available to intervene. The probabilistic nature of the arbitrator reflects the reality that arbitration is not feasible in every trading scenario.

Standard theory predicts that a non-contingent contract produces inefficient (low quality) procurement outcomes (see also Brosig-Koch and Heinrich 2014, Fugger et al. 2019, Katok et al. 2020). Consistent with empirical observation (e.g. Bajari et al. 2014), potential sellers strategically adjust their bids to account for the possibility of payment deductions. There is a contingent contract, defined by the combination of price flexibility and arbitration probability, for which efficient (high quality) trade is the unique equilibrium outcome. The outcome is first-best in the absence of direct arbitration costs and is associated with more aggressive auction bids. A striking implication of the model is that the contingent contract benefits the buyer, but not the

³ In the US, standard construction contracts are provided by the American Institute of Architects, available at <https://www.aiacontracts.org/>. Until 2007, arbitration was the preferred dispute resolution method by default. Since 2007, the contract requires trade parties to check a box specifying the preferred method (e.g. Document A101-2017). Construction case filings at the American Arbitration Association (AAA) continue to increase, up 6% in 2019 (AAA Annual Report, available at <https://www.adr.org/>). Standard building contracts in the UK are produced by the Joint Contracts Tribunal, which also provides rules for the resolution of disputes by arbitration, available at <https://www.jctltd.co.uk/>. A motion to resolve payment disputes in the construction industry via a formal arbitration scheme was recently debated in UK parliament (Construction Retention Deposit Schemes Bill 2017-19).

seller, in equilibrium. That is, when the buyer holds the balance of market power, arbitration serves an efficiency rather than a fairness function.

Uncertainty is a necessary condition to realize the efficiency benefits of a contingent contract. Certain arbitration cannot mitigate the seller moral hazard problem. If the arbitrator is always available, then the seller always has an incentive to marginally undercut his competitor on price at date 0 and deliver low quality. A probabilistic arbitrator, however, changes the seller's marginal bidding incentives. For a given bid, a potential seller may prefer to deliver high quality than to undercut his competitor and deliver low quality, if he expects to be rewarded sufficiently by the arbitrator in relation to the associated costs.

An established literature in experimental economics recognises the potential for explicit incentives to be counterproductive (for a survey of early contributions, see Fehr and Gächter 2000). I therefore test the main predictions of the model using a laboratory experiment. The lab has found long-standing use as a test-bed for more general arbitration mechanisms, because it enables control over factors such as the arbitrator's preferences, costs, and valuations (e.g. Ashenfelter et al. 1992, Bolton and Katok 1998, Pecorino and Van Boening 2001, Deck and Farmer 2007, Deck et al. 2007, Gabuthy et al. 2008). These factors are unobserved in naturally occurring settings. To my knowledge, no prior experiment has analysed the effect of an arbitration mechanism in an auction context, nor with relationship-specific investments.

The experiment results are summarised as follows. Bidding is more aggressive in the presence of an arbitrator, but a contingent contract does not outperform a non-contingent contract from an efficiency standpoint. The contingent contract crowds out buyer reciprocity, although high quality remains on average the most profitable seller strategy. In contrast to the theory, there is some evidence that a contingent contract does improve the seller's position when the arbitrator prefers to award the seller most of the trade surplus. This finding suggests that arbitration can serve a fairness function in competitive procurement interactions, although more data is required to draw firm conclusions. We also uncover that a significant fraction of fair buyers in our experiment display reference-dependent preferences.

The paper continues in Section 2 with a discussion of related studies in disparate literatures on arbitration, renegotiable contracts and procurement auctions. Section 3 outlines the model and equilibrium predictions, with and without a contractual contingency. Section 4 describes the experiment and hypotheses to be tested. Section 5 summarizes the main experiment findings and presents a formal analysis of behaviour at the aggregate and individual levels. Section 6 concludes with a consideration of the external validity of the model and empirics.

2. Arbitration, Renegotiable Contracts and Procurement Auctions

Traditionally, the economics literature has examined distributional implications of arbitration, under the assumption of a fixed surplus to divide (e.g. Farber and Katz 1979, Farber 1980, Crawford 1981, Ashenfelter and Bloom 1984). Gabuthy and Muthoo (2019) drop the fixed surplus assumption. They show that the mere presence of an arbitrator with known preferences can improve incentives of negotiating parties in a bilateral trade relationship to make mutual relationship-specific investments. The arbitrator is called upon to adjudicate disputes with a certain probability. This probability is agreed by the parties in a first stage. Any dispute outcome not settled by the arbitrator is welfare-decreasing. In a second stage, hold-up concerns undermine incentives to make a sunk investment relative to the first best. The size of the inefficiency is influenced by the arbitration probability. In the optimal arrangement, the bargaining powers of the two sides are equalised and rent seeking behaviour is deterred. We consider the efficiency consequences of an exogenous arbitration mechanism, but in a competitive bidding environment.

There is also a connection between our study and economic experiments testing whether renegotiable contracts can solve the hold-up problem. Theory predicts that flexible price contracts are only effective when renegotiation is not possible. Hoppe and Schmitz (2011) report evidence from a lab experiment that the threat of seller aggrievement alone is sufficient to deter opportunistic price revisions by buyers and significantly improve trade outcomes relative to the no-contract benchmark. These findings are consistent with the behavioural contract-theoretic model of Hart and Moore (2008), in which pre-agreed contractual terms act as a reference point in the trade relationship. Further experimental studies in this area are Fehr et al. (2011, 2015), Bartling and Schmidt (2015) and Brandts et al. (2015). Davis and Leider (2018) analyse the contractual hold-up problem in a supply chain setting. Empirically, they observe that option contracts outperform wholesale price contracts, quantity commitment contracts and multiple part pricing across various demand parameterisations.

Like the flexible contract considered by Hart and Moore (2008), the final trading price in the setup considered in this paper is selected from a price range determined under competitive conditions. An important difference between the setup here and the model of Hart and Moore is the timing of actions. Here, the buyer proposes a final trading price after, rather than before, the seller's quality decision, thereby precluding an aggrieved seller from subsequently reducing the product quality at private cost. In this sense, our setup is closer in nature to experimental labor market studies on gift-exchange (e.g. Fehr et al. 1993, Brown et al. 2004), in which high quality is more costly for the seller to deliver than low quality.

The present study is further related to a literature on procurement auctions with renegotiation. Herweg and Schwarz (2018) consider a contingency not specified in the contract

(cost overruns). In their setup, prices can increase relative to the initial contract. Suppliers are endowed with a private cost type, which is observed by the buyer after the auction and before contract renegotiation takes place. The buyer can specify one of two designs during the initial auction, which differ according to their complexity. The cost advantage of the efficient supplier type is increasing in the design complexity. The generalized Nash bargaining solution implies that the buyer should minimise this cost differential by specifying the less complex design up front. The final price is higher if either the cost of the more complex design is higher for the winning supplier, or if the winning supplier is endowed with enough bargaining power. The outcome is always efficient even though the buyer cannot commit not to renegotiate. Their finding of higher prices after auction renegotiation is consistent with Chang et al. (2016), who outline a model in which there is a common unknown production cost component and suppliers are differentiated by wealth constraints.⁴ The ability to renegotiate helps those suppliers who can credibly threaten to default on the contract.

Using a mechanism design approach, Herweg and Schmidt (2020) show that a two-stage auction process relying on an independent arbitrator to verify the payoff consequences of trade contingencies *ex post* can implement an efficient outcome, even when the set of all possible contingencies is not common knowledge. This stands in contrast to the usual assumption in the mechanism design literature that all model parameters are common knowledge when implementing an allocation *ex ante*. The assumption of *ex-post* verifiability enables the arbitrator to complete the mechanism by separating the problems of inducing truthful reporting of design flaws when they are least costly to correct and truthful reporting of production costs. As in the present study, this assumption is motivated by the established role that arbitrators have in real-world dispute resolution of complex goods.

Procurement auctions with moral hazard have received attention in the behavioural operations management literature.⁵ Earlier studies focus on fixed-price contracts in which quality cannot be conditioned on price, but where the seller moral hazard problem can be overcome using either a reputational mechanism (Brosig-Koch and Heinrich 2014) or a buyer-determined mechanism in which the buyer is no longer constrained to choose the lowest auction bid (Fugger et al. 2019). Closely related is Katok et al. (2020), who introduce a flexible-price contract via retainage incentives. Their focus is on the level of retainage in the contract, i.e., the proportion of the contract price withheld by the buyer from the seller after the auction and before product delivery. Using lab experiments, they observe that voluntary retainage incentives can mitigate the

⁴ Shachat and Tan (2015) consider a form of post-auction price renegotiation that can lead to lower prices. Waehrer (1995) and Wang (2000) also examine auction models in which the procurer can re-negotiate with the winning bidder and lower prices emerge as a result.

⁵ For a more comprehensive survey of this literature, see Elmaghraby and Katok (2018).

seller moral hazard problem, as captured by a significant increase in product quality, relative to a fixed-price contract. This finding is explained using a model of fair payment norms, in which a known fraction of buyers in the population are trustworthy and reciprocate quality received with a fair payment. By restricting the bid space, they also show that procurement arrangements in which the retainage level is set too high can lead sellers not to participate in the contracting process. The present paper builds on this study, by incorporating a contingency into the contract.

3. Model and Theory

3.1 Basic setup

I first outline the baseline model with a non-contingent contract. A single buyer seeks to purchase one unit of an indivisible product from a pool of N potential sellers, indexed by i . Thus, sellers compete to deliver the unit. At date 0, the buyer may choose to purchase from the potential seller submitting the lowest priced (winning) bid or reject the opportunity to trade. A decision to purchase signals the beginning of a bilateral trade relationship with the winning potential seller (henceforth the seller). In that case, at date 1 the product quality is determined by the seller and the winning bid is renegotiable. Payoffs are realized at date 2.

The buyer's payoff equals her valuation for the product v minus the final trading price p . The seller's payoff is given by the difference between the final price p and his delivery cost c . The seller's quality choice q maps directly to the valuation and cost schedules, which are common knowledge. The cost schedule is the same across potential sellers and so they are *ex-ante* homogeneous. We assume that price cannot be conditioned on quality at date 0 but quality can be (costly) verified by third parties at date 1. *Ex-post* verifiability is not unrealistic: third-party arbitrators must be able to verify work to settle real-world contractual disputes (see Herweg and Schmidt 2020).

The payoffs of the buyer and seller are summarised as follows,

$$\pi_B = v(q) - p. \quad (1)$$

$$\pi_S = p - c(q). \quad (2)$$

A potential seller that loses the contract at date 0 earns zero payoff. A buyer that rejects the opportunity to purchase at date 0 ends the interaction with zero payoff to all agents.

If trade occurs, then at date 1 the seller can choose to deliver either low product quality ($q = q^L$) or high product quality ($q = q^H$). By discretizing the quality space, we simplify the analysis and restrict attention to contractual performance at the extensive margin. A seller's cost of delivering high quality is strictly greater than his cost of low quality, $c(q^H) > c(q^L)$. Trade is preferred to no trade and high quality maximises surplus, $v(q^H) - c(q^H) > v(q^L) - c(q^L) > 0$.

The seller's quality choice is taken before the final trading price p is determined, but after the *range* of contract prices is determined. The contract price range is given by the interval $[\underline{p}, \bar{p}]$. The upper bound contract price is equal to the winning bid at date 0, $\bar{p} = \min\{b_1, \dots, b_N\}$. The lower bound contract price is determined by the flexibility in the procurement contract, defined by an exogenous parameter $\lambda \in (0,1)$, with $\underline{p} = (1 - \lambda)\bar{p}$. We exclude the boundary cases of $\lambda = 0$, which constitutes a non-renegotiable contract, and $\lambda = 1$, for which bidding is uninformative cheap talk. This specification of the price bounds is without loss of generality. The bid might just as well correspond to the lower bound contract price, with a fixed interval above, and leave the model's predictions unaffected. The important thing is that there is flexibility in the contract price.

The structure of the sequential game proceeds as follows.

Date 0: Contracting.

Stage 1. The potential sellers each submit one sealed bid b_i at a first-price procurement auction. Bids can be greater than or equal to the seller's cost of delivering low product quality. The potential seller that submits the lowest priced bid wins the auction and becomes the seller. If two or more potential sellers submit the same bid, the tie is broken randomly. The winning bid determines the contract price range. The bid profile is public information before the next stage. Any potential seller that submits a losing bid is no longer considered in the interaction.

Stage 2. The buyer observes the contract price range and takes a decision from the action set $A \in \{a^0, a^1\}$, where a^0 represents a decision to reject and a^1 to accept the opportunity to purchase from the seller. If the buyer decides to reject, the interaction ends without a trade.

Date 1: Trade

Stage 3. The seller chooses the product quality level, either low (q^L) or high (q^H) and incurs his delivery cost $c(q)$.

Stage 4. The buyer observes her value for the product and proposes a final trading price from the contract price range.⁶

Date 2: The trade is completed, and payoffs are realized.

Formally, a potential seller i 's strategy has two components: a bid b_i from a discrete price grid $G = \{c(q^L), c(q^L) + \Delta, \dots\}$, where Δ is a small increment, and a quality choice function $q_i(b_i | b_i = \bar{p}, a = a^1)$. The use of a discrete price grid reflects that the (necessary) discrete experiment implementation influences the point prediction when there are just two bidders. A

⁶ The analytical results would be unchanged by switching the order of stages 3 and 4.

buyer's strategy also has two components: a trade decision function, $a(\bar{p})$, and a price proposal function, $y([\underline{p}, \bar{p}], q | a = a^1)$.

The game can be analysed as an extensive-form game of complete information. We are interested in whether there exists a price at which high product quality is implementable in our environment. High quality is implementable if it is preferred to low quality (henceforth, incentive compatible) as part of an equilibrium seller strategy. The solution concept is subgame perfect Nash equilibrium (SPNE) in pure strategies and we apply backward induction. As in standard Bertrand competition models with bounded demand, complete information, fixed marginal costs and zero fixed costs, there is no mixed strategy equilibrium (see e.g. Baye and Morgan 1999).

For the basic setup, there is a unique and inefficient SPNE: $b_i^* = c(q^L)/(1 - \lambda) + \Delta$ for all i or $b_i^* = c(q^L)/(1 - \lambda)$ for at least two i ; $a^* = a^1$; $q_i^* = q^L$; and $y^* = \underline{p}$.

At date 1, the buyer uses her payment discretion to choose the lower bound contract price, independently of the seller's chosen quality level. The seller anticipates this and so delivers low quality to minimise his cost. At date 0, the buyer will accept to purchase at any bid less than or equal to $v(q^L)/(1 - \lambda)$, in anticipation of a non-negative profit. If two or more potential sellers submit a bid equal to $c(q^L)/(1 - \lambda)$, i.e., the cost of low quality marked up in proportion to the degree of price flexibility, then every potential seller earns a profit of zero. Any deviation would also yield zero profit. If all potential sellers submit a bid equal to $c(q^L)/(1 - \lambda) + \Delta$, then each makes a (small) positive expected profit equal to Δ/n . Any deviation would yield zero profit; thus, no potential seller has an incentive to deviate. All trade surplus accrues to the buyer. Our first proposition follows directly.

Proposition 1. With a non-contingent contract, high quality is not an implementable procurement outcome.

Conditional on expectations about population reciprocity, models of social preferences can rationalise high product quality in our environment with a non-contingent, but flexible-price, contract (see Katok et al. 2020 for details). This will serve as the benchmark against which a contingent contract is assessed in the experiment.

3.2 Arbitration

We now write a contingency into the procurement contract, which allows the seller to rely probabilistically on an independent third-party arbitrator to settle any dispute over the buyer's claim on the surplus. An additional stage is included at date 1 as follows.

Stage 5. The seller observes the buyer's price proposal and takes a decision from the action set $D \in \{d^0, d^1\}$, where d^0 represents a decision not to dispute and d^1 a decision to dispute via arbitration.

There is now a third component to a seller i 's strategy: a dispute function $d_i([p, \bar{p}], q_i, y | a = a^1)$. If the seller opts not to dispute, he receives the buyer's proposed trading price and the buyer and seller payoffs are captured by equations (1) and (2). If a dispute is recorded, the arbitration clause is invoked. The probability of arbitration, σ , is pre-determined. With probability $(1 - \sigma)$, the arbitrator is not called upon to adjudicate the dispute and the seller still receives the buyer's price proposal. With probability σ , the arbitrator is called upon and imposes a final trading price based on the following decision rule,

$$p^A(q, y | \mu) = \max\{y, z^q(\mu)\}, \quad (3)$$

where $0 < \mu < 1$ and $z^q(\mu) = \min\{\bar{p}, \mu v(q) + (1 - \mu)c(q)\}$.

The parameter μ is common knowledge to both trade parties prior to date 0 and captures the arbitrator's preference over the share of transaction surplus due to the seller. This corresponds to what the arbitrator deems to constitute a fair settlement after completion of the fact-finding process. Consistent with the arbitration literature, μ is exogenous and a function of how an unbiased arbitrator reaches a decision (see Gabuthy and Muthoo 2019). We abstract from the possibility that μ itself might be quality-specific. Such a formulation would not qualitatively change the arguments, so long as the arbitrator's preferences are commonly known.

If $z^q(\mu)$ is greater than y , the arbitrator settles in favour of the seller and awards a final trading price which is non-decreasing in both the buyer's valuation for the product and the seller's cost incurred. Under this scenario, a fixed and non-negative cost k of arbitration is borne by the buyer. We restrict k to be less than the trade surplus, since this is the only interesting case. If $z^q(\mu)$ is less than or equal to y , the arbitrator settles in favour of the buyer and awards the seller the buyer's proposed price. The cost of arbitration is then borne by the seller.⁷

Let events A and B represent settlement in favour of the buyer and seller respectively. The expected dispute payoffs of buyer and seller are as follows,

$$d_B = v(q) - (1 - \sigma)y - \sigma(p^A + \mathbf{1}_B k). \quad (4)$$

⁷ In practice, arbitration costs may be borne by the unsuccessful party or shared between the dispute parties. To simplify the exposition, we employ the former rule here, but for reasons that will become clear the main insights of the model are unchanged by using a shared cost allocation. Common knowledge about the arbitration cost has real-world application. The German Arbitration Institute, for example, makes its cost schedule publicly available and provides a cost calculator tool on its website (<http://www.disarb.org/en/>).

$$d_S = (1 - \sigma)y + \sigma(p^A - \mathbf{1}_A k) - c(q). \quad (5)$$

We assume risk neutrality primarily to simplify the notation. The results of the model are qualitatively unchanged by substituting into (4) and (5) a constant relative risk aversion (CRRA) utility function, $u(\pi) = \pi^{(1-r)}$ for $\pi \geq 0$ and $u(\pi) = -(-\pi)^{(1-r)}$ for $\pi < 0$, where r measures the degree of relative risk aversion, $r < 0$ implies risk seeking, $r = 0$ implies risk neutrality, and $r > 0$ implies risk aversion.⁸ The expected dispute utilities remain monotonic over the discrete price grid. We show later that the equilibrium predictions still hold in our experiment if we assume risk aversion. For the rest of this section, we stay with the risk neutral case.

At Stage 5, the quality level has already been chosen by the seller. It is never in the interest of a seller to dispute if the buyer proposes a price greater than or equal to $z^q(\mu)$, because he would bear the direct cost of employing the arbitrator if called upon at no monetary gain. In that case, $d_i = d^0$. For any lower price proposal, the seller should dispute since the buyer would bear the direct cost. In this case, $d_i = d^1$. Therefore, we assign conditional probability zero to event A and conditional probability one to event B (conditional on dispute).

At Stage 4, the buyer anticipates this outcome and, having observed the delivered quality, proposes a trading price as follows,

$$y = \begin{cases} \underline{p}, & \text{if } k < \frac{(1-\sigma)}{\sigma} (z^q(\mu) - \underline{p}) \\ \max\{\underline{p}, z^q(\mu)\}, & \text{if } k > \frac{(1-\sigma)}{\sigma} (z^q(\mu) - \underline{p}). \end{cases} \quad (6)$$

At high costs of arbitration k , the buyer prefers to propose a price according to the arbitrator's preferences. At low such costs, she proposes the lower bound price \underline{p} . Since the buyer's dispute payoff is strictly decreasing in p , it is never in her interest to propose a price between \underline{p} and $z^q(\mu)$.

At Stage 3, the seller chooses the quality level. If the winning bid is high enough that $z^H(\mu)$ is less than or equal to \underline{p} , then the arbitrator has no influence on the analysis and high quality is not incentive compatible. The incentive compatibility constraint for high quality is composed of the union of three cases (a proof is contained in Appendix A):

Case (i): $k \geq \frac{(1-\sigma)}{\sigma} (z^H(\mu) - \underline{p})$.

High quality is incentive compatible if $z^H(\mu) - \max\{\underline{p}, z^L(\mu)\} > c(q^H) - c(q^L)$.

Case (ii): $\frac{(1-\sigma)}{\sigma} (z^L(\mu) - \underline{p}) < k < \frac{(1-\sigma)}{\sigma} (z^H(\mu) - \underline{p})$.

⁸ When $r = 1$, the function is replaced by $\ln d$. When $r > 1$, the function must be divided by $(1 - r)$ to preserve increasing utility.

High quality is incentive compatible if $\sigma > \frac{c(q^H) - c(q^L) + \max\{\underline{p}, z^L(\mu)\} - \underline{p}}{z^H(\mu) - \underline{p}}$.

Case (iii): $k \leq \frac{(1-\sigma)}{\sigma} (z^L(\mu) - \underline{p})$.

High quality is incentive compatible if $\sigma > \frac{c(q^H) - c(q^L)}{z^H(\mu) - z^L(\mu)}$.

We infer the following insight: at high arbitration costs, above the threshold defined by case (i), incentive compatibility does not depend on the arbitration probability; at lower arbitration costs, as defined by cases (ii) and (iii), high quality satisfies the incentive compatibility constraint if and only if the probability of arbitration is large enough.

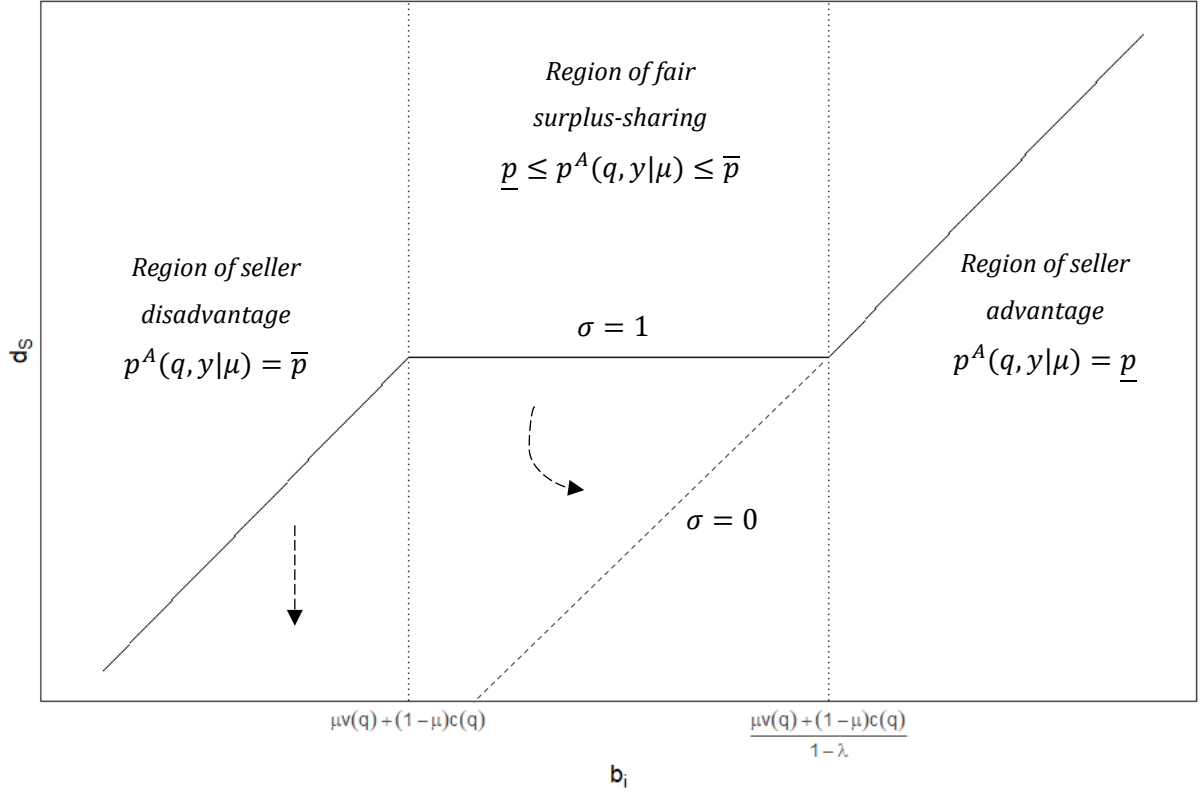
At Stage 2, having observed the contract price range, the buyer anticipates the seller's quality choice and accepts to purchase whenever her own expected payoff is non-negative, since the payoff to rejection is zero. Since the arbitrator cannot award a price below the buyer's proposal, the buyer will accept any winning bid that does not exceed $v(E[q; \bar{p}])/(1 - \lambda)$, where the expectation of quality is determined by the incentive compatible quality level given the probability of arbitration σ and the arbitrator's preference parameter μ . The buyer will reject any bid above this.

We now turn to the bidding incentives at Stage 1. The benchmark for payoff comparison is zero, which is a losing seller's payoff at auction. From (3), the arbitrator's award is constrained by the upper bound contract price \bar{p} . At a high enough arbitration cost k , as defined by the cost threshold in (6), a potential seller always has an incentive to marginally undercut his competitor at auction and increase his probability of winning from $1/n$ to one. When the upper bound contract price \bar{p} is less than $z^q(\mu)$, the seller is guaranteed a trading price equal to \bar{p} . In this cost case, by standard Bertrand arguments, the winning auction bid equals the cost of low quality, or one increment above. High quality is not incentive compatible, and we return to the inefficient outcome analogous to that obtained with a non-contingent contract.

Corollary 1. At high arbitration costs, high quality is not an implementable procurement outcome.

Suppose now that the arbitration cost k is low enough to ensure that the buyer proposes a trading price equal to \underline{p} . Without loss of generality, we can normalize the arbitration cost to zero. In Figure 1, we sketch the seller dispute payoff $d_S(q)$, as a function of seller i 's bidding strategy. The solid line in the figure corresponds to the payoff function given certain arbitration ($\sigma = 1$). This payoff function is quality specific and has two pivot points.

Figure 1. Expected dispute payoff as a function of seller i 's bid.



The lower pivot point in the figure is at a bid equal to $\mu v(q) + (1 - \mu)c(q)$. As bids fall below this level, the arbitrator's preferred settlement cannot be implemented. From (3), the arbitrator imposes a final price equal to the upper bound contract price \bar{p} . I denote this the *region of seller disadvantage*. The upper pivot point is at a bid equal to $[\mu v(q) + (1 - \mu)c(q)] / (1 - \lambda)$. A bid above this level is high enough to guarantee the seller a greater share of the surplus than the arbitrator deems fair. The arbitrator imposes a final price equal to the lower bound contract price \underline{p} . Thus, the arbitrator has no influence on the interaction at bids in this region. I denote this the *region of seller advantage*. For bids between the two pivot points, the arbitrator's preferred settlement is located within the contract price range and so the arbitrator can impose a final trading price to achieve this settlement. I denote this the *region of fair surplus-sharing*.

We can examine comparative statics for μ , λ and σ . At larger values of μ , both the pivot points in Figure 1 shift to the right. Intuitively, a higher bid is required to allocate the seller a greater share of the surplus and the dispute payoff in the *region of fair surplus-sharing* is located at a higher level. At greater degrees of price flexibility λ , there is a wider contract price range. The payoff function shifts down in all regions, due to the larger potential for seller losses if the arbitrator is not called upon following a dispute. The upper pivot point also shifts right and so the *region of fair surplus-sharing* incorporates a wider bid interval.

Let us turn to the probability of arbitration σ . We know from our analysis of Stage 3 that high quality is only incentive-compatible if σ is large enough. Yet this does not tell us whether

high quality can be part of an equilibrium seller strategy. Figure 1 again provides insight. As σ decreases below one, the payoff function in the *region of seller disadvantage* shifts downwards and the payoff function in the *region of fair surplus-sharing* rotates gradually anti-clockwise. Thus, σ determines the slope of the payoff function in the *region of fair surplus-sharing*. When $\sigma = 0$, the two segments form a straight line and the *region of fair surplus-sharing* collapses to a single price. This scenario is shown by the dashed line in the figure.

Incentive compatibility corresponds to the bid regions in which the seller dispute payoff to high quality is above the payoff to low quality. When $\sigma = 0$, the two quality-specific payoff functions never intersect and the payoff function to low quality is strictly above the payoff function to high quality. This led us to the inefficient outcome described in *Proposition 1*. When $\sigma > 0$, the two quality-specific payoff functions either never intersect or they intersect at two points in the price grid G . For high quality to be incentive compatible, we require these payoff functions to intersect.

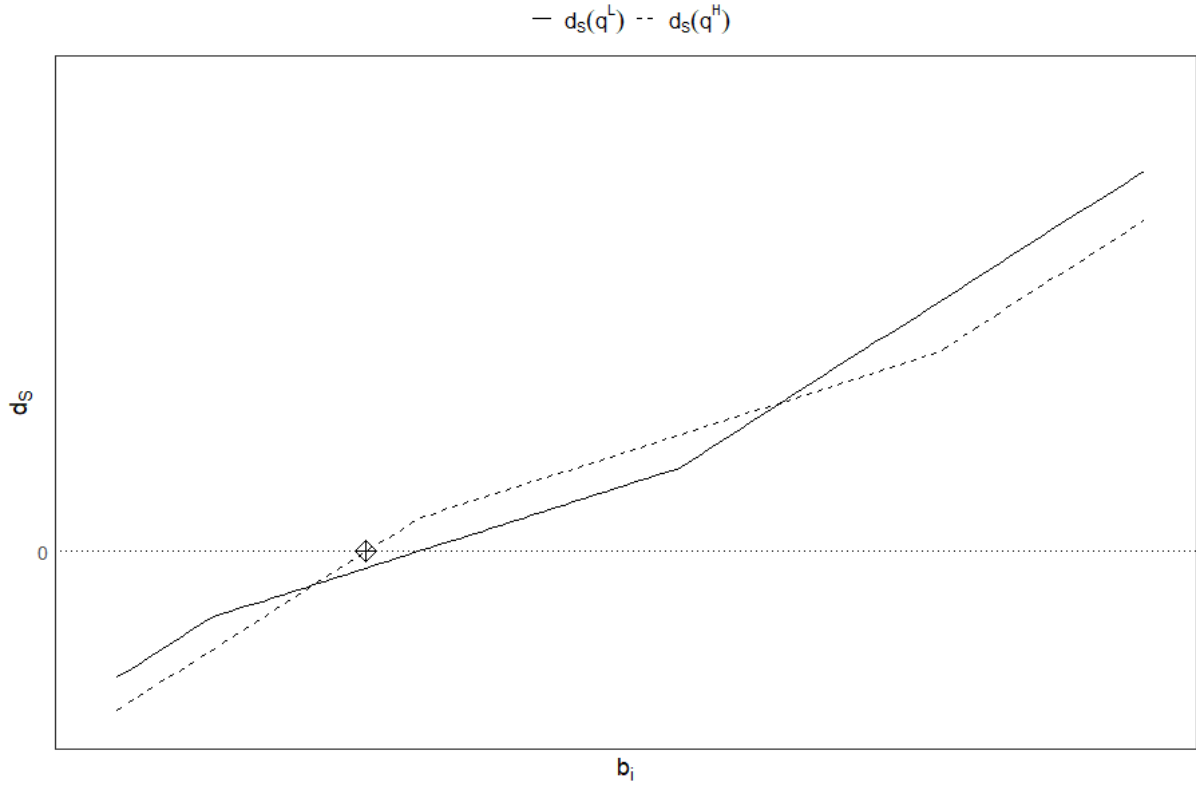
Due to monotonicity of the dispute payoff function, the low quality payoff must be strictly less than zero for some non-empty subset G' of the price grid. For high quality to be an implementable equilibrium outcome, we require that there exists at least one bid within the subset G' for which a seller strategy with a high quality component yields him a non-negative expected payoff. This intuition is formalised in the following condition.

Equilibrium Implementability Condition. $\exists b_i \in G: d_S(q^L) < 0 \leq d_S(q^H)$.

At any bid in G' , the arbitrator would settle in favour of the seller. The lowest bid in G' must be $\underline{b} = \frac{c(q^H) - \sigma z^H(\mu)}{(1-\lambda)(1-\sigma)}$, obtained by setting the seller's dispute payoff to high quality equal to zero.

The *Equilibrium Implementability Condition* is sufficient for the existence of a unique SPNE associated with high quality. In each SPNE, we have $b_i^* = \underline{b} + \Delta$ for all i or $b_i^* = \underline{b}$ for at least two i ; $a^* = a^1$; $q_i^* = q^H$; $y^* = \underline{p}$; and $d_i^* = d^1$. Potential sellers' no deviation incentives during bidding are the same as described in the basic setup, except that now the breakeven bid depends on the cost of high quality and accounts for the probability of arbitration and the arbitrator's preferences in addition to the degree of price flexibility. Since the seller dispute payoff is monotone increasing in price, if the *Equilibrium Implementability Condition* is satisfied, then there exists no alternative equilibrium associated with low quality, i.e., this equilibrium. The buyer achieves the full (larger) surplus in expectation, although her payoff is reduced by the arbitration cost. Seller payoffs are unchanged in expectation from the case of a non-contingent contract.

Figure 2. *Equilibrium Implementability Condition.*



With zero cost of arbitration then, the first-best outcome can be achieved. At any positive arbitration cost, a pareto-improving second-best outcome can be achieved when the cost of arbitration is less than the gains from trading a high quality product.⁹ Our second proposition follows directly.

Proposition 2. Probabilistic arbitration can mitigate the seller moral hazard problem. There exists a contingent contract defined by λ and σ for which high quality is the unique implementable procurement outcome.

Since \underline{b} is at the point where the payoff function to high quality crosses the zero profit line and the slope of the payoff function in the *region of fair surplus-sharing* is shallower than its slope in the other two regions, we can state the following corollary.

Corollary 2. Any contingent contract that implements a high quality procurement outcome is associated with lower equilibrium bids than a non-contingent contract.

⁹ If the seller were to share a fixed proportion α of the arbitration cost, then she would only dispute at Stage 5 if the expected gain from doing so exceeds αk . The seller dispute payoff function intercept would shift upwards and the breakeven bid would adjust accordingly. Sufficiency of the *Equilibrium Implementability Condition* still applies.

As σ approaches one, the buyer's claim on the surplus becomes irrelevant. The seller dispute payoff to low quality remains greater than or equal to zero at all acceptable bids and so the *Equilibrium Implementability Condition* is never satisfied. For any bid strictly above $c(q^L) + \Delta$ in the discrete price grid G , each seller has an incentive to marginally undercut his competitor – analogous to the case with a non-contingent contract – so long as the arbitrator prefers to award the seller a non-negative share of the trade surplus. We obtain a final corollary.

Corollary 3. Certain arbitration cannot mitigate the seller moral hazard problem; high quality is not an implementable procurement outcome.

4. Experiment Setup

4.1 Design

I design a controlled experiment to test the main predictions of the model. In the experimental setting, one buyer faces two potential sellers. The degree of price flexibility λ is set exogenously at 0.75. This permits substantial flexibility in the contract. There are three experimental conditions, which vary according to the probability of arbitration σ and the arbitrator's preference parameter μ (see Table 1). The variable σ is implemented using a between-subjects design. The variable μ is implemented using a within-subjects design.

Table 1 – Treatment matrix.

Treatment	<i>Voluntary</i>	<i>Arbitrator</i>	
Sub-treatment		$\mu = 1/3$	$\mu = 2/3$
Probability of arbitration	$\sigma = 0$	$\sigma = 0.50$	
Price flexibility	$\lambda = 0.75$		

The first condition is the baseline non-contingent contract and so the probability of arbitration is zero. We call this the *Voluntary* treatment. Any payment to the seller above one-quarter of the winning bid is made by the buyer voluntarily in this treatment. In the second and third conditions, the probability of arbitration is one-half. We refer to these two conditions collectively as the *Arbitrator* treatment. The two conditions define two sub-treatments in which the arbitrator either prefers to award one-third or two-thirds of the trade surplus to the seller. By varying the arbitrator's preference parameter within subjects, we conduct a less restrictive test of the arbitration model and consider trade relationships in which the arbitrator favours the

buyer and seller. We deliberately avoid invoking the fairness norms associated with an equal split, which have been observed in prior bargaining studies (e.g. Andreoni and Bernheim 2009).¹⁰

The cost and valuation schedules are common knowledge to all subjects in the experiment (Table 2). Low quality generates a trading surplus of twenty and high quality generates a trading surplus of sixty. Low quality costs the seller thirty to deliver and is valued by the buyer at fifty. High quality costs the seller forty to deliver and is valued by the buyer at one-hundred. Consistent with the investment game literature in experimental economics, the surplus multiplier is three. The bid increment Δ is set at one, the minimum bid at thirty and the maximum at two-hundred. The maximum bid is high enough to ensure that potential sellers can always submit a profitable bid associated with either low or high quality. The maximum bid is low enough to ensure that there is no winning bid at which a self-interested buyer would reject the opportunity to purchase.

Table 2 – Cost and valuation schedules.

Quality level q	$q = q^L$	$q = q^H$
Buyer's valuation: $v(q)$	50	100
Seller's cost: $c(q)$	30	40

The *Equilibrium Implementability Condition* is satisfied in *Arbitrator* for all levels of risk aversion (this is shown in Appendix B). The direct cost of arbitration is set at $k = 2$, which is low enough to ensure that the buyer's equilibrium proposal is always the minimum amount in the contract price range. Our parameter choices are without loss of generality and in Appendix B we also demonstrate the implementable equilibrium quality levels for the full set of possible contracts in our experiment environment. The set of contracts for which high quality is an equilibrium bidding strategy component is decreasing in the degree of seller risk aversion.

All subjects participated in a sequence of thirty trade interactions. Each interaction was divided into distinct phases. There were two phases of every interaction in *Voluntary* treatment and three phases of every interaction in *Arbitrator*. The first two phases were the same across the treatments. In the first phase, the potential sellers in a group simultaneously submitted bids at a first-price procurement auction. At the same time as choosing a bid, each potential seller also chose a quality level, either high or low, to be delivered conditional on submitting the winning

¹⁰ Aside from the normative appeal of the equal split, there is no obvious reason why subjects should behave differently for values of μ other than $1/3$ or $2/3$ (so long as the relevant theoretical proposition applies).

bid. This variant on the strategy method enabled twice as many quality observations to be collected, without changing the strategic nature of the game.¹¹

The potential seller submitting the lower bid in the first phase won the auction, with ties broken at random. The winning and losing bids and the contract price range were then revealed within the group. The lower bound contract price was one-quarter of the winning bid and the upper bound contract price was equal to the winning bid. The auction winner then became the seller for the interaction. The potential seller that submitted the losing bid waited for the interaction to conclude, earning a profit of zero.

In the second phase, the buyer observed the seller's chosen quality level and proposed a price to pay the seller from the contract price range. Since there was no bid available at which the lower bound contract price could exceed the buyer's valuation for low quality, the buyer's acceptance decision was automated in the experiment. In *Voluntary*, the buyer's chosen price was the final price by assumption and the interaction ended at the conclusion of this phase.

In *Arbitrator*, there was a third (non-decision) phase in which an appeal to the arbitrator could be triggered. The arbitrator's price was the amount from between the buyer's proposed price and the winning bid that provided the seller with closest to the arbitrator's preferred surplus division, defined by μ . To provide a clean test of the theory, the seller's appeal decision was automated. If and only if the buyer's proposed price was strictly below the arbitrator's price, an appeal was triggered. In other words, optimal use of the arbitrator was assumed. Following an appeal, the arbitrator was not always available to intervene. Specifically, the arbitrator could only set a final price on one-half of appeals, determined at random. If no appeal was triggered, or if the arbitrator was unavailable, then the outcome remained the same as in *Voluntary*. If an appeal was triggered and the arbitrator available, then the buyer had to pay the cost of arbitration. This cost was not factored into the arbitrator's surplus division calculations.

To illustrate, suppose the winning bid is 128 and the seller chooses high quality. The contract price range is from thirty-two to 128. The buyer proposes a price of sixty-five. In *Voluntary*, the final price is sixty-five. The buyer's profit is thirty-five and the seller's profit is twenty-five. In *Arbitrator*, if $\mu = 1/3$, the arbitrator's price is sixty. No appeal is triggered, and the profit outcomes are the same as in *Voluntary*. If $\mu = 2/3$, the arbitrator's price is eighty and so an appeal is triggered. If the arbitrator is unavailable, profit outcomes are the same as in *Voluntary*. If the arbitrator is available, a fee of two points is levied on the buyer. The buyer's profit is eighteen and the seller's profit is forty.

¹¹ Further benefits were that it helped to preserve anonymity of the winning seller and buyers did not have to wait for the completion of two stages before deciding. A complete strategy method in *Arbitrator* would have required sellers to specify a quality choice conditional on their bid and a decision over whether or not to invoke the arbitration clause, for every possible price. Such an approach was not practical.

4.2 Hypotheses

The equilibrium point predictions are presented in Table 3. In *Voluntary*, low quality is the unique implementable equilibrium outcome, as predicted by **Proposition 1**. In *Arbitrator*, high quality is the unique implementable equilibrium outcome, as predicted by **Proposition 2**. To within one increment, equilibrium bids are 120 in *Voluntary*, 80 in *Arbitrator* for $\mu = 1/3$, and 64 in *Arbitrator* for $\mu = 2/3$; buyers propose the lower bound contract price; buyers earn the full trade surplus and sellers earn zero in expectation.

Table 3 – Equilibrium point predictions.

Measure	Treatment	<i>Arbitrator</i>	
	Sub-treatment	$\mu = 1/3$	$\mu = 2/3$
b_i	<i>Voluntary</i>	120 or 121	80 or 81
q		Low quality	High quality
y		30	20
π_B		19 or 20	59 or 60
π_S		0 or 1	0 or 1

Notation follows from the model: A potential seller's bid is b_i ; the product quality level chosen by the auction winner is q ; the buyer's price proposal is y . The buyer's expected profit is π_B and the seller's expected profit is π_S .

The comparative hypotheses to be tested are summarised as follows:

Hypothesis 1. Bidding. (a) Winning bids are lower in *Arbitrator* than in *Voluntary*; and (b) winning bids are higher when $\mu = 1/3$ than when $\mu = 2/3$.

Hypothesis 2. Efficiency. The realised trade surplus net of the arbitration cost is greater in *Arbitrator* than in *Voluntary*.

Hypothesis 3. Reciprocity. In both *Voluntary* and *Arbitrator*, the buyer proposes the lower bound contract price.

Hypothesis 4. Fairness. Buyer profits are higher in *Arbitrator* than in *Voluntary*; and seller profits are unchanged between *Arbitrator* and *Voluntary*.

4.3 Procedure

The total number of subjects recruited for the experiment was 108. Subjects were recruited using the web-based software hroot (Bock et al. 2014) and allocated at random to one

of the two treatments, *Voluntary* or *Arbitrator*. Each treatment included six independent cohorts and no subject participated in more than one treatment. Every cohort had nine human subjects, three buyers and six sellers, who were matched into groups of one buyer and two sellers using a stranger matching protocol for thirty trade interactions. An algorithm ensured no subject played with the same two players in consecutive interactions, and subjects were informed of this. In procurement, firms typically compete for contracts within a fixed pool of potential sellers. While buyers and potential sellers vary between interactions, they often meet again at some future date. Thus, our matching protocol is preferable to perfect stranger matching. There were two cohorts (18 subjects) of the same treatment in every session.¹² To minimize the possibility of tacit collusion, we did not inform subjects about the cohort size.

In *Arbitrator*, the sequence of trade interactions was broken into two blocks of fifteen interactions, which constituted the two sub-treatments. To control for order effects, the block sequence in this treatment followed a crossover design, with one half of the subjects in a session assigned to the sequence $\mu = \{1/3, 2/3\}$ and one half of the subjects in a session assigned to the sequence $\mu = \{2/3, 1/3\}$. Where relevant, we refer to these two sequences in the analysis below as *Arbitrator-LH* and *Arbitrator-HL*, respectively.

To facilitate subject understanding, the task was presented using a cover story. A shipping company sought to procure an engine for its cargo ship. The choice of a shipping company was deliberate. Subjects are less likely to have experience from daily life with a shipping company, rather than say, an airline (see Alekseev et al. 201, on the merits and pitfalls of using a meaningful frame in economic experiments). One-third of subjects assumed the role of the shipping company (buyer). Two-thirds assumed the role of a supplier (seller). At the beginning of every interaction, each buyer was matched with two sellers. The potential seller that won the auction delivered the engine to the buyer. Whether the engine was of low or high quality depended on the winning seller's quality choice. The losing seller earned zero profit for the interaction.

The probabilistic nature of the arbitrator was explained in the experiment using a die roll. Subjects were informed that if an appeal to the arbitrator were to be triggered, the computer would roll a fair six-sided die. If the die came up one to three, the arbitrator would be unavailable, and the seller would receive the buyer's proposed price. If the die came up four to six, the arbitrator would be available, and the seller would receive the arbitrator's price instead.¹³ A fee of two points would then be levied on the buyer. The arbitrator's preferred surplus division was explained to subjects in terms of a two-to-one buyer (seller) to seller (buyer) profit ratio.

¹² The first cohort of the *Voluntary* treatment had to be re-run, with 9 (new) subjects in the session, after discovery of a comprehension issue. Data for the excluded cohort is available on request.

¹³ This randomization was implemented successfully: the arbitrator was available on 49.3 percent of appeals in the experiment, which is not significantly different from 50 percent (p -value = 0.690, two-tailed binomial test).

In *Arbitrator*, sellers were informed about the arbitrator's price associated with each combination of bid and quality level. Similarly, buyers were informed about the arbitrator's price associated with each proposal and whether an appeal would be triggered. At the end of each interaction, feedback was provided about the outcomes of a subject's own interaction group only, including on bids, quality, the price proposal, the arbitration process (in *Arbitrator*), the final price and profits. This feedback was provided to all groups simultaneously, to prevent subjects inferring information about the identities of others in their interaction group. Private feedback remained available throughout the session in a history table to facilitate learning, since the current design permits various permutations (many bids, two quality levels, many price proposals, three arbitration possibilities). The losing seller only observed the auction outcome. Neither buyer nor winning seller observed the losing seller's quality choice.

The experiments were computerised and programmed in oTree (Chen et al. 2016). To aid replicability and ensure common knowledge, video instructions were created and played at the start of an experimental session.¹⁴ Subjects had to answer a set of comprehension questions correctly before proceeding to two trial rounds, in which they were guided through the decision screens specific to their role. Roles were assigned randomly before the trial rounds and remained fixed. Each seller had their own auction in the trial rounds and these rounds were non-incentivised. The main experiment followed immediately. At the end of a session, subjects completed a survey to elicit demographic information, risk attitudes and trust attitudes.¹⁵

Subjects received money for their participation, which was paid in private and in cash at the end of a session. This was the only incentive offered for participation. In this type of experiment, there is potential for losses. To mitigate against this, subjects assigned to a buyer role began each interaction with a non-refundable endowment of five points. Subjects assigned to a seller role began each interaction with a non-refundable endowment of ten points.¹⁶ Any profit or loss was added to or subtracted from the endowment.

Sessions lasted sixty to seventy-five minutes. The total points accumulated from all rounds were multiplied by a pre-determined exchange rate of forty points per one British pound. Subjects were informed that the minimum amount they could leave the experiment with is £4 (160 points).¹⁷ A breakdown of the subject earnings by role and treatment is provided in Table 4. Those subjects assigned to a buyer role earned sizeably more than those assigned to the role of a

¹⁴ A copy of the instructions and links to the video recordings can be found in Appendix E.

¹⁵ Since the predictions do not depend on risk preferences, we do not employ an incentivized elicitation. Summary statistics from the post-experiment survey can be found in Appendix F.

¹⁶ The relative endowment sizes compensate for the expectation that, on average, each seller participates in half as many transactions as the buyer.

¹⁷ One subject encountered a limited liability problem. The results are qualitatively unchanged by inclusion or exclusion of the cohort in which this subject participated. For completeness, we include this cohort.

seller. This large payoff inequality is not important from a theoretical perspective, since the model does not assume any form of social preferences. Indeed, it is a direct prediction of the model due to the market imbalance. Empirically, however, social preferences may influence behaviour. We will consider this possibility explicitly when analysing the experiment results.

Table 4 – Mean subject earnings by role and treatment.

	<i>Voluntary</i>	<i>Arbitrator</i>
Buyer	£28.46	£27.66
Seller	£9.54	£9.92

5. Experiment Results

In this section, I present the findings of the lab experiment. First, I conduct an aggregate analysis. The cohort is the independent observation and is used for all statistical comparison tests in this sub-section. Second, I analyse the individual bidding strategies pursued by sellers, in relation to the equilibrium predictions. Third, I estimate a mixture model to demonstrate how buyer behaviour in the experiment might be rationalized by reference-dependent preferences.

5.1 Aggregate findings

The main results of the experiment are summarised in Table 5. Median cohort averages are provided for winning bids, the frequency of high product quality, the buyer's proposed price as a share of trade surplus, the frequency of dispute, final trading prices and profits. An overall trade efficiency measure is provided, which captures the realised percentage of attainable trade surplus net of the arbitration cost incurred. The summary statistics are presented for both the pooled *Arbitrator* data and sub-treatments.

Auction outcomes. The theory predicts that winning bids are lower when the contract contains a contingency and that the equilibrium bid is a function of the arbitrator's preferences. The data strongly supports the first prediction, but not the second.

Table 5 – Summary of experiment outcomes.

Measure	Treatment	<i>Voluntary</i>		<i>Arbitrator</i>	
	Sub-treatment		$\mu = 1/3$	$\mu = 2/3$	Pooled
(i) <i>Winning bid</i>		105.64	65.29	72.96	69.24
(ii) <i>Product quality</i>		0.45	0.43	0.53	0.48
(iii) <i>Proposal / surplus</i>		0.13	-0.18	-0.05	-0.15
	Low quality	0.13	-0.23	-0.13	-0.18
	High quality	0.13	-0.11	-0.05	-0.10
(iv) <i>Rel. freq. dispute</i>			0.80	0.81	0.83
	Low quality		0.74	0.77	0.79
	High quality		0.82	0.84	0.83
(v) <i>Final price</i>		38.93	37.95	42.25	40.78
	Low quality	32.68	30.78	35.24	33.29
	High quality	47.59	46.51	51.11	49.91
(vi) <i>Buyer profit</i>		33.85	33.8	32.32	31.55
	Low quality	17.32	18.35	13.93	15.96
	High quality	52.41	52.83	48.21	49.56
(vii) <i>Seller profit</i>		4.79	3.35	8.01	6.42
	Low quality	2.68	0.78	5.24	3.29
	High quality	7.59	6.51	11.11	9.91
(viii) <i>Efficiency</i>		63%	61%	67%	64%

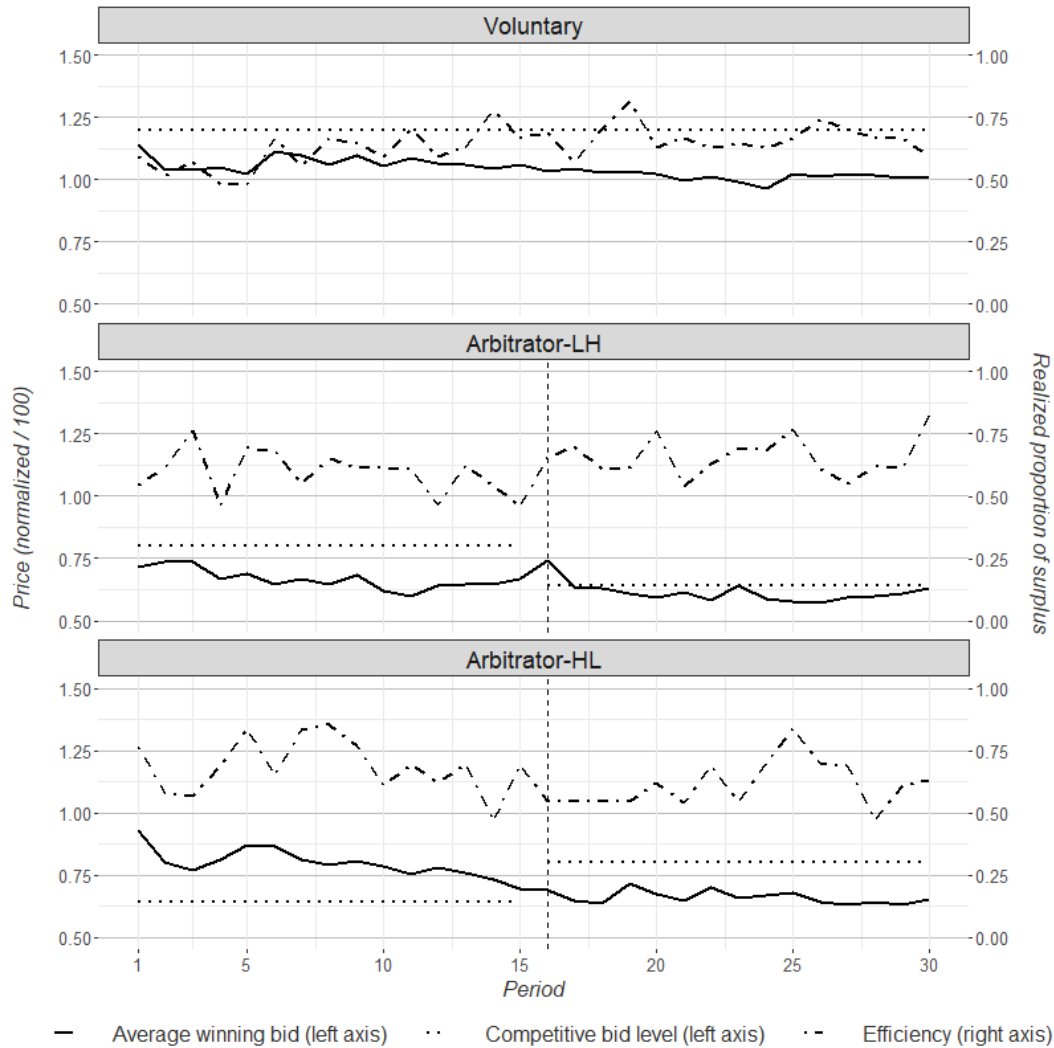
Notes: The table summarises the main outcomes of the lab experiment. All values are median cohort averages, based on six independent cohorts per treatment. (i) *Winning bid* is the lowest bid submitted by potential sellers in the auction phase; (ii) *Product quality* measures the relative frequency that high quality is chosen for delivery by the winning seller; (iii) *Proposal / surplus* measures the buyer's price proposal as a share of trade surplus; (iv) *Rel. freq. dispute* measures how often an appeal to the arbitrator is triggered (in *Arbitrator* only); (v) *Final price* is the trading price exchanged, irrespective of whether this price is determined by the buyer or an arbitrator; (vi) *Buyer profit* is the trade profit per round for the buyer; (vii) *Seller profit* is the trade profit per round for the winning seller; and (viii) *Efficiency* measures the realised percentage of attainable trade surplus net of any arbitration cost incurred.

Both bids and winning bids (the lower-order bidding statistic) differ between the two treatments. In *Voluntary*, the average bid submitted in the auction is 115.77 and in the pooled *Arbitrator* data, 78.55. The average winning bids are 105.64 and 69.24, respectively. Both differences are statistically significant in the direction predicted by Hypothesis 1 (p -value < 0.01 for both comparisons, Wilcoxon-Mann-Whitney tests, one-tailed).¹⁸ Pairwise comparisons

¹⁸ One-tailed statistical tests are used for all comparisons in which the theory predicts a direction.

between the *Voluntary* treatment and the *Arbitrator* sub-treatments yield similar statistical differences.¹⁹ There is a pronounced downward trend in winning bids over time in *Arbitrator*, which is not apparent in *Voluntary*. This trend difference in bids is discernible from Figure 3, which plots the evolution of winning bids and trade efficiency over time.

Figure 3. Winning bids and trade efficiency over time.



Notes: The vertical dashed line in the *Arbitrator* treatment panels represents the change in the arbitrator's preference parameter μ after period fifteen.

The cumulative bid distribution functions are presented in Appendix B. The distribution of bids in *Voluntary* is stochastically larger than in *Arbitrator* (p -value < 0.001 , Kolmogorov-Smirnov test, one-tailed). In *Voluntary*, bids are not significantly different from the equilibrium point prediction of 120 (p -value = 0.563, one-sample Wilcoxon Signed-Rank test, two-tailed).

¹⁹ *Voluntary* vs. $\mu = 1/3$, p -value < 0.01 for both; *Voluntary* vs. $\mu = 2/3$, p -value = 0.021 and p -value < 0.01 .

Result 1(a). We find support for Hypothesis 1(a). A contingent contract induces more aggressive bidding and lower auction prices than a non-contingent contract.

The theory is less successful at explaining bidding behaviour within-subjects in *Arbitrator*. When the arbitrator favours the buyer ($\mu = 1/3$), rather than the seller ($\mu = 2/3$), we expect sellers to adjust their bids upwards to compensate for the lower expected price in the event of winning the auction and the arbitrator being available. This is not observed in the data. Average bids are 74.71 when $\mu = 1/3$ and 81.06 when $\mu = 2/3$. Winning bids are 65.29 and 72.96, respectively. Neither difference is significant (p -value = 0.563 and p -value = 0.313, Wilcoxon Signed-Rank tests, two-tailed). The failure to adjust bids as expected in response to a change in μ is not concealed by learning. If we restrict attention to experienced sellers only – classified as periods 11 to 15 and 26 to 30 – neither the bids nor winning bids are significantly different (p -value = 1.00 and p -value = 0.563, Wilcoxon Signed-Rank tests, two-tailed). This is also apparent from the bottom two panels of Figure 3. The vertical dashed line represents the change in the arbitrator's preference parameter μ after period fifteen, given our crossover design. Regardless of the sequence in which the arbitrator's preferences are presented, winning bids begin at around 80 and converge to around 64 (recall that these are in fact the equilibrium predictions when $\mu = 1/3$ and $\mu = 2/3$, respectively).

The finding is not a consequence of behavioural spillover effects. Spillover effects are possible in *Arbitrator* because μ is varied within subjects between the first and second half of the experiment (see Bednar et al. [2012] for a discussion of the importance of accounting for spillovers in economic experiments). The results in this section are qualitatively unchanged if we only consider data from periods one to fifteen and so only consider between-subjects variation.²⁰ This suggests that any spillover effects, if present, did not significantly alter subject behaviour.

Result 1(b). We reject Hypothesis 1(b). Potential sellers fail to adjust their bids based on expected values in response to a change in the contractual contingency.

Product quality. The theory predicts that high quality trading relationships are more likely when the contract contains a contingency. Unlike for bids, the prediction is independent of the arbitrator's preferences. There is at most weak evidence for this in the data.

High quality trades are observed across the winning bid distributions in both treatments. The relative frequencies of high product quality delivered by the winning seller in the *Voluntary* and *Arbitrator* treatments are forty-five and forty-eight percent, respectively. This difference is not significant (p -value = 0.350, Wilcoxon-Mann-Whitney test, one-tailed). There is no significant

²⁰ Summary statistics for the spillover analyses can be found in Appendix B.

difference in quality between the *Arbitrator* sub-treatments (p -value = 0.313, Wilcoxon Signed-Rank test, two-tailed).

Figure 3 reveals that trade efficiency fluctuates in a similar band over time in all three contractual conditions. Average trade efficiency is sixty-three percent in *Voluntary* and sixty-four percent in *Arbitrator*. The difference is not statistically significant (p -value = 0.531, Wilcoxon-Mann-Whitney test, one-tailed). There is no significant pairwise difference between *Voluntary* and either of the *Arbitrator* sub-treatments (*Voluntary* vs. $\mu = 1/3$, p -value = 0.803 and vs. $\mu = 2/3$, p -value = 0.294, Wilcoxon-Mann-Whitney tests, one-tailed).

Result 2. We reject Hypothesis 2. A contingent contract does not significantly increase trade efficiency relative to a non-contingent contract.

Price proposals. The theory developed earlier assumes standard preferences and so predicts that buyers always propose the minimum price in the contract price range. Given the vast experimental economics literature documenting social preferences, it is unsurprising that this assumption is falsified behaviourally. Buyer reciprocity, however, is significantly undermined when the contract contains a contingency.

In *Voluntary*, the buyer on average proposed a price equal to 37.6 percent of the winning bid. In *Arbitrator*, the average proposal is forty-two percent, but of a lower average price. Both are significantly above the minimum required one-quarter (p -value = 0.016 for both comparisons, one-sample Wilcoxon Signed-Rank tests, one-tailed) and attest to the existence of social preferences in the buyer population.²¹ There is a sustained gap between lower bound contract price and the average buyer proposal in the experiment over time (the time-series is contained in Appendix B). The average final trading price in *Arbitrator* is 40.78, which coincides nearly exactly with the equilibrium prediction of seller compensation equal to the delivery cost of high quality. By contrast, in *Voluntary* the final trading price is 38.93, significantly above the cost of low quality (p -value = 0.031, one-sample Wilcoxon Signed-Rank test, two-tailed).

The level of buyer reciprocity diverges between the treatments. Aggregate price proposals in *Voluntary* imply a seller profit equal to 13.26 percent of the transaction surplus. The picture is very different in *Arbitrator*, where proposals imply a seller loss equal to 14.65 percent of the transaction surplus. The difference is highly significant (p -value < 0.01, Wilcoxon-Mann-Whitney test, two-tailed). The arbitration clause was triggered on four out of five transactions.

Revealingly, while in *Voluntary* buyers propose that the seller receives a similar share of the surplus independently of the quality level, in *Arbitrator* buyers offer significantly less in exchange for low product quality than for high quality (p -value = 0.031, Wilcoxon Signed-Rank

²¹ Since proposals cannot be below one-quarter of the winning bid, we use a one-tailed test.

test, two-tailed). This implies that buyers are less forgiving of a seller's reluctance to deliver high quality with a contingent contract.

Result 3. We reject Hypothesis 3. A contingent contract crowds out buyer reciprocity and the arbitrator acts as a partial substitute for reciprocity in the determination of trading prices.

Buyer and seller profits. The standard theory predicts that sellers compete away trading profits and the buyer appropriates the trade surplus, ruling out a fairness function of arbitration. To some extent, this prediction is borne out in the data.

In both the *Voluntary* and *Arbitrator* treatments, buyers earn significantly higher profits than sellers (p -value = 0.016 for both comparisons, Wilcoxon Signed-Rank tests, one-tailed). The gains from high quality trading relationships are also apparent. Buyers and sellers each earn roughly three times as much from high quality trades as low quality trades, in line with the surplus multiplier. Yet buyers do not benefit from the presence of an arbitrator. In fact, buyer profits are remarkably stable across experimental conditions, at 33.85 in *Voluntary*, 33.80 in *Arbitrator* when $\mu = 1/3$ and 32.32 when $\mu = 2/3$. Neither pairwise difference is significant.²²

The average seller profit is 4.79 in *Voluntary* and 6.42 in *Arbitrator*. This increase is not significant (p -value = 0.818, Wilcoxon-Mann-Whitney test, two-tailed). It does, however, mask underlying differences between the *Arbitrator* sub-treatments. Whereas seller profits are 3.35 when $\mu = 1/3$, they are 8.01 when $\mu = 2/3$, although the latter is not significantly different from the *Voluntary* treatment level (p -value = 0.310, Wilcoxon-Mann-Whitney test, two-tailed). We find evidence that this may be due to a lack of statistical power – see the robustness check below. More data is required to make concrete inferences on whether, and the conditions under which, an arbitration mechanism serves a fairness function in competitive procurement interactions.

Result 4. We find partial support for Hypothesis 4. Buyers earn significantly more than sellers but do not benefit from a contingent contract; there is weak evidence that sellers benefit from a contingent contract when the arbitrator's preferences are in their favour.

Robustness check: Since the number of independent observations per treatment is small, low statistical power might be a concern for any null result in the hypothesised direction. We have disaggregated data from 540 matching groups in each treatment, which yields greater statistical power to detect effects than the more conservative mean comparison tests used above. Results 1 to 4 are qualitatively unchanged if we conduct a regression analysis on the matching group level data, accounting for intra-cohort dependencies and small-sample considerations

²² *Voluntary* vs. $\mu = 1/3$, p -value = 0.758 and vs. $\mu = 2/3$, p -value = 0.758 (Wilcoxon-Mann-Whitney tests, one-tailed); $\mu = 1/3$ vs. $\mu = 2/3$, p -value = 1.00 (Wilcoxon Signed-Rank test, two-tailed).

using the wild cluster bootstrap method (Cameron et al. 2008). The results of this exercise are contained in Appendix C.

5.2 Seller bidding strategies

In this sub-section, I analyse seller bidding strategies formulated during the auction phase of the experiment. There is data from 1,070 strategies submitted in *Voluntary* and 1,059 strategies submitted in *Arbitrator*.²³ High quality is a component of forty-two percent of bidding strategies in *Voluntary*. In *Arbitrator*, this is true of forty-six percent of seller strategies when $\mu = 1/3$, and fifty-five percent of seller strategies when $\mu = 2/3$. The latter is a significant increase over *Voluntary* (p -value = 0.066, Wilcoxon-Mann-Whitney test based on cohort averages, one-tailed). Thus, whereas a contingent contract is not able to significantly improve the winning product quality delivered, it does promote higher quality among the pool of potential sellers, particularly when the arbitrator is known to favour the seller.

As a first step, it is instructive to examine average bids that accompany the respective quality levels at the cohort level. In *Voluntary*, the median bid for low quality strategies is 118.35, very close to the equilibrium point prediction. The median bid for high quality strategies is significantly lower at 109.71 (p -value = 0.031, Wilcoxon Signed-Rank test, two-tailed). This suggests that, with a non-contingent contract, deviation to a lower bid may have been driven by a belief that buyers view high quality as a “gift” that they are more likely to reciprocate.

In *Arbitrator*, by contrast, the median bid for high quality strategies – at 81.86 – is significantly higher than for low quality strategies – at 74.46 (p -value = 0.063, Wilcoxon Signed-Rank test, two-tailed). Similar qualitative differences are observed in both the *Arbitrator* sub-treatments (see summary statistics in Appendix B). This suggests that potential sellers anticipated the lower levels of buyer reciprocity with a contingent contract and adjusted their bid–quality combinations to reflect the implied difference in delivery cost. Conditional on bids, potential sellers chose the incentive compatible quality level in fifty-six percent of auction decisions in *Arbitrator*.²⁴ This measure of incentive compatibility is based on the expectation of strictly self-interested buyers. If sellers expect some degree of buyer reciprocity, the delivery of high quality might be justified. Nevertheless, in *Arbitrator* buyer reciprocity is attenuated and so the assumption of self-interested buyers is more appropriate.

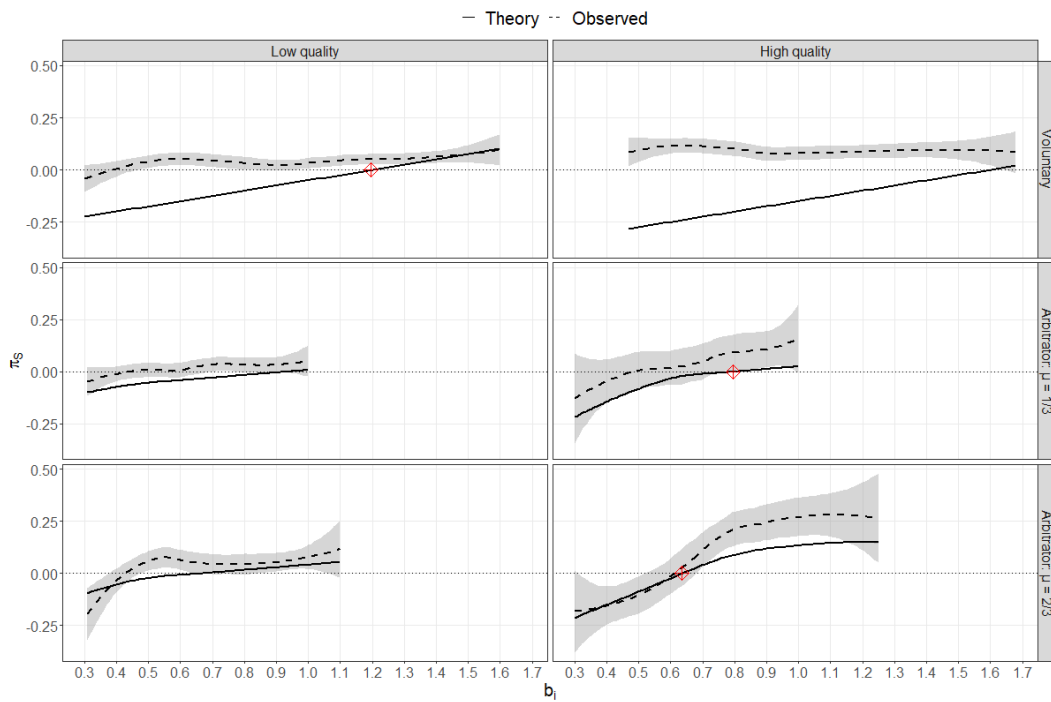
To check the implications of sub-optimal seller bidding strategies for the efficiency Hypothesis 2, in Figure 4 I compare *actual* versus *predicted* seller profits per quality level, as a

²³ The difference is due to the use of a hard time-out protocol in the experimental sessions. In the event of timing out, the default bid was 200, the quality level was decided at random and the price proposal was set equal to the winning bid. The rate of data loss is less than one percent, which seems a reasonable trade-off given the benefits of the protocol for minimizing delays during a session.

²⁴ This calculation assumes risk neutrality. In *Voluntary*, incentive-compatible choice is always low quality.

function of the winning bid. The SPNE point in each experimental condition is marked with a red diamond plus. Consistent with our earlier results, buyer reciprocity is stable across bids in *Voluntary*. On average, observed seller profits remain flat above the break-even level, independent of the bid and quality choice. In *Arbitrator*, seller profits track the prediction as a function of the winning bid, at a level weakly above the theory. For the median winning bid in each *Arbitrator* sub-treatment, high quality is incentive compatible (see figure in Appendix B).

Figure 4. Seller profits as a function of the winning bid: Actual versus predicted.



Notes: The lines are loess smoothers and the shaded regions for the observed data are 95% confidence intervals. The red diamond plus is the equilibrium bid level.

As a further investigation into the determinants of bidding strategies, I employ two complementary parametric methods (Table 6). First, to examine the factors affecting seller bids *independently* of quality, I use random effects regressions with three levels of dependencies and controlling for a range of covariates (columns 1 and 2, see also Moffatt 2015, Chapter 4.7). The dependent variable is the bid of subject i in cohort j in period t . To account for intra-session correlation, the variance is estimated at subject and cohort levels.²⁵ We observe a positive

²⁵ The exact regression specification is $y_{ijt} = \beta_0 + \beta_1 x_{ijt} + \beta_2 z_i + \beta_3 o_j + \beta_4 t + u_i + v_j + \varepsilon_{ijt}$, with variances $V(u_i) = \sigma_u^2$, $V(v_j) = \sigma_v^2$ and $V(\varepsilon_{ijt}) = \sigma_\varepsilon^2$. The three levels are $i = 1, \dots, 36$ (sellers), $j = 1, \dots, 6$ (cohorts) and $t = 1, \dots, 30$ (periods). Independent variables are specific to the subject-cohort-period (x_{ijt}), subject only (z_i) or cohort only (o_j).

dependency of current bids on the once-lagged competitor's bid, which captures the strategic complementarities of Bertrand competition. Bids trend downwards over time in both treatments.

Table 6 – Determinants of seller bidding strategies.

Dependent variable Estimation method Treatment	Bid <i>Multilevel</i>		Prob. High Quality <i>IV</i>	
	<i>Voluntary</i>	<i>Arbitrator</i>	<i>Voluntary</i>	<i>Arbitrator</i>
<i>Competitor's bid</i> $t-1$	0.222*** (0.024)	0.075*** (0.024)		
<i>Bid</i>			-0.016** (0.006)	0.002 (0.007)
<i>Period</i>	-0.302*** (0.063)	-0.510*** (0.077)	0.007 (0.009)	0.002 (0.010)
$\mu = 2/3$		1.44 (1.28)		0.149 (0.159)
<i>OrderHL</i>		8.37 (9.38)		0.178 (0.490)
<i>Constant</i>	87.62*** (13.58)	85.38*** (19.95)	1.415* (0.859)	-2.71 (1.79)
Control variables	Yes	Yes	Yes	Yes
Cohort fixed effects	No	No	Yes	Yes
Observations	1,037	990	1,028	970
Subjects	36	36	36	36
Cohorts	6	6	6	6
Wald χ^2	149.8	85.08	51.53	45.57
χ^2 p -value	0.000	0.000	0.000	0.000
σ_v	15.01	11.08		
σ_u	9.60	5.44		
σ_ε	16.68	20.05		

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The models in columns 1 and 2 are estimated using multilevel mixed effects linear regression and include two random effects intercepts that capture intra-session correlation at the subject and cohort levels, respectively. Coefficient estimates are presented, with standard errors in parentheses. The models in columns 3 and 4 are estimated using instrumental variable (IV) probit regression. Coefficient estimates are on the z-score scale, with robust standard errors clustered at the subject level. Cohort fixed effects are included. *Bid* is the continuous endogenous regressor and the excluded instrument is the once-lagged bid. The IV first stage is reported in Appendix B. The following control variables (not shown) are included: dummy for being female; dummy for being an economics and finance major; two Likert scales for self-reported willingness to take risks in general and in financial matters; dummy for reporting trust in strangers; and a generalised trust index. *OrderHL* is a dummy variable to indicate that the cohort followed the sequence $\mu = \{2/3, 1/3\}$ in the *Arbitrator* treatment. Period 1 is excluded from the models in column 1 and 3; periods 1 and 16 are excluded from the models in columns 2 and 4 due to the within-subjects design.

Second, to examine the factors affecting seller quality choices, I estimate instrumental variable probit regressions in which quality is the dependent variable and bid is the endogenous regressor, which is instrumented for by the once-lagged bid (columns 3 and 4). The instrumental variable estimator is used to mitigate the issue of simultaneity between bids and quality choices inherent in the strategy method application. The instrument passes recognised strength tests – see Appendix B for diagnostic statistics. Standard errors are adjusted using the Huber-White

sandwich estimator of variance at the subject level. Cohort fixed effects are included in these regressions, based on a test of their joint significance (for both treatments, p -value < 0.001, two-tailed Wald test). We find support for our earlier observation that lower bids are more likely to be accompanied by high quality in *Voluntary*, but that this relationship breaks down with the introduction of an arbitrator. Reassuringly, the order in which the values of μ are presented to sellers in *Arbitrator* has no significant effect on either bid or quality decisions.²⁶

5.3 Reference-dependent buyers

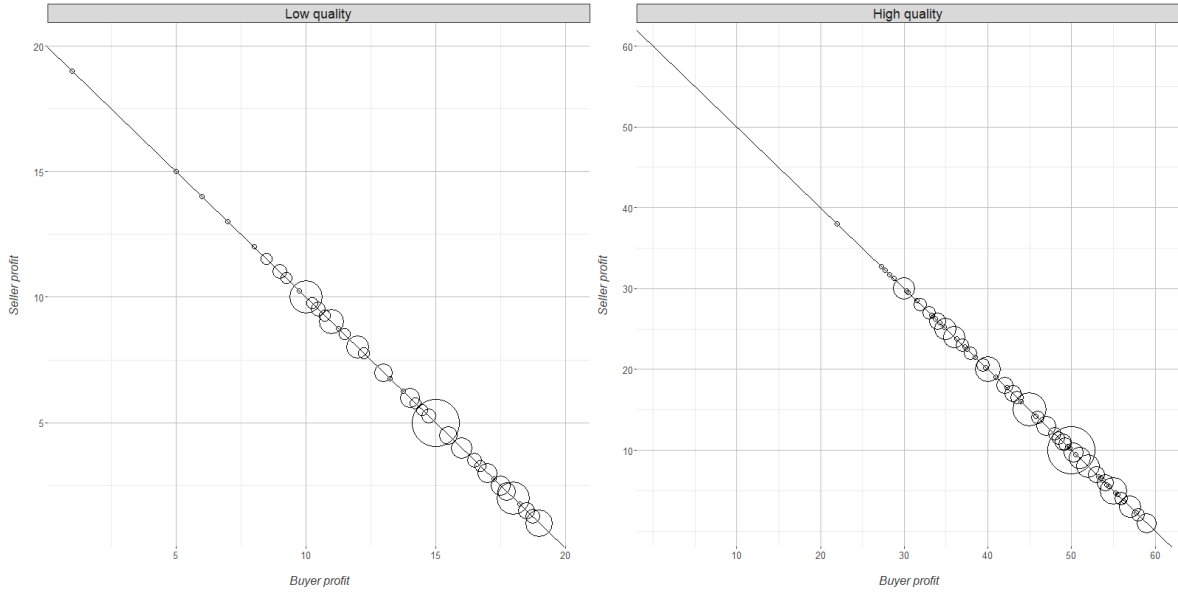
In this sub-section, I analyse buyer behaviour during the price proposal phase of the experiment. There are eighteen buyers in each treatment, each of whom is observed over thirty trading periods. We have data from 535 price proposal decisions in each treatment.

In *Voluntary*, only 23.7 percent of buyer choices are of the lower bound contract price. This frequency varies by product quality, rising to 31.4 percent of choices in exchange for low quality and falling to less than fifteen percent in exchange for high quality. This implies that potential sellers' beliefs about the buyer engaging in gift-exchange type behaviour are justified. In Figure 5, we plot the range of surplus allocations in *Voluntary*, for all transactions in which the seller earns non-negative profits (423 out of 535). A further three transactions are excluded in which the buyer incurred a loss. These losses occurred early on and are likely attributed to learning. The sizes of the open circles in the figure are weighted by the number of observations at the allocation. The left panel corresponds to low quality trades and the right to high quality trades. In both panels, there are concentrations at the salient fifty-fifty surplus split. The most common allocations are found at three-to-one and five-to-one buyer to seller ratios.

In *Arbitrator*, proposals of the lower bound contract price are observed more frequently: 46.8 percent of choices when $\mu = 1/3$ and 50.8 percent of choices when $\mu = 2/3$. These frequencies are stable across quality levels. Thus, a seller in this treatment could expect to encounter a strictly self-interested buyer on roughly one-half of transactions, regardless of the quality level delivered. Around one in five buyer choices in *Arbitrator* are "mimicking". That is, the buyer proposes the arbitrator's interim preferred price p_{it}^A , given the winning bid and quality level delivered. Note that the arbitrator's interim preferred price does not necessarily coincide with the arbitrator's final price, because the arbitrator is constrained to award at least the buyer's proposal - see equation (3) in the model. The frequency of mimicking choices was again stable across quality levels. Most of the remaining price proposals are in the region between the lower bound contract price and the arbitrator's preferred price. Less than four percent of price proposals awarded the seller more than what the arbitrator deemed a fair surplus split.

²⁶ No significant order effects are found for buyer behaviour either (see Appendix B).

Figure 5. Buyer surplus allocations in the *Voluntary* treatment.



Notes: Based on individual buyer decisions over the course of the experiment. The sizes of the open circles are weighted by the number of surplus allocations at the coordinate.

In both treatments, buyers learn to follow the equilibrium strategy as the experiment progresses. Non-parametric bootstrap tests of a positive time trend in the number of lower bound price proposals show strong significance across conditions.²⁷ The standard theory cannot, however, explain the persistence of mimicking choices in *Arbitrator*. When $\mu = 1/3$, no time trend is observed in the average number of such choices (p -value = 0.499). When $\mu = 2/3$, there is a significant *positive* time trend (p -value = 0.015).

One explanation for the persistence of mimicking choices is that some buyers have reference-dependent social preferences over outcomes. In the experiment, subjects are informed in advance about the arbitrator's preferred surplus allocation and when and how this would change. Buyers with a concern for fairness may have anchored their proposals on the *arbitrator's interim preferred price* $p_{it}^{A'}$ and perceived this as the fair reference point in each trade. If the reference point is determined by recently held expectations (cf. Kőszegi and Rabin 2006), then we expect the proportion of mimicking buyer choices to be stable across the *Arbitrator* sub-treatments. We refer to this as the *reference-dependent argument*. Plausibility of this argument is suggested by the raw choice frequencies: 22.3 percent of choices are mimicking when $\mu = 1/3$ and 20.7 percent when $\mu = 2/3$, respectively. Such an argument could also explain why price proposals, as a share of trade surplus, are higher when the arbitrator favoured the seller (see

²⁷ Based on 999 replications, *Voluntary*: Mean = 0.365, S.E. = 0.187, p -value = 0.027; *Arbitrator*, $\mu = 1/3$: Mean = 0.763, S.E. = 0.271, p -value < 0.001; *Arbitrator*, $\mu = 2/3$: Mean = 0.553, S.E. = 0.271, p -value = 0.026.

Table 5). Given our within-subjects design, the opposite would obtain if fair buyers evaluated payoffs relative to a fixed reference point, as in usual models of social preferences (e.g. Fehr and Schmidt 1999, Bolton and Ockenfels 2000).

To test the reference-dependent argument, I separate buyers in *Arbitrator* into two behavioural types, self-interested (*Self*) or fair (*Fair*). I then estimate the parameters underlying each type's price proposal function using a finite mixture model. This structural approach will help us to identify the determinants of buyer behaviour in the experiment. The mixture model estimation procedure is adapted from Moffatt (2015) and full details are contained in Appendix D. The observation that buyers in *Arbitrator* tend to either offer the lower bound contract price or mimic the arbitrator's interim preferred price $p_{it}^{A'}$ is used to define the type strategies. Self-interested buyers propose the lower bound contract price. For fair buyers, we specify a latent model in which price proposals anchor on $p_{it}^{A'}$ with a normally distributed error term of variance s^2 . We expect the coefficient estimate on $p_{it}^{A'}$ for this type to be close to one. An implicit assumption is that subjects do not switch between types.²⁸

A two-limit tobit model is appropriate, where the limits correspond to the lower and upper bounds of the contract price range and are specific to each buyer and period. The two buyer types are represented in the model by the mixing fractions fr_{self} and fr_{fair} . The maximum likelihood parameter estimates obtained from this procedure are summarized in Table 7. Bootstrapped standard errors are calculated, with resampling clusters at the cohort level. The fraction of self-interested buyers in *Arbitrator* is estimated to be 0.611 and the fraction of fair buyers to be 0.389 (p -value < 0.001). The proportion of fair types in the seminal study of Berg et al. (1995) has been estimated at around forty percent (see Bolton and Ockenfels 2000). A similar estimate, using a least absolute deviation method, was obtained in Katok et al. 2020. Consistent with the reference-dependent argument, the coefficient estimate on $p_{it}^{A'}$ at 0.971 is not significantly different from one (p -value = 0.853, two-sided Wald test).

Panel (a) of Figure 6 displays the posterior type probabilities obtained from estimation of the mixture model, as a function of the number of lower bound proposals. The model fully separates buyers into the two types. A large cluster of buyers in *Arbitrator* never choose the lower bound price and these subjects fall naturally into the fair type category. Self-interested types are identified as choosing the equilibrium (minimal) price on more than ten transactions. Panel (b) of the figure presents the histogram of price proposals by type, as a share of the trade surplus. For fair types, the largest densities can be found at proposals equal to 1/3 and 2/3 of the surplus.

²⁸ Preliminary evidence to support these type definitions is obtained from raw strategy count data, which are presented in Appendix B. Buyers appear to gravitate towards one strategy during the experiment, rather than mixing between different strategies.

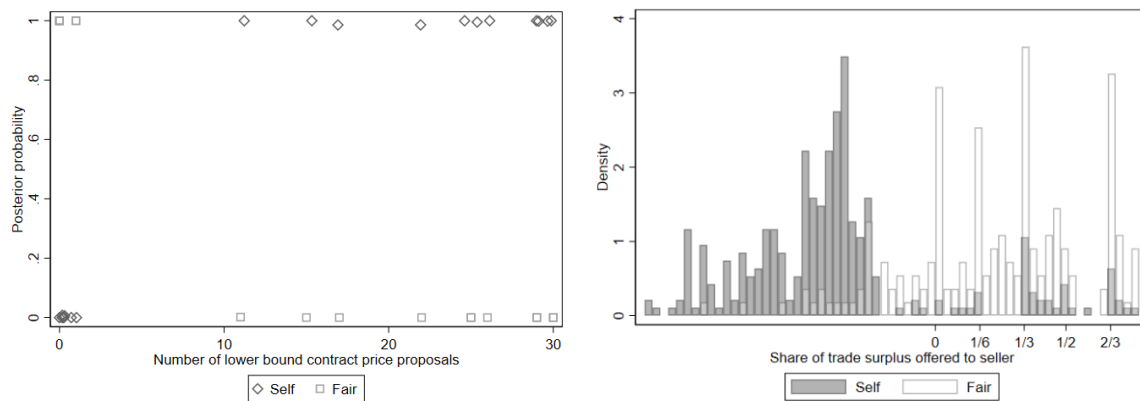
We exclude any proposal equal to the winning bid (35 observations) because these do not necessarily reflect a buyer's true preference. If we include these observations, the mass at 1/3 of the surplus increases. Weaker norms are also apparent at 1/6 of the surplus and the breakeven level. By comparison, there is only a small mass at the 50:50 split.

Table 7 – Maximum likelihood estimates from the finite mixture two-limit Tobit model of buyer behaviour in the *Arbitrator* treatment.

Fair buyers	
$p_{it}^{A'}$	0.971 (0.158)***
Constant	-3.64 (5.76)
s	8.51 (6.54)
Mixing fractions	
fr_{self}	0.611 (0.084)***
fr_{fair}	0.389 (0.084)***
Observations	535
Wald χ^2	37.97
χ^2 p -value	0.000
(Subjects, Periods)	(18, 30)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Bootstrapped standard errors in parentheses, clustered at the cohort level, based on 999 bootstrap replications.

Figure 6. Finite mixture model validation.



(a) Posterior type probabilities.

(b) Histogram of buyer type proposals.

Notes: Panel (a) contains a jittered scatter plot of posterior type probabilities from the finite mixture two-limit Tobit model. Panel (b) contains a histogram of buyer price proposals as a share of the trade surplus, excluding upper censored price proposals and one outlying observation in which a buyer offered more than the full trade surplus to the seller.

6. Conclusion

Arbitration is a common method of alternative dispute resolution in procurement interactions characterised by moral hazard. By embedding a contingency into the contract, trade parties can avoid the prohibitive costs of lengthy court proceedings and align incentives to maximise the gains from trade. I use a combination of theory and lab experiments to analyse the effect of a contractual contingency in such an environment. The seller is determined by a price-based procurement auction and the contract price is renegotiable. In theory, with a non-contingent contract, trade is inefficient. A contingent contract, meanwhile, can establish welfare-enhancing trust, with increased bidding competition and improved product quality.

A notable implication of the equilibrium analysis is that, in a broad sense, arbitration does not benefit the seller in our environment – expected seller profits are zero, with or without a contingent contract. Buyers, who hold the balance of market power, do benefit. Our model can rationalize why powerful corporations may favour the use of arbitration agreements, from simple regard to their own self-interest. Another interesting implication arises from our finding that uncertainty is a necessary condition for the effectiveness of a contingent contract. That is, if the auction selection rule is price-based, then making an arbitrator available with 100 percent probability is (i) always inefficient and (ii) welfare-decreasing if there exists an efficient contingent contract with a probabilistic arbitrator. Since the total arbitration cost increases in the relative frequency with which an arbitrator is employed, any industry scheme guaranteeing the availability of an arbitrator during trade may incur an unnecessary and counter-productive expense. Potential sellers can use such information to lower their bids and increase their probability of winning the contract. That sellers may lower their bids, in expectation of recovering costs through a dispute mechanism, was recognised in a recent survey of industry practitioners.²⁹

The experiment results demonstrate that the contingent contract is robust in the following sense: despite the potential sellers' failure to adjust their bids as expected in response to a change in the arbitrator's preferences, high quality remains incentive compatible on average. Despite this, we find that a contingent contract crowds out buyer reciprocity and cannot outperform a non-contingent contract from a trade efficiency standpoint. This is consistent with prior evidence from labour market experiments on the demotivating effects of explicit incentives (e.g. Fehr and Gächter 2000, Fehr et al. 2007), and with principal-agent literature on the hidden costs of incentive devices (e.g. Falk and Kosfeld 2006). There is also evidence of reference-

²⁹ 2019 International Arbitration Survey, p 7: "Some interviewees noted that they had observed project participants bidding a lower price upon the expectation of recovering sums through variation orders which, if disputed, could be arbitrated." Available at <https://www.pinsentmasons.com/thinking/special-reports/international-arbitration-survey>.

dependency in buyers' payment decisions (e.g. Kőszegi and Rabin 2006), which enables the contingent contract to weakly improve the position of sellers in the experiment.

Both the non-contingent and contingent contracts tested in the lab imply large efficiency gains over a rigid, fixed-price contract, which performs poorly in procurement auctions with moral hazard (Brosig-Koch and Heinrich 2014, Fugger et al. 2019). There are, of course, alternative mechanisms to mitigate the seller moral hazard problem, such as buyer-determined auctions and systems of reputation. Since there is a cost involved in organising an arbitration scheme, these mechanisms may be complementary and/or preferable in certain settings. We also employed an optimal dispute resolution mechanism in the experiment, as implemented by a computer program. Whether or not trade parties make optimal use of dispute resolution is an interesting question for future work. An alternative avenue would be to assess how payment norms emerge endogenously if a human arbitrator(s) were introduced into the experiment. Third-party punishment behaviours have been observed to serve a useful function in enforcing distributional norms (e.g. Fehr and Fischbacher 2004). Whether social norms emerge in a competitive procurement environment, and whether the incentives offered by third parties are enough to promote efficient trade relationships, is a question best answered empirically.

An important limitation of our setup is that the seller is endowed with no bargaining power during trade. The assumption that buyers are endowed with full bargaining power is often used in the contract-theoretic literature to create the most severe hold-up problem (e.g. Hart and Moore 1999, Hoppe and Schmitz 2011). Our model might be extended to allow for an investigation of a seller's relative bargaining power by applying the Nash bargaining solution at Stage 4. A further assumption is that the arbitrator's preferred settlement is a known and deterministic function of the transaction price and quality. In practice, this decision is likely to involve some noise, due for example to difficulty in establishing the facts of the case. This *caveat* should be kept in mind if drawing policy implications, which may overstate the effects of an arbitrator in real-world dispute resolution. To help mitigate against this, introducing a stochastic component into the arbitrator's preferences would be a useful extension in future work.

References

- Alekseev, A., Charness, G., & Gneezy, U. (2017). Experimental methods: When and why contextual instructions are important. *Journal of Economic Behavior & Organization*, 134, 48-59.
- Allison, J. R. (1990). Five ways to keep disputes out of court. *Harvard Business Review*, 68(1), 166-177.
- Andreoni, J., & Bernheim, B. D. (2009). Social Image and the Fifty-Fifty Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica*, 77(5), 1607-1636.
- Ashenfelter, O., Currie, J., Farber, H. S., & Spiegel, M. (1992). An Experimental Comparison of Dispute Rates in Alternative Arbitration Systems. *Econometrica*, 60(6), 1407-1433.
- Ashenfelter, O. C., & Bloom, D. E. (1984). Models of Arbitrator Behavior: Theory and Evidence. *American Economic Review*, 74(1), 111-124.
- Bajari, P., Houghton, S., & Tadelis, S. (2014). Bidding for incomplete contracts: An empirical analysis of adaptation costs. *American Economic Review*, 104(4), 1288-1319.
- Bartling, B., & Schmidt, K. M. (2015). Reference points, social norms, and fairness in contract renegotiations. *Journal of the European Economic Association*, 13(1), 98-129.
- Baye, M. R., & Morgan, J. (1999). A Folk Theorem for One-Shot Bertrand Games. *Economics Letters*, 65(1), 59-65.
- Bazerman, M. H., & Gillespie, J. J. (1999). Betting on the future: The virtues of contingent contracts. *Harvard Business Review*, 77, 155-162.
- Bednar, J., Chen, Y., Liu, T. X., & Page, S. (2012). Behavioral spillovers and cognitive load in multiple games: An experimental study. *Games and Economic Behavior*, 74(1), 12-31.
- Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, 71, 117-120.
- Bolton, G. E., & Katok, E. (1998). Reinterpreting Arbitration's Narcotic Effect: An Experimental Study of Learning in Repeated Bargaining. *Games and Economic Behavior*, 25(1), 1-33.
- Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1), 166-193.
- Brandts, J., Ellman, M., & Charness, G. (2015). Let's talk: How communication affects contract design. *Journal of the European Economic Association*, 14(4), 943-974.
- Brosig-Koch, J., & Heinrich, T. (2014). Reputation and mechanism choice in procurement auctions: An experiment. *Production and Operations Management*, 23(2), 210-220.
- Brown, M., Falk, A., & Fehr, E. (2004). Relational contracts and the nature of market interactions. *Econometrica*, 72(3), 747-780.

- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, 90(3), 414-427.
- Chang, W.-S., Salmon, T. C., & Saral, K. J. (2016). Procurement auctions with renegotiation and wealth constraints. *Economic Inquiry*, 54(3), 1684-1704.
- Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88-97.
- Cooter, R., Marks, S., & Mnookin, R. (1982). Bargaining in the shadow of the law: A testable model of strategic behavior. *The Journal of Legal Studies*, 11(2), 225-251.
- Crawford, V. P. (1981). Arbitration and Conflict Resolution in Labor-Management Bargaining. *American Economic Review*, 71(2), 205-210.
- Crawford, V. P. (1985). The role of arbitration and the theory of incentives. In A. E. Roth (Ed.), *Game-Theoretic Models of Bargaining*. Cambridge, MA: Cambridge Univ. Press.
- Davis, A. M., & Leider, S. (2018). Contracts and Capacity Investment in Supply Chains. *Manufacturing & Service Operations Management*, 20(3), 403-421.
- Deck, C., Farmer, A., & Zeng, D.-Z. (2007). Amended Final-Offer Arbitration over an Uncertain Value: A Comparison with CA and FOA. *Experimental Economics*, 10(4), 439-454.
- Deck, C. A., & Farmer, A. (2007). Bargaining over an Uncertain Value: Arbitration Mechanisms Compared. *Journal of Law, Economics, and Organization*, 23(3), 547-579.
- Drahozal, C. R., & Ware, S. J. (2010). Why Do Businesses Use (or Not Use) Arbitration Clauses. *Ohio State Journal on Dispute Resolution*, 25(2), 433-476.
- Eisenberg, T., Miller, G. P., & Sherwin, E. (2008). Arbitration's Summer Soldiers: An Empirical Study of Arbitration Clauses in Consumer and Nonconsumer Contracts Symposium: Empirical Studies of Mandatory Arbitration. *University of Michigan Journal of Law Reform*, 41(4), 871-896.
- Elmaghraby, W., & Katok, E. (2018). Behavioral research in competitive bidding and auction design. In K. Donohue, E. Katok, & S. Leider (Eds.), *The Handbook of Behavioral Operations*. Hoboken, NJ: Wiley.
- Falk, A., & Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, 96(5), 1611-1630.
- Farber, H. S. (1980). An analysis of final-offer arbitration. *Journal of Conflict Resolution*, 24(4), 683-705.
- Farber, H. S., & Katz, H. C. (1979). Interest Arbitration, Outcomes, and the Incentive to Bargain. *Industrial and Labor Relations Review*, 33(1), 55-63.
- Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2), 63-87.

- Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3), 159-181.
- Fehr, E., Hart, O., & Zehnder, C. (2011). Contracts as reference points—experimental evidence. *American Economic Review*, 101(2), 493-525.
- Fehr, E., Hart, O., & Zehnder, C. (2015). How do informal agreements and revision shape contractual reference points? *Journal of the European Economic Association*, 13(1), 1-28.
- Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics*, 108(2), 437-459.
- Fehr, E., Klein, A., & Schmidt, K. M. (2007). Fairness and contract design. *Econometrica*, 75(1), 121-154.
- Fugger, N., Katok, E., & Wambach, A. (2019). Trust in procurement interactions. *Management Science*, 65(11), 5110-5127.
- Gabuthy, Y., Jacquemet, N., & Marchand, N. (2008). Does Resorting to Online Dispute Resolution Promote Agreements? Experimental Evidence. *European Economic Review*, 52(2), 259-282.
- Gabuthy, Y., & Muthoo, A. (2019). Bargaining and Hold-Up: The Role of Arbitration. *Oxford Economic Papers*, 71(1), 292-308.
- Hart, O., & Moore, J. (1999). Foundations of Incomplete Contracts. *Review of Economic Studies*, 66(1), 115-138.
- Hart, O., & Moore, J. (2008). Contracts as reference points. *Quarterly Journal of Economics*, 123(1), 1-48.
- Herweg, F., & Schmidt, K. M. (2020). Procurement with Unforeseen Contingencies. *Management Science*, 66(5), 2194-2212.
- Herweg, F., & Schwarz, M. A. (2018). Optimal Cost Overruns: Procurement Auctions with Renegotiation. *International Economic Review*, 59(4), 1995-2021.
- Hoppe, E., & Schmitz, P. (2011). Can contracts solve the hold-up problem? Experimental evidence. *Games and Economic Behavior*, 73(1), 186-199.
- Katok, E., Shachat, J., & Walker, M. J. (2020). Trust and Trustworthiness in Procurement Contracts with Retainage. Unpublished manuscript.
- Kőszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics*, 121(4), 1133-1165.
- Moffatt, P.G. (2015) *Experimetrics: Econometrics for Experimental Economics*. Palgrave Macmillan.
- Pecorino, P., & Van Boening, M. (2001). Bargaining and Information: An Empirical Analysis of a Multistage Arbitration Game. *Journal of Labor Economics*, 19(4), 922-948.

- Shachat, J., & Tan, L. (2015). An experimental investigation of auctions and bargaining in procurement. *Management Science*, 61(5), 1036-1051.
- Waehrer, K. (1995). A Model of Auction Contracts with Liquidated Damages. *Journal of Economic Theory*, 67(2), 531-555.
- Wang, R. (2000). Bidding and Renegotiation in Procurement Auctions. *European Economic Review*, 44(8), 1577-1597.

Appendix A: Incentive compatibility constraint with a contingent contract.

There are three cases to consider for the incentive compatibility of high quality at Stage 3 of our competitive procurement interaction, which depend on the cost of arbitration k in relation to the arbitration probability σ , the arbitrator's preferences μ and the lower bound contract price \underline{p} . The arbitrator's decision rule z^q is defined in equation (3) of the main text.

Case (i): $k \geq \frac{(1-\sigma)}{\sigma} (z^H(\mu) - \underline{p})$.

High quality is incentive compatible if $z^H(\mu) - \max\{\underline{p}, z^L(\mu)\} > c(q^H) - c(q^L)$.

Proof. The buyer proposes a trading price equal to $z^H(\mu)$ in exchange for high quality and $\max\{\underline{p}, z^L(\mu)\}$ in exchange for low quality. The seller opts not to invoke the arbitration clause at Stage 5 and so incentive compatibility follows directly from the associated delivery costs $c(q)$. ■

Case (ii): $\frac{(1-\sigma)}{\sigma} (z^L(\mu) - \underline{p}) < k < \frac{(1-\sigma)}{\sigma} (z^H(\mu) - \underline{p})$.

High quality is incentive compatible if $\sigma > \frac{c(q^H) - c(q^L) + \max\{\underline{p}, z^L(\mu)\} - \underline{p}}{z^H(\mu) - \underline{p}}$.

Proof. The buyer proposes a trading price equal to \underline{p} in exchange for high quality and $\max\{\underline{p}, z^L(\mu)\}$ in exchange for low quality. The seller opts to invoke the arbitration clause at Stage 5 only if he delivers high quality. In that scenario, the buyer bears the direct cost of arbitration. The incentive-compatibility condition follows directly from the associated delivery costs $c(q)$ and the probability of arbitration σ . ■

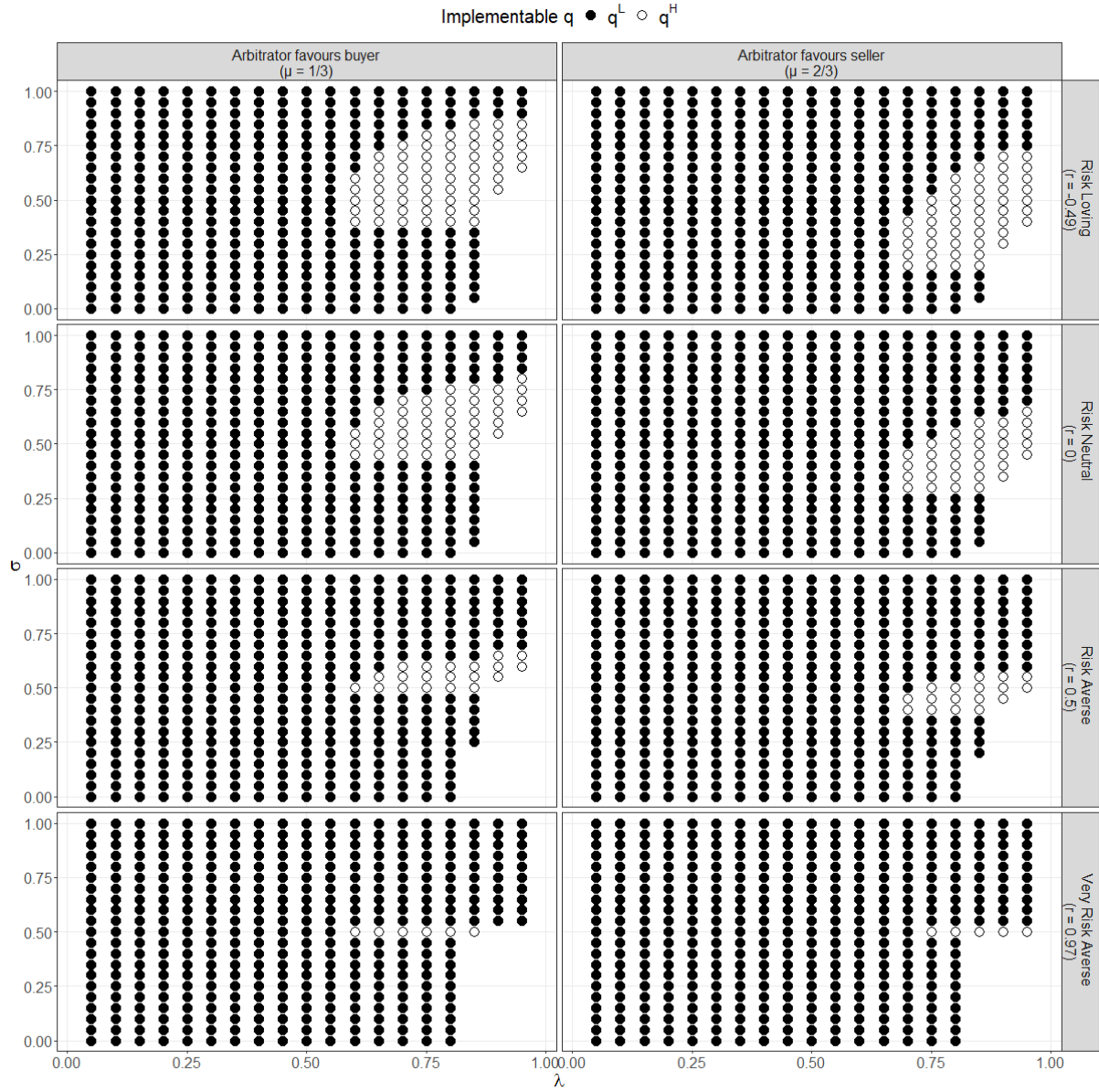
Case (iii): $k \leq \frac{(1-\sigma)}{\sigma} (z^L(\mu) - \underline{p})$.

High quality is incentive compatible if $\sigma > \frac{c(q^H) - c(q^L)}{z^H(\mu) - z^L(\mu)}$.

Proof. This case exists if and only if $z^L(\mu) \geq \underline{p}$. The buyer proposes a trading price equal to \underline{p} both in exchange for high quality and low quality. The seller opts to invoke the arbitration clause at Stage 5 regardless of the quality level. The buyer again bears the direct cost of arbitration and the incentive-compatibility condition follows directly. ■

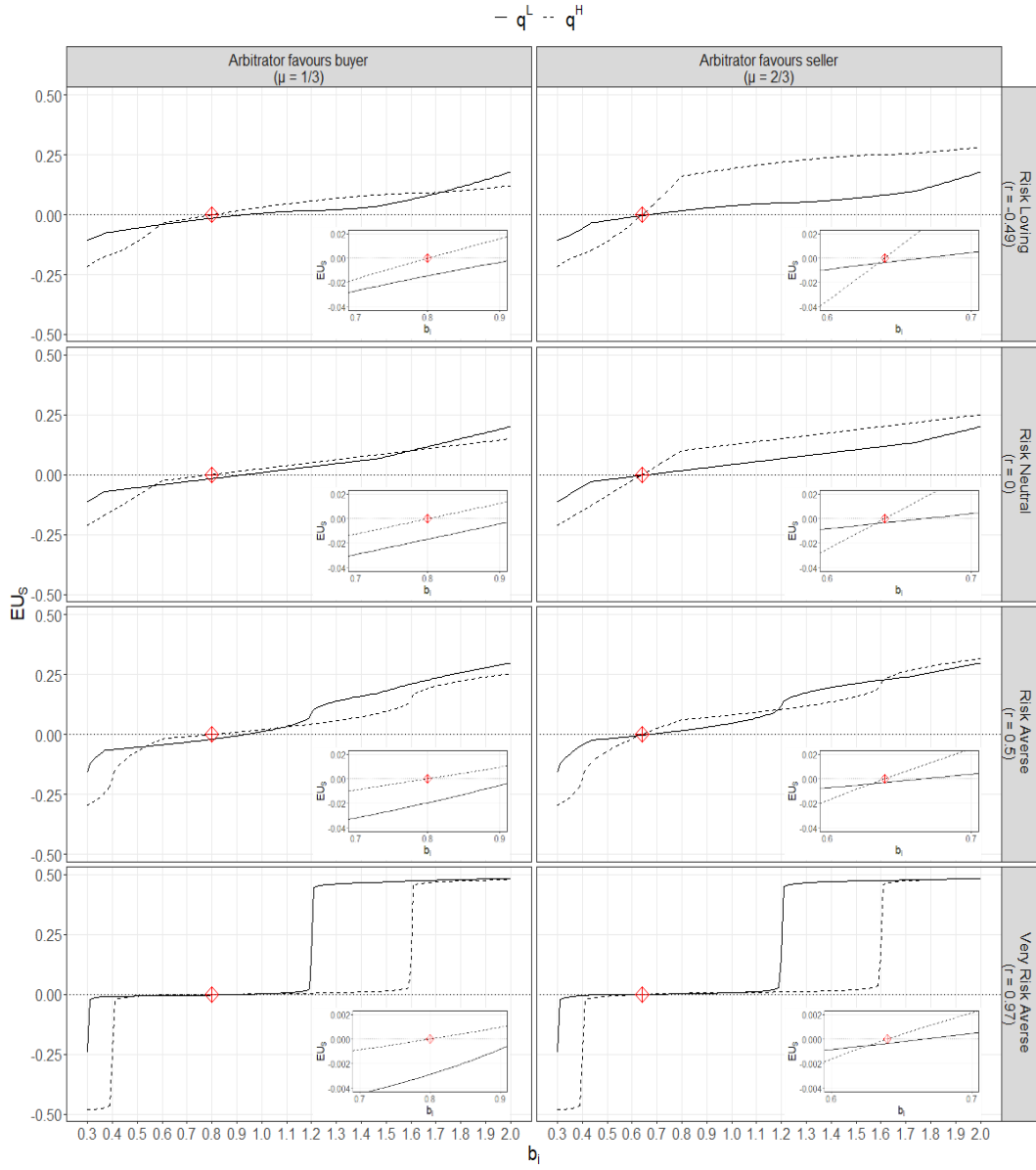
Appendix B: Additional figures and tables.

Figure A. 1. Implementable equilibrium quality by contract for the CRRA utility case (experiment parameters are without loss of generality).



Notes: The figure assumes that potential sellers formulate their bidding strategies based on the CRRA utility function $u(\pi_s) = \pi_s^{(1-r)}/(1-r)$ for $\pi_s \geq 0$ and $u(\pi) = -(-\pi_s)^{(1-r)}/(1-r)$ for $\pi_s < 0$, where r measures the degree of relative risk aversion. A contract is defined by the probability of arbitration σ and the degree of flexibility λ . The empty regions correspond to contracts for which no profitable bidding strategy exists, given the maximum bid limit in the experiment.

Figure A. 2. Expected seller dispute utilities conditional on submitting the winning bid (*Equilibrium Implementability Condition* is satisfied in *Arbitrator* for all levels of risk aversion).



Notes: The figure assumes that potential sellers formulate their strategies based on the CRRA utility function $u(\pi_s) = \pi_s^{(1-r)}/(1-r)$ for $\pi_s \geq 0$ and $u(\pi) = -(-\pi_s)^{(1-r)}/(1-r)$ for $\pi_s < 0$, where r measures the degree of relative risk aversion. The red diamond plus is the equilibrium bid level. Inset in each panel is the subset of the discrete price grid in which the *Equilibrium Implementability Condition* is satisfied.

Figure A. 3. CDFs of seller auction bids in the experiment (supports Results 1(a) and 1(b)).

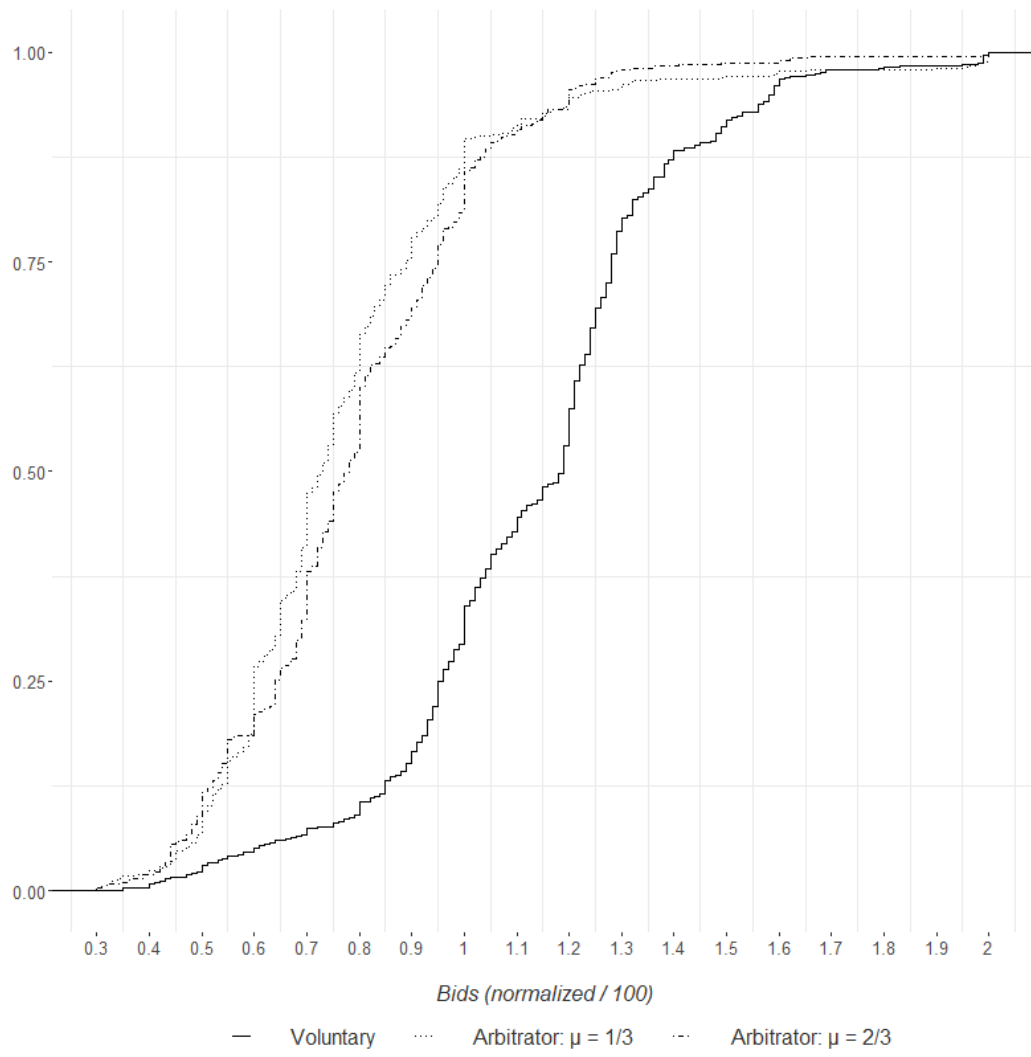
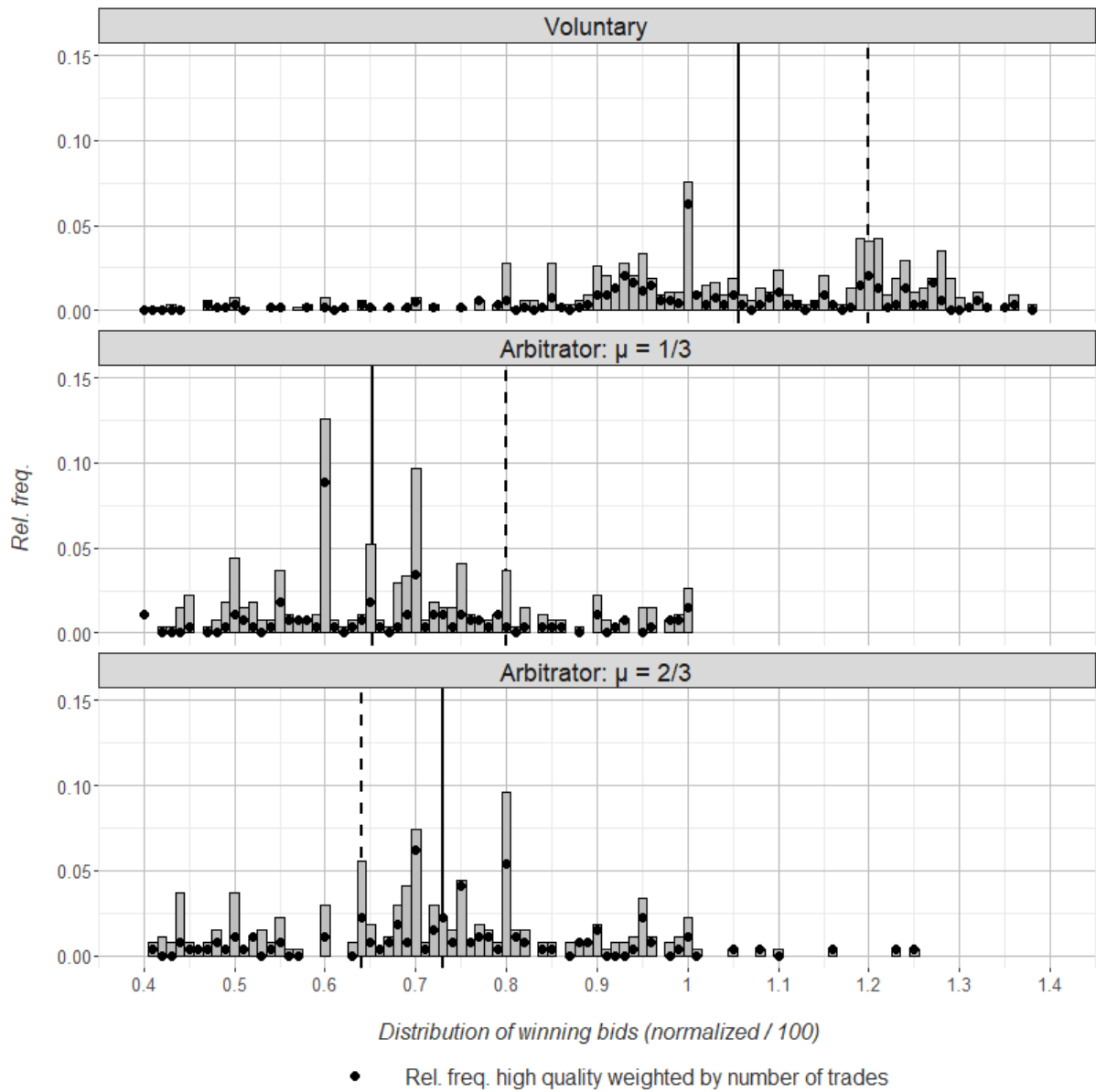
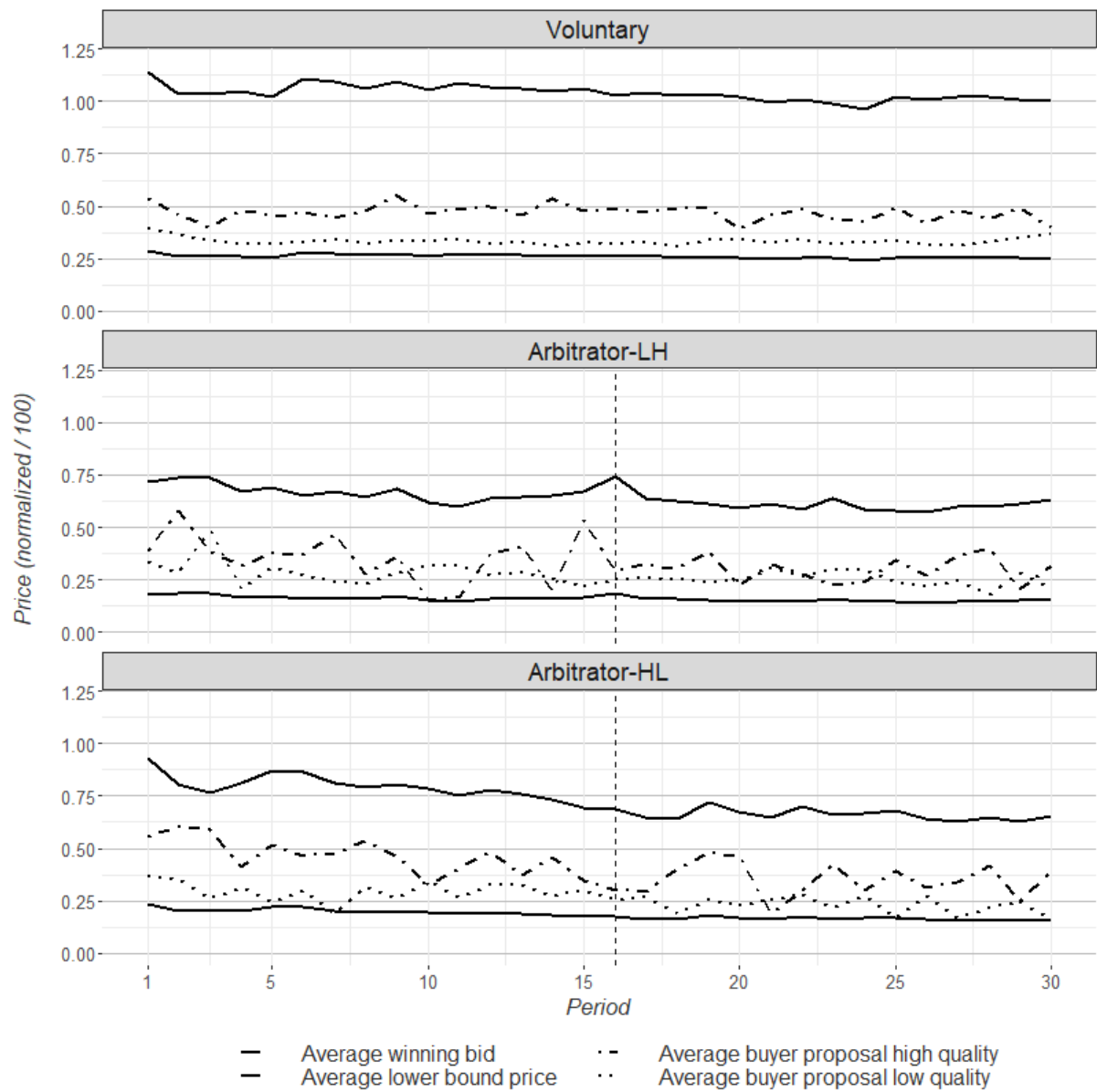


Figure A. 4. Distribution of winning bids and the relative frequency of high quality in the experiment.



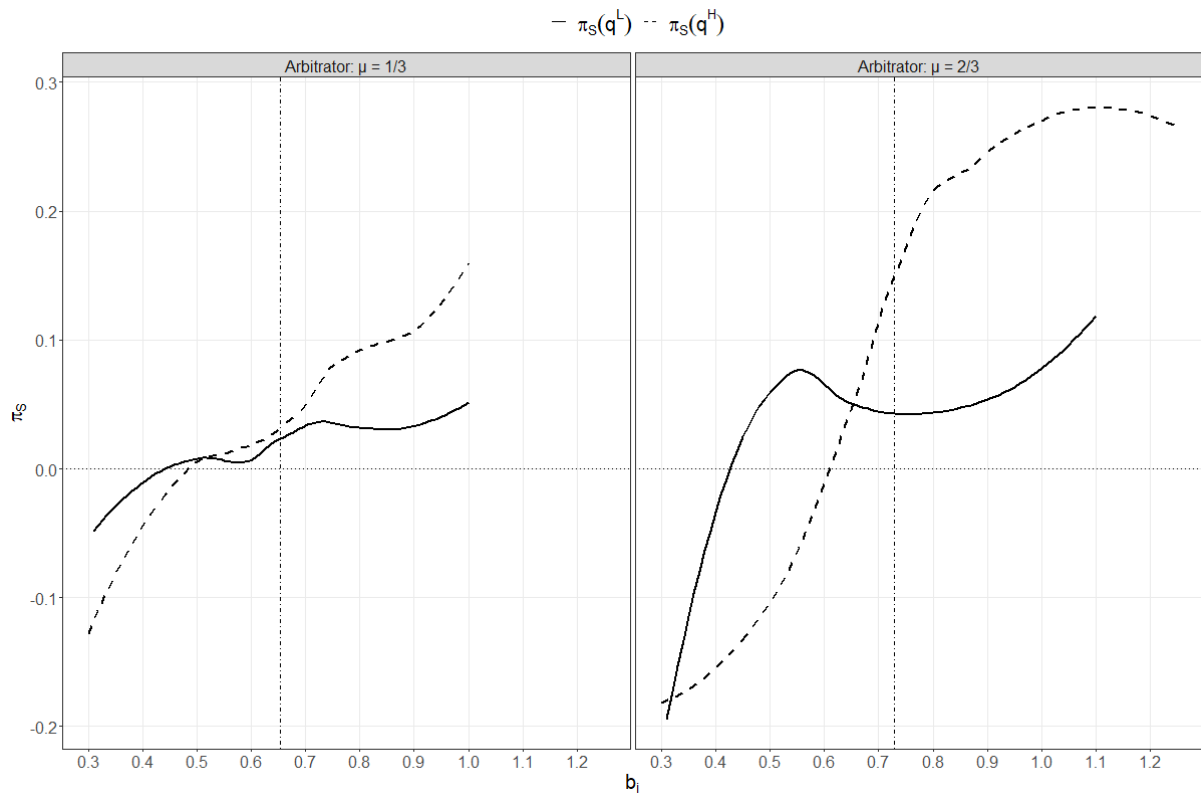
Notes: The vertical dashed line is the equilibrium bid level. The vertical solid line is the median winning bid observed in the experiment.

Figure A. 5. Contract price ranges and buyer reciprocity over time in the experiment.



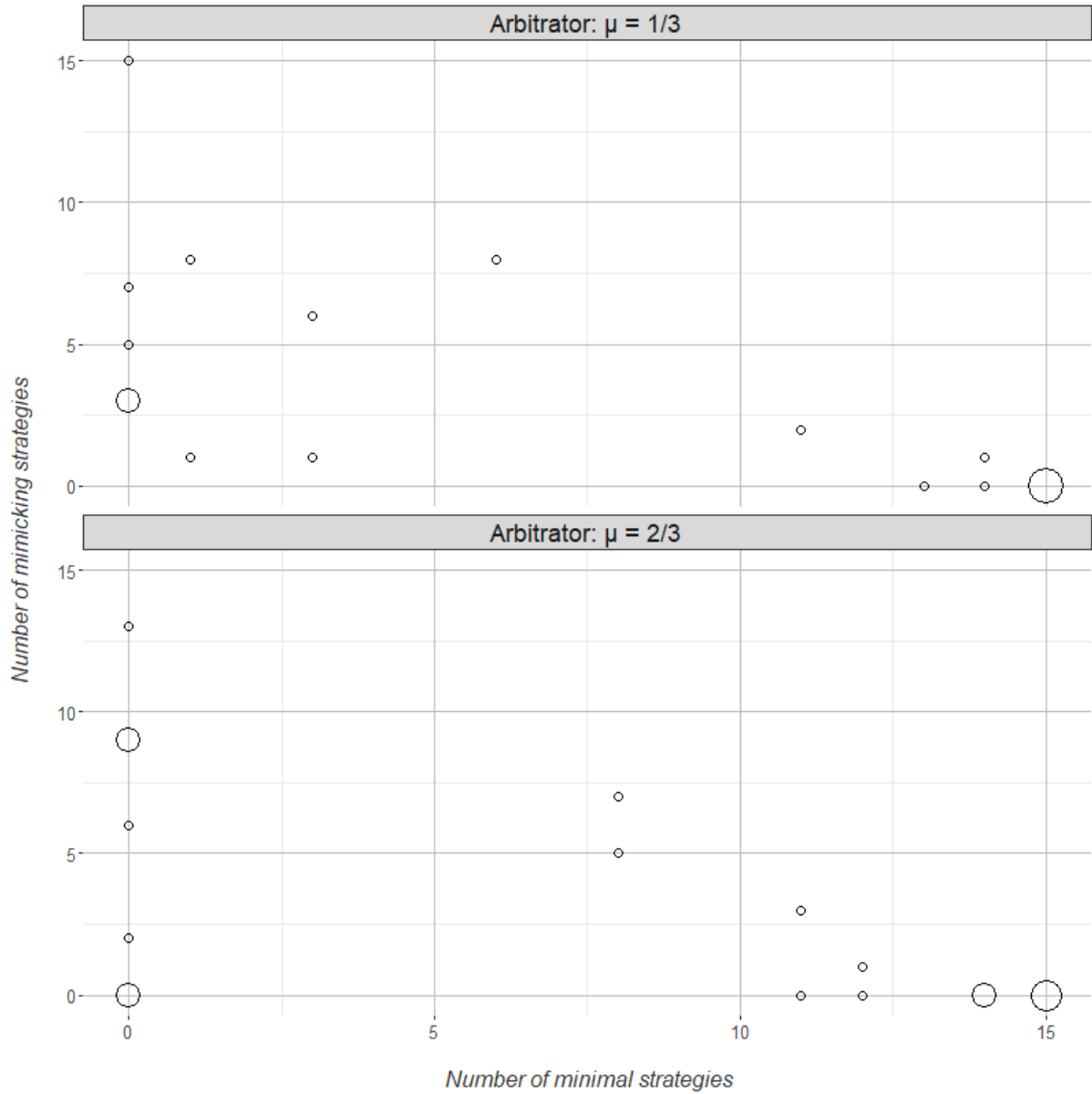
Notes: The vertical dashed line in the *Arbitrator* treatment panels represents the change in the arbitrator's preference parameter μ after period fifteen.

Figure A. 6. Seller profits as a function of the winning bid observed in the *Arbitrator* treatment:
High quality is incentive compatible on average.



Notes: The vertical dot-dash line is the median winning bid observed in the respective experiment condition.

Figure A. 7. Individual buyer strategy counts in the *Arbitrator* treatment.



Notes: Based on individual buyer decisions over the course of the experiment. The size of the open circles is weighted by the number of buyers at the coordinate. Minimal strategies are defined as buyer choices to propose the lower bound contract price. Mimicking strategies are defined as buyer choices to propose the arbitrator's interim preferred price $p_{it}^{A'} \pm 2$. We permit a tolerance of two points either side of the arbitrator's interim preferred price to reflect the granularity of the buyer's price proposal slider. This slider was set at increments of 0.25 points, which did not always permit the arbitrator's preferred price to be selected exactly.

Table A. 1 – Summary of experiment outcomes from periods one to fifteen only in the *Arbitrator* treatment (relates to behavioural spillover analysis in Section 5.1).

Measure	Treatment	<i>Arbitrator-LH</i> ($\mu = 1/3$)	<i>Arbitrator-HL</i> ($\mu = 2/3$)
(i) <i>Winning bid</i>		71.16	75.00
(ii) <i>Product quality</i>		0.42	0.56
(iii) <i>Proposal / surplus</i>		-0.16	0.01
	Low quality	-0.19	0.04
	High quality	-0.12	-0.04
(iv) <i>Rel. freq. dispute</i>		0.78	0.70
	Low quality	0.72	0.68
	High quality	0.85	0.79
(v) <i>Final price</i>		39.22	53.79
	Low quality	34.86	37.24
	High quality	44.12	60.35
(vi) <i>Buyer profit</i>		31.80	23.19
	Low quality	14.50	12.09
	High quality	55.04	38.56
(vii) <i>Seller profit</i>		4.78	16.01
	Low quality	4.86	7.24
	High quality	4.12	20.35
(viii) <i>Efficiency</i>		60%	69%

Notes: The table summarises the main outcomes of the *Arbitrator* treatment based on data from periods one to fifteen only. For $\mu = 1/3$, this corresponds to sessions assigned to *Arbitrator-LH* and for $\mu = 2/3$ to sessions assigned to *Arbitrator-HL*. This precludes the possibility of behavioural spillover effects between the first and second half of the experiment. All values are median cohort averages, based on three independent cohorts per treatment. (i) *Winning bid* is the lowest bid submitted by potential sellers in the auction phase; (ii) *Product quality* measures the relative frequency that high quality is chosen for delivery by the winning seller; (iii) *Proposal / surplus* measures the buyer's price proposal as a share of trade surplus; (iv) *Rel. freq. dispute* measures how often an appeal to the arbitrator is triggered (in *Arbitrator* only); (v) *Final price* is the trading price exchanged, irrespective of whether this price is determined by the buyer or an arbitrator; (vi) *Buyer profit* is the trade profit per round for the buyer; (vii) *Seller profit* is the trade profit per round for the winning seller; and (viii) *Efficiency* measures the realised percentage of attainable trade surplus net of any arbitration cost incurred.

Table A. 2 – Average seller bids by quality level in the experiment.

Treatment	<i>Voluntary</i>	<i>Arbitrator</i>	
Sub-treatment		$\mu = 1/3$	$\mu = 2/3$
Low quality strategies	118.35	73.29	78.95
High quality strategies	109.71	78.66	85.38

Notes: The table presents seller bids from the strategy method data, by the quality level selected. All values are median cohort averages, based on six independent cohorts per treatment.

Table A. 3 – Frequencies of buyer choices in the *Arbitrator* treatment.

Sub-treatment	$\mu = 1/3$	$\mu = 2/3$
Minimal	0.468	0.508
Low quality	0.471	0.496
High quality	0.465	0.518
Mimicking	0.223	0.207
Low quality	0.232	0.216
High quality	0.211	0.199

Notes: The table displays relative frequencies of buyer choices to propose the lower bound contract price (minimal) and the arbitrator's interim preferred price $p_{it}^{A'} \pm 2$ (mimicking). We permit a tolerance of two points either side of the arbitrator's interim preferred price to reflect the granularity of the buyer's price proposal slider. This slider was set at increments of 0.25 points, which did not always permit the arbitrator's preferred price to be selected exactly. The missing categories are price proposals between minimal and mimicking, and price proposals above mimicking.

Table A. 4 – IV first stage estimates: Determinants of quality level in seller bidding strategies (relates to Section 5.2 in the main text).

Dependent variable Estimation method Treatment	<i>Bid</i> <i>IV</i>	
	<i>Voluntary</i>	<i>Arbitrator</i>
<i>Bid_{t-1}</i>	0.603*** (0.050)	0.412*** (0.066)
<i>Period</i>	-0.129 (0.082)	-0.279** (0.120)
$\mu = 2/3$		0.654 (1.411)
<i>OrderHL</i>		-5.712** (2.618)
<i>Constant</i>	39.004*** (6.646)	68.752*** (11.402)
Control variables	Yes	Yes
Cohort fixed effects	Yes	Yes
Observations	1,028	970
Subjects	36	36
Cohorts	6	6
F test of excluded instruments	142.04***	38.44***
Cragg-Donald Wald F statistic	615.13	181.72

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Coefficient estimates, with robust standard errors clustered at the subject level. Cohort fixed effects are included. The following control variables (not shown) are included: dummy for being female; dummy for being an economics and finance major; two Likert scales for self-reported willingness to take risks in general and in financial matters; dummy for reporting trust in strangers; and a generalised trust index. *OrderHL* is a dummy variable to indicate that the cohort followed the sequence $\mu = \{2/3, 1/3\}$ in the *Arbitrator* treatment. Periods 1 and 16 are excluded due to the within-subjects design.

Table A. 5 – Determinants of buyer price proposals (no significant order effects).

Dependent variable Estimation method Treatment	Price proposal <i>Multilevel</i>	
	<i>Voluntary</i>	<i>Arbitrator</i>
<i>Lower bound contract price</i>	0.493*** (0.077)	1.347*** (0.168)
<i>Product quality</i>	13.168*** (0.649)	9.389*** (0.988)
<i>Period</i>	-0.037 (0.036)	-0.263*** (0.062)
$\mu = 2/3$		1.428 (0.978)
<i>OrderHL</i>		8.940 (6.475)
<i>Constant</i>	45.201*** (17.029)	-62.141** (25.577)
Control variables	Yes	Yes
Observations	535	535
Subjects	18	18
Cohorts	6	6
Wald χ^2	454.6	287.7
χ^2 p-value	0.000	0.000
σ_v	0.000	0.000
σ_u	6.881	9.549
σ_ε	7.118	11.09

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The models are estimated using mixed effects linear regression and include two random effects intercepts that capture intra-session correlation at the subject and cohort levels, respectively. Coefficient estimates are presented, with standard errors in parentheses. The following control variables (not shown) are included: dummy for being female; dummy for being an economics and finance major; two Likert scales for self-reported willingness to take risks in general and in financial matters; dummy for reporting trust in strangers; and a generalised trust index. *OrderHL* is a dummy variable to indicate that the cohort followed the sequence $\mu = \{2/3, 1/3\}$ in the *Arbitrator* treatment.

Appendix C: Robustness check of aggregate experiment findings.

In this section of the appendix, I validate Results 1 to 4 (see Section 5.1 of the main text) using an OLS regression procedure based on disaggregated data at the matching group level. There are 540 matching groups in each treatment. Data from all periods are included and we incorporate a time trend. There are dummies for each of the *Arbitrator* sub-treatments. The *Voluntary* treatment is omitted. Robust standard errors are clustered at the cohort level to control for intra-cohort dependencies – matching groups are randomly reconstituted each round within a cohort. Since the number of clusters is small (six), standard asymptotic inference is liable to over-reject (Cameron et al. 2008). Thus, I also present p -values using wild cluster bootstrap tests of linear hypotheses (Roodman et al. 2019). The results are contained in Table A. 6.

Column 1 of the table supports Results 1(a) and 1(b). On average, winning auction bids are 34 to 38 points lower in the respective *Arbitrator* sub-treatments than in the *Voluntary* treatment (both p -values < 0.01 , cluster bootstrap- t tests). There is some evidence of a decreasing trend in bids over time (p -value = 0.098, cluster bootstrap- t test). The difference in winning bids between the *Arbitrator* sub-treatments is not in the direction predicted by the theory.

Result 2 states that there is no significant difference in trade efficiency between a contingent and non-contingent contract. Since this insight is based on a null finding, it is important to check whether the null result still obtains in the matching group data. This is of particular interest for the comparison between *Voluntary* and *Arbitrator* when $\mu = 2/3$. Indeed, a power calculation suggests that this pairwise comparison at the cohort-level only has around 13 percent power to detect an effect size of the magnitude observed in the data.¹ Column 2 of the table supports Result 2. There is no significant increase in trade efficiency with a contingent contract than without ($\mu = 1/3$, p -value = 0.399; $\mu = 2/3$, p -value = 0.699; score bootstrap- t tests). This does not change significantly over time in the experiment (p -value = 0.518, score bootstrap- t test). We fail to reject a linear hypothesis test that the coefficient estimates on the *Arbitrator* sub-treatment dummies are equal (p -value = 0.404, score bootstrap- t test).

Column 3 of the table supports Result 3. Buyer price proposals as a proportion of trade surplus are significantly lower with a contingent contract, both when $\mu = 1/3$ (p -value < 0.01 , cluster bootstrap- t test) and when $\mu = 2/3$ (p -value = 0.011, cluster bootstrap- t test). This is consistent with a crowding out of buyer reciprocity. The period indicator in column 3 is also negative and significant (p -value < 0.01 , cluster bootstrap- t test), indicating decreasing levels of buyer reciprocity over time.

¹ Mean (sd) in *Voluntary* of 0.639 (0.096) and in *Arbitrator* $\mu = 2/3$ of 0.674 (0.098). All power calculations are based on 1000 replications and assume a normal distribution.

Table A. 6 – Regression analysis of matching group data.

Dependent variable	<i>Winning bid</i>	<i>Trade efficiency</i>	<i>Proposal / surplus</i>	<i>Buyer Profit</i>	<i>Seller Profit</i>
Validation of Result	1	2	3	4	4
<i>Arbitrator</i> ($\mu = 1/3$)	-37.63*** (7.56) [0.009]	-0.151 (0.162) [0.399]	-0.308*** (0.039) [0.006]	-0.431 (2.32) [0.870]	-1.70 (1.05) [0.153]
<i>Arbitrator</i> ($\mu = 2/3$)	-33.52*** (8.28) [0.013]	0.157 (0.335) [0.699]	-0.215*** (0.039) [0.011]	-1.84 (3.33) [0.597]	3.96 (2.53) [0.230]
<i>Period</i>	-0.46 (0.28) [0.081]	0.008 (0.103) [0.518]	-0.008*** (0.001) [0.005]	0.373** (0.108) [0.007]	-0.258** (0.097) [0.077]
<i>Constant</i>	111.26*** (3.51)	0.440*** (0.085)	0.262*** (0.050)	27.48*** (1.58)	9.050*** (1.89)
Observations	1,080	1,070	1,070	1,070	1,070
Wald test statistic	41.61***	26.97***	51.76***	9.40**	4.50*

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Coefficient estimates are presented, with robust standard errors in parentheses clustered at the cohort level to correct for intra-cohort correlation. The omitted experimental condition is the *Voluntary* treatment. In square brackets are wild cluster bootstrap p -values after imposing the zero null on the bootstrap data generating process, with 999 bootstrap samples drawing from the six-point Webb distribution – which is preferred to alternative distributions when the number of clusters is less than ten – obtained using the *boottest* command in Stata (Roodman et al. 2019). The models in columns 1, 3, 4 and 5 are estimated using OLS regression. The model in column 2 was estimated using fractional logistic regression and coefficient estimates are on the logit scale. The Wald test statistic is based on a test of joint regression significance (F-statistics except for column 2, which is a chi-squared statistic). The number of observations is higher in the first column because we have auction data from all matching groups, whereas trade data from 10 matching groups was lost due to a hard time-out protocol.

Columns 4 and 5 of the table are in line with Result 4. Buyer profits do not differ significantly with a contingent contract than without and there is no significant difference between the *Arbitrator* sub-treatments (p -value = 0.625, cluster bootstrap-t test). Of more interest is the comparison in seller profits between *Voluntary* and *Arbitrator* when $\mu = 2/3$. There was only weak evidence at the aggregate level to suggest that the contingent contract increases seller profits under these conditions, but this test only has around 23 percent power to detect an effect size of the magnitude observed in the aggregate data.² At the matching group level, the evidence for a positive effect is stronger but still not significant at conventional thresholds (p -value = 0.230, cluster bootstrap-t test). We also fail to reject a linear hypothesis test that the coefficient estimates on the dummy variables for the *Arbitrator* sub-treatments are equal in column 5 (p -value = 0.163, cluster bootstrap-t test). The buyer-seller profit differential widens over the course of the experiment, with a significant positive time trend in the specification for buyer profits (p -value < 0.01, cluster bootstrap-t test) and a significant negative time trend in the specification for seller profits (p -value = 0.077, cluster bootstrap-t test).

² Mean (sd) in *Voluntary* of 5.09 (3.85) and in *Arbitrator* $\mu = 2/3$ of 9.06 (9.27).

Appendix D: Mixture model estimation procedure (relates to Section 5.3 in the main text).

The finite mixture model estimation procedure used here is adapted from Moffatt (2015). There are 18 buyers in our *Arbitrator* treatment, each of whom has been observed over 30 trading periods. Buyer i 's price proposal in period t is given by y_{it} . The variable y_{it} is bounded by the contract price range, $[\underline{p}_{it}, \bar{p}_{it}]$, which is determined in the auction phase of the period and is observed before the buyer makes her decision. The arbitrator's interim preferred price is given by:

$$p_{it}^{A'} = \max \left\{ \underline{p}_{it}, \min \left\{ \bar{p}_{it}, \mu_t v(q_{it}) + (1 - \mu_t) c(q_{it}) \right\} \right\}. \quad (\text{A.1})$$

This price is determined by the seller's product quality choice q_{it} , the arbitrator's exogenous preference parameter μ_t , which varies during the experiment, and the cost and valuation schedules. It is constrained by the bounds of the contract price range. The experiment parameter values are described in the main text.

The buyer's preferred price is y_{it}^* , which depends on her type. We assume that there are two buyer types: self-interested (*self*) and fair (*fair*). The model precludes subjects from switching between types. Self-interested buyers propose the lower bound contract price, \underline{p}_{it} . This type consists of strictly self-interested economic agents. Fair buyers are assumed to exhibit reference-dependent fairness preferences. For this type, we specify the following latent model:³

$$y_{it}^* = \beta_0 + \beta_1 p_{it}^{A'} + \varepsilon_{it, \text{fair}} \quad (\text{A.2})$$

$$i = 1, \dots, 18, \quad t = 1, \dots, 30, \quad \varepsilon_{it, \text{fair}} \sim N(0, s^2).$$

Where the arbitrator's interim preferred price is the fair reference point in a trading period. We expect the parameter β_1 to be close to one for fair buyers. The error term is assumed to be normally distributed, with variance s^2 .

A two-limit tobit model (Nelson 1976) is appropriate, where the limits correspond to the lower and upper bounds of the contract price range and are specific to each buyer and period. The relationship between the buyer's preferred price proposal and actual price proposal is determined by the following censoring rules:

For self-interested buyers:

$$y_{it} = \underline{p}_{it} \quad \forall t \quad (\text{A.3a})$$

³ A period indicator was found not to be a significant contributor to the latent model and so was dropped from the equation.

For fair buyers:

$$y_{it} = \begin{cases} \underline{p}_{it} & \text{if } y_{it}^* \leq \underline{p}_{it} \\ y_{it}^* & \text{if } \underline{p}_{it} < y_{it}^* < \underline{\bar{p}}_{it} \\ \underline{\bar{p}}_{it} & \text{if } y_{it}^* \geq \underline{\bar{p}}_{it} \end{cases} \quad (\text{A.3b})$$

We further incorporate the possibility that buyers deviate on occasion from their preferred price by incorporating a tremble parameter ω . In any given period, with probability ω a buyer may lose concentration and choose a price proposal at random. The price proposal increment in the experiment was set at 0.25 points and so the cardinality of the set of possible price proposals is $|Y_{it}| = 1 + (\underline{\bar{p}}_{it} - \underline{p}_{it})/0.25$. The importance of this parameter is likely decreasing during the experiment as subjects gain experience in the decision-making environment and so we employ the specification:

$$\omega_{it} = \omega_0 \exp [\omega_1 (\text{Period})] \quad (\text{A.4})$$

where ω_0 is the initial tremble probability and ω_1 is the rate of decline, which we expect to be negative. These tremble parameter estimates, however, turn out not to be statistically different from zero and so we do not report on them further here.

The likelihood contributions for a single price decision are as follows, where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal cumulative distribution and probability density functions respectively:

$y = \underline{p}$:

$$P(y_{it} = \underline{p}_{it} | i = \text{self}) = 1 - \frac{\omega_{it}}{|Y_{it}|} \quad (\text{A.5a})$$

$$P(y_{it} = \underline{p}_{it} | i = \text{fair}) = (1 - \omega_{it}) \Phi\left(-\frac{\beta_0 + \beta_1 p_{it}^{A'}}{s}\right) + \frac{\omega_{it}}{|Y_{it}|}$$

$\underline{p} < y < \underline{\bar{p}}$:

$$f(y_{it} | i = \text{self}) = \frac{\omega_{it}}{|Y_{it}|} \quad (\text{A.5b})$$

$$f(y_{it} | i = \text{fair}) = (1 - \omega_{it}) \frac{1}{s} \phi\left(\frac{y_{it} - \beta_0 - \beta_1 p_{it}^{A'}}{s}\right) + \frac{\omega_{it}}{|Y_{it}|}$$

$y = \underline{\bar{p}}$:

$$P(y_{it} = \bar{p}_{it} | i = self) = \frac{\omega_{it}}{|Y_{it}|}$$

$$P(y_{it} = \bar{p}_{it} | i = fair) = (1 - \omega_{it}) \left[1 - \Phi \left(\frac{\bar{p}_{it} - \beta_0 - \beta_1 p_{it}^{A'}}{s} \right) \right] + \frac{\omega_{it}}{|Y_{it}|}$$
(A.5c)

The two buyer types are represented in the model by the mixing fractions fr_{self} and fr_{fair} . The likelihood contribution for a buyer i is:

$$L_i = fr_{self} \prod_{t=1}^{30} P(y_{it} = \underline{p}_{it} | self)^{I_{y_{it}=\underline{p}}} f(y_{it} | self)^{I_{\underline{p} < y_{it} < \bar{p}}} P(y_{it} = \bar{p}_{it} | self)^{I_{y_{it}=\bar{p}}}$$

$$+ fr_{mimic} \prod_{t=1}^{30} P(y_{it} = \underline{p}_{it} | fair)^{I_{y_{it}=\underline{p}}} f(y_{it} | fair)^{I_{\underline{p} < y_{it} < \bar{p}}} P(y_{it} = \bar{p}_{it} | fair)^{I_{y_{it}=\bar{p}}}$$
(A.6)

Where $I(.)$ is an indicator function for the subscripted expression and the conditional probabilities/densities are substituted from (A.5).

Finally, the sample log-likelihood is obtained as follows:

$$Log L = \sum_{i=1}^{18} \log(L_i)$$
(A.7)

on maximization of which we obtain maximum likelihood estimates for the five parameters $\beta_0, \beta_1, s, \omega_0, \omega_1$ and one of the two mixing fractions (from which the other is deduced using the delta method).

Posterior probabilities for the two types can then be calculated as:

$$P(i = self | y_{i1}, \dots, y_{i30})$$

$$= \frac{fr_{self} \prod_{t=1}^{30} P(y_{it} = \underline{p}_{it} | self)^{I_{y_{it}=\underline{p}}} f(y_{it} | self)^{I_{\underline{p} < y_{it} < \bar{p}}} P(y_{it} = \bar{p}_{it} | self)^{I_{y_{it}=\bar{p}}}}{L_i}$$
(A.8a)

$$P(i = fair | y_{i1}, \dots, y_{i30})$$

$$= \frac{fr_{fair} \prod_{t=1}^{30} P(y_{it} = \underline{p}_{it} | fair)^{I_{y_{it}=\underline{p}}} f(y_{it} | fair)^{I_{\underline{p} < y_{it} < \bar{p}}} P(y_{it} = \bar{p}_{it} | fair)^{I_{y_{it}=\bar{p}}}}{L_i}$$
(A.8b)

The estimation is conducted in Stata 16, using the d0 estimator to account for the panel structure of the data. A starting value for the mixing fraction is obtained from the frequency table of buyer choices in Table A. 3. Starting values for the five parameters are (0, 1, 4, 0.11, -0.05), obtained from a process of trial and error. Bootstrapped standard errors are calculated, clustered at the cohort level based on 999 bootstrap replications.

Appendix E: Experiment instructions.

The experiment video instructions can be found online at the [external link: [project repository](#)]. A written version is provided below. Note: Horizontal sliders were used for the suppliers' bidding decision and the shipping company's price proposal. The initial values of these sliders were set at random in each interaction to avoid anchoring bias.

You are now participating in an experiment in the economics of decision making. Based on your decisions and the decisions of other participants in the experiment, you can earn money which will be paid to you in private and in cash. Read these instructions carefully.

Please turn to silent mode your cell phone or any other electronic device that you have brought with you. These electronic devices must be stored out of sight and not used for the duration of the experiment. Please do not communicate with other participants. If at any point you have a question, raise your hand and we will answer you as soon as possible.

How you earn money

The experiment today consists of **30 decision-making rounds**. You will be paid for all these rounds. You earn money during the experiment by accruing points in each round. All the points that you earn during the experiment are converted to pounds sterling at the following rate:

40 points = £1

At the end of the experiment, you will be asked to complete a short questionnaire. Upon completion, you will immediately receive the monetary amount that you earned. The minimum amount that you will leave today's session with is £4.

Summary of the experiment procedure

For today's session, one third of you will be randomly assigned to the role of a **shipping company**, and two thirds of you assigned to the role of a **supplier**. You will be informed to which role you are assigned before the first round. You remain in the same role for the entire experiment.

In each of the 30 rounds, a shipping company is matched with two suppliers in the room. This is called a matching group. The shipping company is looking to buy an engine from **one** of the two suppliers in his/her matching group.

Note: Your matching group changes in each round. You are not matched with the same persons in consecutive rounds.

All subjects assigned to a shipping company role begin each round with an endowment of 5 points. All subjects assigned to a supplier role begin each round with an endowment of 10 points.

----Voluntary----

There are two phases in each round:

1. Auction phase
2. Price-setting phase

----Arbitrator----

There are three phases in each round:

1. Auction phase
2. Price proposal phase
3. Arbitration phase.

In the auction phase, it is decided which supplier delivers the engine. Each supplier submits a bid. The **lower** bid wins the auction. The winning supplier at the auction delivers the engine. The losing supplier at the auction exits the round.

The engine delivered to the shipping company can be of **high or low quality**. This is decided by the winning supplier during the auction and determines the shipping company's value for the engine and the winning supplier's delivery cost (see Table 1).

Table - 1: Engine value and cost schedule

	Low quality	High quality
Shipping company's value	50	100
Winning supplier's cost	30	40

----Voluntary----

In the price-setting phase, the shipping company observes the engine quality delivered. Then, the shipping company chooses a final price to pay the winning supplier. This price can be any amount from **one quarter** of the winning auction bid up to and including the winning auction bid.

At the end of the round, the points earned are calculated as follows:

Shipping company earns 5 + Profit, where Profit = value - final price
Winning supplier earns 10 + Profit, where Profit = final price - cost
Losing supplier earns 10 (Profit = 0)

Note: Profit can be negative. Make your decisions carefully.

We now explain the experiment procedure in detail.

Auction phase:

- The two suppliers in a matching group each submit a bid at an auction.
- This bid can be any whole number from 30 up to and including 200.
- Neither of the two suppliers can see the other's choice of bid.
- The supplier that submits the **lower** bid wins the auction. In the event of a tie, the computer chooses the winner at random.
- The winning bid is not necessarily the final price paid by the shipping company. Instead, the winning bid determines the **price range** that the winning supplier can receive:
➔ the **minimum price** in this range is one quarter of the winning bid.

➔ the **maximum price** in this range is the winning bid.

- At the same time as submitting a bid, each supplier also selects an engine quality (high or low) to be delivered in the event of winning the auction.
- The winning supplier at the auction proceeds to the next phase and incurs his/her engine delivery cost.
- The losing supplier at the auction exits and earns the round endowment of 10 points (a profit of zero).

Computer interface:

In the auction phase, each supplier submits his/her bid by adjusting a slider in the upper part of the computer screen. As you adjust the slider, underneath you see the price ranges associated with the different possible bids. The first number in the brackets is the minimum price, which is one quarter of the selected bid. The second number is the maximum price, which is the bid itself.

A supplier submits his/her quality by selecting one of two options in the lower part of the screen.

On the right-hand side of the screen appears the engine value and cost schedule.

Suppliers should attempt to make their decisions in the prescriptive time. For the first five rounds, you have 60 seconds to make your decisions. After round five, you have 30 seconds to make your decisions. If you cannot make your decisions in time, the default bid is 200 and the engine quality is selected at random.

Auction phase computer screen:

Time left to complete this page: 0:26

Round 1 out of 30

You are a Supplier.

1) Please select a bid.

Adjust the slider to your preferred bid between 30 and 200.
The price range for that bid will appear underneath.

128

Price range

[32 , 128]

2) Please select an engine quality.

You will only deliver the engine if you win the auction.

☒ Low quality

☐ High quality

Submit

Engine value/cost schedule

	Low quality	High quality
Shipping company's value	50	100
Winning supplier's cost	30	40

Consider the following examples for the auction phase:

1. Supplier A bids 160 and selects low quality. Supplier B bids 170 and selects high quality.
 - ➔ Supplier A wins the auction.
 - ➔ The winning bid is 160.

➔ The price range is [40, 160].

➔ The engine is low quality

2. Suppliers A and B both bid 128. Supplier A selects high quality and B selects low quality.

➔ Suppliers A and B each have 50% probability of winning the auction.

➔ The winning bid is 128.

➔ The price range is [32, 128].

➔ If supplier A wins, the engine is high quality

➔ If supplier B wins, the engine is low quality.

Price-setting phase:

- The shipping company is informed about the price range and the engine quality selected by the winning supplier.
- The shipping company is not informed about the quality selected by the losing supplier.
- The shipping company receives his/her engine value.
- The shipping company chooses a final price from the price range to pay the winning supplier.

Computer interface:

In the price-setting phase, the shipping company's price choice is submitted by adjusting a slider. As you adjust the slider, underneath you will see the profits associated with each price.

The shipping company should attempt to make his or her decision in the prescriptive time. For the first five rounds, you have 60 seconds to make your decision. After round five, you have 30 seconds to make your decision. If you cannot make your decision in time, the default price choice is the maximum price in the price range.

Price-setting phase computer screen:

Time left to complete this page: 0:25

Round 1 out of 30

You are a Shipping Company.

Winning bid:	128
Price range:	[32, 128]
Engine:	High quality

Please choose a price from the price range.

Adjust the slider to your preferred price choice then click "Submit". The profits associated with each choice will appear underneath.

72

- Shipping Company Profit = 28
- Winning Supplier Profit = 32

Submit

Engine value/cost schedule

	Low quality	High quality
Shipping company's value	50	100
Winning supplier's cost	30	40

At the conclusion of the price-setting phase, the round ends and profits are realised.

Consider the following example scenarios:

1. The winning bid is 128 and so the price range is [32, 128].

The winning supplier selected a low quality engine.

If the shipping company chooses a final price of 45:

➔ Shipping Company Profit = $50 - 45 = 5$

➔ Winning Supplier Profit = $45 - 30 = 15$

➔ Losing Supplier Profit = 0

2. The winning bid is 128 and so the price range is [32, 128].

The winning supplier selected a high quality engine.

If the shipping company chooses a final price of 65:

➔ Shipping Company Profit = $100 - 65 = 35$

➔ Winning Supplier Profit = $65 - 40 = 25$

➔ Losing Supplier Profit = 0

Information Display

Much of the information displayed to you during the experiment is stored in a table in the lower portion of the computer screen. Here is an example of this table filled with made up numbers for demonstration purposes. The information shown is not from an actual experiment with people.

Round	Winning Bid	Losing Bid	Quality	Final Price	Profit	Balance
3	45	185*			0.00	56.00
2	75*	131	Low	34.00	4.00	56.00
1	128*	175	High	72.00	32.00	42.00

*Your Bid

The table provides detailed information for each round. Suppliers see the winning and losing auction bids and which of these is their own bid. The shipping company and winning supplier also see the winning supplier's engine quality delivered and the final price. All participants see their profit for the round and total accrued points balance.

Comprehension quiz

Please answer the questions below. Raise your hand if you require assistance. Once you answer all the questions correctly, you will be assigned your role and guided through two training rounds. The main part of the experiment then begins. Note: The comprehension questions and training rounds have no influence on your payment.

- Consider the following scenario.

The winning auction bid is 140 and so the price range is [35, 140]. First, suppose the winning supplier selected a low quality engine, which is valued by the shipping company at 50 and costs the winning supplier 30. The shipping company chooses a final price of 44.

1. What is the Shipping Company Profit? 6
2. What is the Winning Supplier Profit? 14
3. What is the Losing Supplier Profit? 0

Now suppose the winning supplier selected a high quality engine, which is valued by the shipping company at 100 and costs the winning supplier 40. The shipping company still chooses a final price of 44.

4. What is the Shipping Company Profit? 56
5. What is the Winning Supplier Profit? 4

---- Arbitrator ----

In the price proposal phase, the shipping company observes the engine quality delivered. Then, the shipping company proposes a final price to pay the winning supplier. The shipping company's proposed price can be any amount from **one quarter** of the winning auction bid up to and including the winning auction bid.

In the arbitration phase, an appeal is triggered if the shipping company's proposed price is below the price that an arbitrator would set. The arbitrator's price depends on its favoured profit ratio, which changes during the experiment. Whenever an appeal is triggered, the arbitrator is available with **50% probability**. If the arbitrator is available, the shipping company must pay the arbitrator's price instead of the price proposal. Precise details will be given below.

At the end of the round, the points earned are calculated as follows:

Shipping company earns $5 + \text{Profit}$, where $\text{Profit} = \text{value} - \text{final price}$

Winning supplier earns $10 + \text{Profit}$, where $\text{Profit} = \text{final price} - \text{cost}$

Losing supplier earns 10 ($\text{Profit} = 0$)

Note: Profit can be negative. Make your decisions carefully.

We now explain the experiment procedure in detail.

Auction phase:

- The two suppliers in a matching group each submit a bid at an auction
- This bid can be any whole number from 30 up to and including 200.
- Neither of the two suppliers can see the other's choice of bid.
- The supplier that submits the **lower** bid wins the auction. In the event of a tie, the computer chooses the winner at random.
- The winning bid is not necessarily the final price paid by the shipping company. Instead, the winning bid determines the **price range** that the winning supplier can receive:
 - ➔ the **minimum price** in this range is **one quarter** of the winning bid.
 - ➔ the **maximum price** in this range is the **winning bid**.
- At the same time as submitting a bid, each supplier also selects an engine quality (high or low) to be delivered in the event of winning the auction.
- The winning supplier at the auction proceeds to the next phase and incurs his/her engine delivery cost.
- The losing supplier at the auction exits and earns the round endowment of 10 points (a profit of zero).

Computer interface:

In the auction phase, each supplier submits his/her bid by adjusting a slider in the upper part of the computer screen. As you adjust the slider, underneath you see the price ranges associated

with the different possible bids. The first number in the brackets is the minimum price, which is one quarter of the selected bid. The second number is the maximum price, which is the bid itself.

A supplier submits his/her quality by selecting one of two options in the lower part of the screen. As you click on each option, underneath you see the arbitrator's price associated with your combination of bid and quality level selected.

On the right-hand side of the screen appears the arbitrator's favoured profit ratio and the engine value and cost schedule.

Suppliers should attempt to make their decisions in the prescriptive time. For the first five rounds, you have 60 seconds to make your decisions. After round five, you have 30 seconds to make your decisions. If you cannot make your decisions in time, the default bid is 200 and the engine quality is selected at random.

Auction phase computer screen:

Time left to complete this page: **0:28**

Round 1 out of 30

You are a Supplier.

1) Please select a bid.

Adjust the slider to your preferred bid between 30 and 200.
The price range for that bid will appear underneath.

128

Price range

[32 , 128]

2) Please select an engine quality.

You will only deliver the engine if you win the auction.

☒ Low quality
☐ High quality

- Arbitrator's price (given bid/quality): **43.33**
- Appeal triggered if proposal receive is below this.

Submit

Arbitrator favours a 2 to 1
supplier to shipping company
profit ratio in this round

Engine value/cost schedule

	Low quality	High quality
Shipping company's value	50	100
Winning supplier's cost	30	40

Consider the following examples for the auction phase:

1. Supplier A bids 160 and selects low quality. Supplier B bids 170 and selects high quality.
 - ➔ Supplier A wins the auction.
 - ➔ The winning bid is 160.
 - ➔ The price range is [40, 160].
 - ➔ The engine is low quality
2. Suppliers A and B both bid 128. Supplier A selects high quality and B selects low quality.
 - ➔ Suppliers A and B each have 50% probability of winning the auction.

- ➔ The winning bid is 128.
- ➔ The price range is [32, 128].
- ➔ If supplier A wins, the engine is high quality
- ➔ If supplier B wins, the engine is low quality

Price proposal phase:

- The shipping company is informed about the price range and the engine quality selected by the winning supplier.
- The shipping company is not informed about the quality selected by the losing supplier.
- The shipping company receives his/her engine value.
- The shipping company proposes a final price from the price range to pay the winning supplier.

Computer interface:

In the price proposal phase, the shipping company's proposed price is submitted by adjusting a slider. As you adjust the slider, underneath you will see the profits associated with each price. You will also see the price that the arbitrator would set for each proposal and whether an appeal to the arbitrator would be triggered.

The shipping company should attempt to make his or her decision in the prescriptive time. For the first five rounds, you have 60 seconds to make your decision. After round five, you have 30 seconds to make your decision. If you cannot make your decision in time, the default price proposal is the maximum price in the price range.

Price proposal phase computer screen:

Time left to complete this page: **0:28**

Round 1 out of 30

You are a Shipping Company.

Winning bid:	128
Price range:	[32, 128]
Engine:	High quality

Please propose a price from the price range.

Adjust the slider to your preferred price proposal then click "Submit". The profits and the arbitrator's price associated with each proposal will appear underneath.

72

- Shipping Company Profit = 28
- Winning Supplier Profit = 32
- Arbitrator's price: **80.00**
- Proposal would trigger an appeal? **Yes**

Submit

Arbitrator favours a 2 to 1
supplier to shipping company
profit ratio in this round

Engine value/cost schedule

	Low quality	High quality
Shipping company's value	50	100
Winning supplier's cost	30	40

Arbitration phase:

- The outcome of the arbitration phase depends on the decisions taken in the first two phases.
- An arbitrator observes the winning bid, the shipping company's proposed price and the proposed shipping company and winning supplier profits.
- The arbitrator then sets its own price. The arbitrator cannot set any price, however. The arbitrator's price can be from between the shipping company's proposed price and the winning bid.
- The arbitrator's price is set to bring profits as close as possible to a favoured profit ratio, which varies between the first half and the second half of the experiment:
 - ➔ **EITHER** the arbitrator favours a 2 to 1 **shipping company** to supplier profit ratio, i.e., the arbitrator favours an outcome in which the shipping company's profit is double the winning supplier's profit.
 - ➔ **OR** the arbitrator favours a 2 to 1 **supplier** to shipping company profit ratio, i.e., the arbitrator favours an outcome in which the winning supplier's profit is double the shipping company's profit.
- You will be informed about the arbitrator's favored profit ratio immediately before the first round, and again when this ratio changes at the beginning of the sixteenth round.
- If the **shipping company's proposed price is below the arbitrator's price**, then an appeal to the arbitrator is automatically triggered. If no appeal is triggered, then the final price remains the shipping company's proposed price.
- Following an appeal, the computer rolls a standard six-sided die to determine whether the arbitrator is available:
 - ➔ If the die comes up **1, 2 or 3**, the arbitrator is **unavailable**, and the final price paid by the shipping company remains the shipping company's proposed price.
 - ➔ If the die comes up **4, 5 or 6**, the arbitrator is **available**, and the final price paid by the shipping company is the arbitrator's price. In this situation, a 2 point arbitrator fee is levied on the shipping company.

At the conclusion of the arbitration phase, the round ends and profits are realised.

Consider the following example scenarios:

1. The winning bid is 128 and so the price range is [32, 128].

The winning supplier selected a low quality engine.

The shipping company proposes a price of 39.

If the arbitrator favours a 2 to 1 **shipping company** to supplier profit ratio, then the arbitrator's price is also 39, because the shipping company's proposed price already gives the winning supplier more than half the shipping company's profit. No appeal is triggered.

- ➔ Shipping Company Profit = $50 - 39 = 11$
- ➔ Winning Supplier Profit = $39 - 30 = 9$
- ➔ Losing Supplier Profit = 0

2. The winning bid is 128 and so the price range is [32, 128].

The winning supplier selected a high quality engine.

The shipping company proposes a price of 65.

If the arbitrator favours a 2 to 1 **supplier** to shipping company profit ratio, then the arbitrator's price is 80, at which price the winning supplier would earn double the shipping company's profit. The shipping company's proposed price is below the arbitrator's price and so an appeal is triggered.

Suppose that the arbitrator is unavailable and so the final price remains 65:

- ➔ Shipping Company Profit = $100 - 65 = 35$
- ➔ Winning Supplier Profit = $65 - 40 = 25$
- ➔ Losing Supplier Profit = 0

Suppose instead that the arbitrator is available and so the final price is 80. An arbitrator fee of 2 points is subtracted from the shipping company's profit.

- ➔ Shipping Company Profit = $100 - 80 - 2 = 18$
- ➔ Winning Supplier Profit = $80 - 40 = 40$
- ➔ Losing Supplier Profit = 0

Information Display

Much of the information displayed to you during the experiment is stored in a table in the lower portion of the computer screen. Here is an example of this table filled with made up numbers for demonstration purposes. The information shown is not from an actual experiment with people.

Round	Bids Winning / Losing	Quality	Proposal	Arbitrator Appeal / Available	Final Price	Profit	Balance
3	45 / 131*					0.00	75.00
2	75* / 131	Low	45.00	No / -	45.00	15.00	75.00
1	128* / 175	High	72.00	Yes / Yes	80.00	40.00	50.00

*Your Bid

The table provides detailed information for each round. Suppliers see the winning and losing auction bids and which of these is their own bid. The shipping company and winning supplier also see the winning supplier's engine quality, the shipping company's proposed price, whether an appeal is triggered and the arbitrator's availability, and the final price. All participants see their profit for the round and total accrued points balance.

Comprehension quiz

Please answer the questions below. Raise your hand if you require assistance. Once you answer all the questions correctly, you will be assigned your role and guided through two training rounds. The main part of the experiment then begins. Note: The comprehension questions and training rounds have no influence on your payment.

- Consider the following scenario.

The winning auction bid is 140 and so the price range is [35, 140]. First suppose the winning supplier selected a low quality engine, which is valued by the shipping company at 50 and costs the winning supplier 30. The shipping company proposes a price of 44.

Let the arbitrator favour a 2 to 1 supplier to shipping company profit ratio. The arbitrator's preferred price is no higher than the shipping company's proposal (and remember it cannot be any lower), because a price of 44 already gives the winning supplier more than double the shipping company's profit. No appeal is triggered. The final price remains 44.

1. What is the Shipping Company Profit? *6*
2. What is the Winning Supplier Profit? *14*
3. What is the Losing Supplier Profit? *0*

Now suppose the winning supplier selected a high quality engine, which is valued by the shipping company at 100 and costs the winning supplier 40. The shipping company still proposes a price of 44.

Again, let the arbitrator favour a 2 to 1 supplier to shipping company profit ratio. The arbitrator's preferred price is now 80, which is higher than the shipping company's proposal and so an appeal is triggered.

The die is rolled and the arbitrator turns out to be available. The final price is now 80. An arbitrator fee of 2 points should be subtracted from the shipping company's profit.

4. What is the Shipping Company Profit? *18*
5. What is the Winning Supplier Profit? *40*

Appendix F: Post-experiment questionnaire and subject pool characteristics.

Age: Interval variable.

Years.

Mean 22.71, Median 22, Standard deviation 4.03, Minimum 18, Maximum 39

Gender: Categorical variable.

Male 50.93%; Female 48.15%; Other 0.00%; Prefer not to Say 0.93%.

Field of studies: Categorical variable.

Arts and Education 16.67%; Economics and Finance 11.11%; Business and Management 13.89%; Social Sciences and Law 10.19%; Medicine and Health Sciences 14.81%; Engineering and Natural Sciences 33.33%; Not a Student 0.00%.

Nationality: Categorical variable:

Central and Eastern Asia 2.78%; Central and Western Africa 4.63%; Central, South America and the Caribbean 1.85%; Europe (excl. UK) 29.63%; Middle East and North Africa 2.78%; North America 1.85%; Oceania 0.00%; South and Eastern Africa 0.93%; South-East Asia 6.48%; Southern Asia 3.70%; UK 45.37%.

Income: Categorical variable.

When you were 16 years of age, what was the income of your parents in comparison to other families in your country?

Far below average 4.63%; Below average 15.74%; Average 37.04%; Above average 36.11%; Far above average 6.48%

Risk Indices:

Based on Dohmen et al. (2011). Likert scale from 0 “Completely unwilling to take risks” to 10 “Completely willing to take risks”.

1) Are you generally a person who is fully willing to take risks or do you try to avoid taking risks?

Mean 6.30, Median 7, Standard deviation 2.35, Minimum 0, Maximum 10

2) How would you rate your willingness to take risks in financial matters?

Mean 4.70, Median 5, Standard deviation 2.53, Minimum 0, Maximum 10

Trust Index: Average of three variables.

Questions taken from the "General Social Survey" consistent with the approach of Glaeser et al. (2000).

- 1) Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?

"Most people can be trusted" 29.63% or "Can't be too careful" 70.37%

- 2) Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?

"Would try to be fair" 31.48% or "Would take advantage of you" 68.52%

- 3) Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?

"Try to be helpful" 38.89% or "Just look out for themselves" 61.11%

Trust Strangers: Dummy variable.

You can't count on strangers anymore.

"More or less disagree" 38.89% or "More or less agree" 61.11%

Additional references

- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3), 522-550.
- Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring Trust. *Quarterly Journal of Economics*, 115(3), 811-846.
- Nelson, F. D. (1976). On a General Computer Algorithm for the Analysis of Models with Limited Dependent Variables. *Annals of Economic and Social Measurement*, 5(4), 493-509.
- Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., & Webb, M. D. (2019). Fast and wild: Bootstrap inference in Stata using boottest. *The Stata Journal*, 19(1), 4-60.