Università degli Studi di Padova dSEA

Riccardo Camboni
DSEA, University of Padova

Luca Corazzini
Department of Economics, University of Venice "Ca' Foscari"

Stefano Galavotti
DEMDI, University of Bari

Paola Valbonesi
DSEA, University of Padova and HSE-NRU, Moscow

BIDDING ON PRICE AND QUALITY: AN EXPERIMENT ON THE COMPLEXITY OF SCORING AUCTIONS

December 2019

Marco Fanno Working Papers - 243

Bidding on price and quality: An experiment on the complexity of scoring auctions

Riccardo Camboni[†], Luca Corazzini[‡], Stefano Galavotti[§], Paola Valbonesi[¶]

December 11, 2019

Abstract

We run an experiment on procurement auctions in a setting where both quality and price matter. We compare two unidimensional treatments in which the buyer fixes one dimension (quality or price) and sellers compete on the other, with three bidimensional treatments (with different strategy spaces) in which sellers submit a price-quality bid and the winner is determined by a *score* that linearly combines the two offers. We find that, with respect to the theoretical predictions, the bidimensional treatments significantly underperform, both in terms of efficiency and buyer's utility. We attribute this result to the higher strategic complexity of these treatments and test this intuition by fitting a structural Quantal Response Equilibrium model with risk aversion to our experimental data. We find very similar estimates for the risk aversion parameter across all treatments; instead, the error parameter, which captures deviations between the observed bids and the payoff-maximizing ones, is larger in the bidimensional treatments than in the unidimensional ones. Our evidence suggests that increasing the dimensionality and the size of the suppliers' strategy space increases their tendency to make suboptimal offers, thus undermining the theoretical superiority of more complex mechanisms.

JEL Code: D44; H11; H57.
Keywords: scoring auctions; multidimensional auctions; complexity; bidding behaviour; Quantal Response Equilibrium.

[*]We thank Alessandro Bucciol, Uri Gneezy, Marco Guerzoni, Antonio Nicolò, Salvatore Nunnari, Fausto Pacicco, Giancarlo Spagnolo, Steve Tadelis, Orestis Troumpounis, Luigi Vena, Andrea Venegoni, Claudio Zoli, participants at the AMEC in St. Petersburg, the European ESA Meeting in Dijion, the ESA World Meeting in Vancouver, the SIOE Conference in Stockholm, and seminars in Padova and Verona for useful discussions and comments. We are indebted to Gregory Lewis and Patrick Bajari for sharing their scoring auctions' real data. We are especially grateful to Joachim Vosgerau for the opportunity to run the experiment at the Bocconi Experimental Laboratory for the Social Sciences (BELSS, Bocconi University, Milan). Iuliana Iuras and Marco Magnani provided excellent research assistance during our experimental sessions in BELSS. Financial support from the Department of Economics and Management, University of Padova, PRAT, CPDA158913/2015; from Centro Studi di Economia e Tecnica dell'Energia "Giorgio Levi Cases", University of Padova; and from Cariparo, Progetto di Eccellenza 2017, is gratefully acknowledged. No funding sources were involved in the study design, in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit the article for publication. All errors are ours.

[†]Corresponding Author - Department of Economics and Management, University of Padova, Italy. Address: Via del Santo 33, 35123, Padova, Italy. email: riccardo.camboni@unipd.it.

[‡]Department of Economic Sciences, University of Venice "Ca' Foscari", Italy.

[§]Department of Economics, Management and Corporate Law, University of Bari, Italy.

[¶]Department of Economics and Management, University of Padova, Italy; and National Research University - Higher School of Economics, Russian Federation, NRU-HSE.

1 Introduction

In procurement markets, suppliers compete for the right to sell goods or to provide services to a buyer. Usually, the object of the transaction is an item that is yet to be produced or a service that will be delivered in the future; sometimes, it has to be totally or partially designed ad hoc and then realized from scratch. In many of these cases, a number of valuable attributes of the procurement contract – technical characteristics, delivery lead time, payment conditions – can be negotiated ex-ante. Hence, the procurement problem for the buyer is not only to select a supplier but also to choose what to procure with the goal of obtaining the best compromise between the object's value and the financial disbursement.

The design of the tender procedure is central to achieving this goal. Two auction mechanisms are usually adopted in practice. In the simplest mechanism, the buyer defines the minimal technical requirements in the call for tender and then lets suppliers bid on price only, awarding the contract to the lowest-price seller: this corresponds to a standard first-price auction, where bidders commit to procure a good with certain pre-specified characteristics. Alternatively, the buyer may adopt a scoring (or multi-attribute) auction in which participants submit a multidimensional bid comprising a price and a number of non-price attributes; these elements are then mapped, usually in a linear combination fashion, into a *score*, and the supplier that earns the highest score is awarded the contract.

Scoring auctions are increasingly used in Europe and the United States. In Europe, scoring auctions are commonly referred to as the "most economically advantageous tender" ("MEAT"). The European Union Directive 2014/24/EU supports moving away from contracts awarded through first price auctions to tenders based on MEAT. According to the Tender European Daily (TED) data, in 2016, 72% of all auctions above the value of 150,000 euros used MEAT as the awarding criterion. In the United States, scoring auctions have been adopted to award highway maintenance and transportation construction. Known as "cost-plus-time" or "A+B" bidding, the mechanism works as a bidimensional scoring auction with time to completion as the non-price dimension (e.g., Lewis and Bajary, 2011; Gupta et al., 2015).

This gradual shift from first-price to scoring auctions seems to have been informed by the economic theory: Asker and Cantillon (2008) show that, when a buyer is concerned with both price and non-price elements of the offer, a scoring auction is superior to a first-price auction with fixed attribute levels. The intuition is straightforward: When suppliers are heterogeneous, a scoring auction promotes competition, in that it allows suppliers to submit offers that best match their productive skills, favoring the attributes on which they have a competitive advantage. Given the relevance of the procurement market, it is then important to test whether and to what extent this theoretical prediction survives experimentally. Our conjecture is that a scoring auction is arguably a more complex strategic environments than a first-price auction in that bidders have to reason multidimensionally, and this greater complexity may prompt behavioral responses that are worth investigating. In addition, given the increasing demand of innovative solutions by procurers, it is also worth exploring alternative mechanisms beyond the widely used first-price and scoring auctions. A thorough understanding of the determinants of the behavior by bidders in these contexts may help the market designer find the awarding procedure that better matches the procurer's needs.

To address these questions, we design an experiment in which a buyer wants to procure an object for which both the price and a non-price dimension, to which we refer as *quality*, matter. Potential sellers have private information on a cost parameter that shifts their (convex) cost functions for quality provision. We implement five treatments in the lab: Two treatments resemble those mainly used in procurement markets. In the scoring rule auction (*SRA*), or simply scoring auction, sellers submit a bidimensional offer comprising a price and a quality bid, which are then linearly combined according to a publicly announced *scoring rule*, and the seller whose score is the highest is awarded the contract. In the first-price auction (*FPA*), the level of quality is imposed by the buyer, sellers compete on price only, and the seller who submits the lowest price wins the auction.

To complete the picture, we also consider another unidimensional treatment, the first-quality auction (*FQA*), in which the buyer announces the price she will pay for the contract, sellers compete on quality only, and the seller who submits the highest quality bid wins the auction. Notice that this awarding mechanism, though apparently uncommon, is not a mere theoretical construct: the European Union Directive on public procurement envisages that "The cost element may also take the form of a fixed price or cost on the basis of which economic operators will compete on quality criteria only" (Directive 2014/24/EU, art. 67, second paragraph).[1] Moreover, public calls for research grants often takes the form of a competition on quality only: for example, the European Research Council states that "Proposals are evaluated by selected international peer reviewers who assess them on the basis of excellence as the sole criterion".

Finally, we also implement two treatments that lie halfway between the *SRA* on one hand and the *FPA* and *FQA* on the other: In these treatments, called *SRA2p* and *SRA2q*, sellers bid on both price and quality, but one of the two (price in *SRA2p*, quality in *SRA2q*) is constrained to a binary choice. To allow for a sensible comparison of the results across treatments, we chose the exogenous parameters to maximize the buyer's expected utility, assuming risk neutrality and equilibrium behavior by sellers.

Our experimental results show a trade-off between optimality and complexity: While the *SRA* is theoretically superior to all other treatments in terms of buyer's surplus and efficiency, this is no longer true in the lab. In particular, the *FQA* performs as well as the *SRA* (which does not perform better than the two bidimensional treatments with binary choice on one dimension), whereas the *FPA* has the worst performance, as predicted by the theory.

To shed light on these findings, we then turn to the analysis of bids. In all treatments (except in *FPA*, where quality is fixed), we detect a clear tendency by sellers to submit a higher level of quality than the one theoretically predicted, a tendency that goes in the direction of improving efficiency and the buyer's surplus. This overbidding on the quality dimension accounts for the overperformance of *FQA* and suggests a potential risk aversion justification. However, in the bidimensional treatments, such overbidding is accompanied by two countervailing effects: first, a significant frequency of contracts are inefficiently allocated to the highest-cost supplier; second, a higher-than-predicted level of submitted quality tends to be accompanied by an even stronger upward adjustment of prices, which eventually reduces the buyer's surplus. Our conjecture is that these two effects, both of which negatively affect the performance of the bidimensional treatments, may be related to the suppliers' response to more complex environments.

To corroborate this intuition, we fit to our data a structural Quantal Response Equilibrium (QRE) model with two parameters: an error parameter that measures the degree at

[1] We are aware of a few examples in Italy in which this awarding rule has been used in the procurement of care services such as accommodation for asylum seekers and psycho-pedagogical activities for kids in primary schools.

which suppliers play suboptimal strategies and a risk aversion parameter (assuming Constant Relative Risk Aversion utilities for sellers). Across treatments, we obtain remarkably similar estimates for the risk aversion parameter. Estimates for the error parameter are consistent with our intuition: As we move from unidimensional to bidimensional treatments, we observe increasing deviations of actual bids from the payoff-maximizing ones. Moreover, in the bidimensional treatments, errors are less significant in *SRA2p* and *SRA2q*, where one of the two dimensions is simplified to a binary choice. However, while the QRE model fits the data quite well in four of the five treatments, the fit in *SRA* is less satisfactory. In addition, our two-parameter QRE model cannot account for the overbidding in quality that we observe in this treatment. We then re-estimate an augmented QRE model in which sellers may have a distorted perception of the cost-benefit trade-off that is associated with a marginal change in their quality bid. We show that allowing for this mis-perception significantly improves the model's fit for *SRA*, while the other treatments are unaffected.

The rest of the paper is organized as follows: Section 2 discusses the literature to which we contribute, while Section 3 describes our experimental design. Theory and testable predictions are presented in Section 4, and the experimental results are shown in Section 5. Section 6 introduces and estimates a structural Quantal Response Equilibrium model to organize our data. Section 7 concludes, elaborating on the consistency of our results with the empirical evidence on real-world scoring auctions, and drawing some policy implications.

2 Related literature

The theoretical properties of scoring auctions were first derived by Che (1993) in a framework in which only one non-price attribute (quality) is relevant, while the price enters linearly in the scoring function. Under these conditions, the optimal scoring rule will under-reward quality relative to the buyer's true preferences. Asker and Cantillon (2008, 2010) generalize the analysis to a situation in which sellers' types are multidimensional, and several non-price attributes matter to the buyer. They also compare the scoring auction with other common procedures used to award multi-attribute contracts. In particular, they show that, in terms of the buyer's surplus, the scoring auction strictly dominates a price-only auction with minimum quality standards.

Unfortunately, testing these theoretical predictions in the field is difficult given the heterogeneity of contracts in a typical dataset. Still, few empirical studies compare the performance of scoring auctions with respect to alternative awarding procedures. Cameron (2000) compares the scoring mechanism with a more flexible approach in which the public buyer reveals the bid-evaluation criteria in general terms, and the bidding process is used to shortlist a set of suppliers to bargain with. Using a dataset of ninety-three long-term electricity purchase contracts awarded in five US states, Cameron finds that the scoring auction obtains an 18% price reduction with respect to the flexible mechanism, but a 50% increase in the probability of breaking the contract. Hyytinen et al. (2018) investigate the change in the awarding of cleaning services in Sweden, which went from a discretionary beauty contest to explicit award rules, notably first-price and scoring auctions. They show that this change of regime did not produce the expected reductions in procurement costs and propose two concurring explanations: that the new rules increased entry costs, and that, in the discretionary regime, an implicit favoritism for in-house units led the other suppliers to strategically bid more aggressively. Interestingly, their data record no significant difference between first-price and scoring auctions in the price paid by the public buyer: In fact, they observe that the cleaning service is an almost homogeneous product for which there is little scope for moving away from a policy of granting the contract to the lowest-cost supplier.

Lewis and Bajary (2011) compare the scoring and first-price auctions used by the California Department of Transportation to award more than 1,300 highway construction projects between 2003 and 2008. The quality component that enters into the scoring rule is the number of days to complete the project. In a framework that is significantly similar to our experimental design, they find that projects awarded using the scoring auction are a little more expensive than those awarded with a price-only auction and are completed much sooner. Using a dollar-value estimate of the negative externality to commuters caused by each day of work, they conclude that the users' welfare gain from using a scoring auction instead of a first-price largely outweighs the increase in the procurement cost. To fully assess the consequences in terms of social welfare, they then structurally estimate the contractors' cost, assuming optimal behavior at the bidding stage, and conclude that scoring auctions generate a significantly larger social welfare than first-price auctions, so they should always be adopted. Moreover, even a policy of small incentives (i.e., a small weight to the quality component in the scoring function) meant to reduce the procurement extra-cost would be welfare improving.

The empirical studies reviewed above are rich in insightful results. However, by clearing the analysis of confounding aspects (e.g., contract heterogeneity) and by controlling for suppliers' costs, an experimental approach like the one we suggest would improve our understanding of the effectiveness of scoring auctions as compared to other awarding procedures. However, the experimental literature on scoring rule and multi-attribute auctions is scant. Moreover, most papers focus almost exclusively on the performance of the various awarding mechanisms, and, unlike our paper, do not deeply analyze the suppliers' side.

Chen-Ritzo et al. (2005) run an experiment involving an English reverse auction in which sellers submit three-dimensional bids (price, quality and lead time), and the buyer does not fully disclose how bids are mapped into the score but only provides feedback to suppliers in terms of a marginal score (i.e., how a one-unit change in quality/lead time from the current bid would change the score). They find that the three-attribute auction is effective in increasing both the buyer's and the sellers' surplus, although differences are less pronounced than predicted (in fact, in three cases out of eight, they do not detect significant differences). Strecker (2010) studies the effect of revealing information in an English auction with three attributes and finds that efficiency is greater when the scoring rule is fully disclosed than when only limited information is provided to sellers; however, the buyer's surplus is not significantly affected by the information-revelation policy. Bichler (2000) employs an experimental setting that mimics the financial market to assess the performance of three multi-attribute mechanisms – a first-score sealed bid like our *SRA*, a second-score sealed bid, and a first-score open-cry auction – with respect to a single-attribute mechanism. In his setting, the buyer solicits an offer for a call option on a certain index or share traded in the Vienna Stock Exchange, and the quality element is represented by the volatility of the underlying index or share. Bichler finds that the buyer achieves higher utility in the multi-attribute mechanism than in the single-attribute mechanisms, whereas the level of efficiency is similar. In the multi-attribute auctions, the first-score sealed bid auction performs better than the second-score and the open-cry auctions. Albano et al. (2018) study how changing the weights attached to the price and quality component in the scoring rule affects suppliers' behavior. Unlike our experiment, in their setting, the quality is exogenously and randomly assigned to each supplier prior to competing, so their mechanism reduces to a unidimensional auction with a non-neutral awarding rule. They find that the scoring rule that gives a larger weight to quality is far more efficient.

Our paper also contributes to the literature on how individuals act in complex strategic environments. One context in which the issue of complexity has been repeatedly raised and analyzed is that of multi-unit combinatorial auctions (see, e.g., the survey by (Kwasnica and Sherstyuk, 2013)). In that context, complexity takes two forms: complexity in the winner's determination problem (i.e., the computational burden of finding the revenue-maximizing allocation for given bids), and complexity in the supplier's choice (i.e., the cognitive difficulty of selecting a good offer). With reference to the latter problem, Kagel et al. (2010) show that, in combinatorial clock auctions, suppliers tend to myopically bid on a small number of packages, which may negatively affect efficiency.[2] Scheffel et al. (2012) reach similar conclusions, finding that suppliers use simple heuristics to select packages and arguing that this approach has to do with cognitive limits in terms of the number of items on which people can simultaneously concentrate. Our paper shows that, even in the apparently simpler context of a single-unit auction, a high degree of complexity in the form of a multiple number of dimensions on which suppliers are called to think and bid, may affect the social welfare.

3 Experimental design

3.1 Baseline game and treatments

The baseline game considered in our experiment consists of a procurement scoring auction with incomplete information (henceforth, denoted by *SRA*). Two sellers participate in an auction to sell an object to a buyer. The sellers simultaneously place their bids, consisting of two integer numbers: the quality of the object, denoted by q, and the price at which the seller is willing to sell it, denoted by p. The submitted quality is constrained to be a number between 0 and 70; the set of admissible prices varies with the submitted quality: In particular, the buyer is willing to pay no more than $p^{\max}(q) = q + 50$ for an object of quality q.

Each seller's bid (q, p) is then mapped into a score s that linearly combines quality and price according to the following *scoring rule*:

$$s(q, p) = 50 + 2q - p. \tag{1}$$

Observe that the scoring rule (1) rewards quality and penalizes price. The coefficients attached to q and p in (1) are set optimally (in a sense that will be explained in the next section). The constant term is added to avoid negative scores.[3]

The seller whose score is the highest wins the auction, and ties are broken randomly. The winning seller is paid the submitted price p but has to bear the cost of providing the submitted quality. Specifically, the winner's monetary payoff is:

$$m(q, p) = p - C(q), \tag{2}$$

where

$$C(q; \theta) = \frac{q^2}{4\theta}. \tag{3}$$

[2]Similarly, Kwasnica et al. (2005) refer to the "*computational complexity of the bidders' problem*" as a potential cause of reduction in an auction's efficiency.

[3]In fact, since only $q \geq 0$ and $p \leq p^{\max}(q) = q + 50$ are admissible, it is $s(q, p) \geq 50 + 2q - p^{\max}(q) = q \geq 0$.

On the other hand, the loser of the auction earns nothing.

Notice that $C(q;\theta)$, the cost of providing an object of quality q, is increasing and convex in q and depends on a parameter, θ, that is idiosyncratic to each seller and that identifies the seller's "type". Notice also that the cost is strictly decreasing in the type θ, which can then be interpreted as an indicator of the seller's productive efficiency. At the beginning of the auction, the types are independently drawn from a discrete uniform distribution whose support is given by all the integers from 1 to 10. Each seller observes the realization of her own type but not that of her opponent. Everything else is common knowledge.

Along with the baseline game *SRA* just described, we implement four additional treatments in which the size and the dimensionality of sellers' strategy sets are gradually reduced.

In two treatments, *FPA* (which stands for first-price auction) and *FQA* (which stands for first-quality auction), sellers bid on one dimension only – price in *FPA* and quality in *FQA* – while the other dimension is set exogenously by the experimenter. Specifically, in *FPA*, sellers are constrained to deliver quality $\bar{q} = 16$ (and to bear the associated cost defined by (3) if they win) and simply submit a price bid. The awarding rule is the same as in *SRA*, but since quality is fixed, the lowest-price seller wins. In *FQA*, the buyer commits to pay the price $\bar{p} = 32$ to the winner, and sellers compete on quality only. Since the price is fixed, the seller who offers the highest quality wins the auction (and bears the cost associated with the submitted quality, as defined by (3)).

In the remaining two treatments, named *SRA2q* (which stands for scoring rule auction with two qualities) and *SRA2p* (which stands for scoring rule auction with two prices), sellers' bids are two-dimensional, like in *SRA*, but one dimension – quality in the former treatment and price in the latter – is constrained to a dichotomous choice. Specifically, in *SRA2q* sellers can submit one of two possible quality levels, either $q_L = 9$ or $q_H = 40$, whereas the price bid can be any (integer) value between 0 and $p^{\max}(q)$. In *SRA2p*, the only admissible prices are $p_L = 12$ and $p_H = 65$, whereas any (integer) quality no greater than 70, for $p_L = 12$, and included between 15 and 70, for $p_H = 65$, can be submitted.[4] As in *SRA*, the winner of the auction is the seller whose score, as defined by (1), is the highest.

The parameters $\bar{q} = 16$ for *FPA*, $\bar{p} = 32$ for *FQA*, $q_L = 9$ and $q_H = 40$ for *SRA2q*, and $p_L = 12$ and $p_H = 65$ for *SRA2p*, have been chosen optimally in a sense that is explained in the next section.

In the rest of the paper, we will often use the phrase "two-dimensional auctions" (or simply "scoring auctions") to encompass *SRA*, *SRA2q* and *SRA2p*, and "one-dimensional auctions" for *FPA* and *FQA*.

3.2 Procedures

Upon their arrival, subjects were randomly assigned to a computer terminal. Instructions[5] were distributed at the beginning of the experiment and read aloud. Before the experiment started, subjects answered a number of control questions to ensure they understood the instructions and the consequences of their choices. When necessary, answers to these questions were checked and explained. In each session, subjects participated in fifteen consecutive repetitions (or periods) of the game. At the beginning of the experiment, the computer randomly formed four rematching groups of six subjects each. The composition of the rematching groups was kept constant throughout the session. In every period, subjects were randomly

[4] $q = 15$ is the minimum level of quality which satisfies $p^{\max}(q) = q + 50$ for $p_H = 65$.

[5] Instructions were originally written in Italian. The English translation is reported in the Appendix.

and anonymously divided into pairs. Pairs were randomly formed in every period within rematching groups. Subjects were told that pairs were randomly formed in such a way that they would never interact with the same opponent in two consecutive periods.[6]

To facilitate decision-making, before submitting their final choice(s), subjects could use a "user-friendly" interface, to simulate as many times as they wished the consequences of their provisional choices on the experimental dimensions (the score associated with that quality/price bid, the cost they would have borne in case of winning, and their earnings). At the end of every period, the outcome was summarized on the screen, along with information about subjects' overall earnings in the period.

We ran three sessions for each of the five treatments, each involving twenty-four subjects, thus generating twelve independent observations at the rematching group level. The experiment took place at the Bocconi Experimental Laboratory for Social Sciences (BELSS) of Bocconi University, Milan, between December 2017 and January 2018. Most participants were undergraduate students who were recruited by means of the SONA recruitment system (http://www.sona-systems.com/default.aspx) from a pool of around 3000 registered users. The experiment was computerized using the z-Tree software (Fischbacher, 2007). Prices, costs and earnings during the experiment were expressed in tokens. At the end of the experiment, the number of tokens a subject had obtained during the experiment was converted at an exchange rate of 1 euro per seven tokens, and monetary earnings were paid in cash privately. Subjects started the experiment with a balance of twenty tokens to cover the possibility of losses. On average, subjects earned 14.47 euro for sessions that lasted seventy minutes, including the time for instructions and payments. Before leaving the laboratory, subjects completed a short questionnaire containing questions on their socio-demographics and their perceptions of the experimental task.

4 Theory and predictions

Our experimental results will be compared to the predictions derived from a benchmark model of risk neutral suppliers and equilibrium behavior.[7] Specifically, we consider a model in which:

- each seller's utility function coincides with her monetary payoff, which is equal to (2) in case of winning, to zero otherwise;
- sellers play the (symmetric) Bayes-Nash equilibrium of the auction.

Since we are also interested in the performance of the various treatments in terms of welfare, we set the following utility function for the buyer:

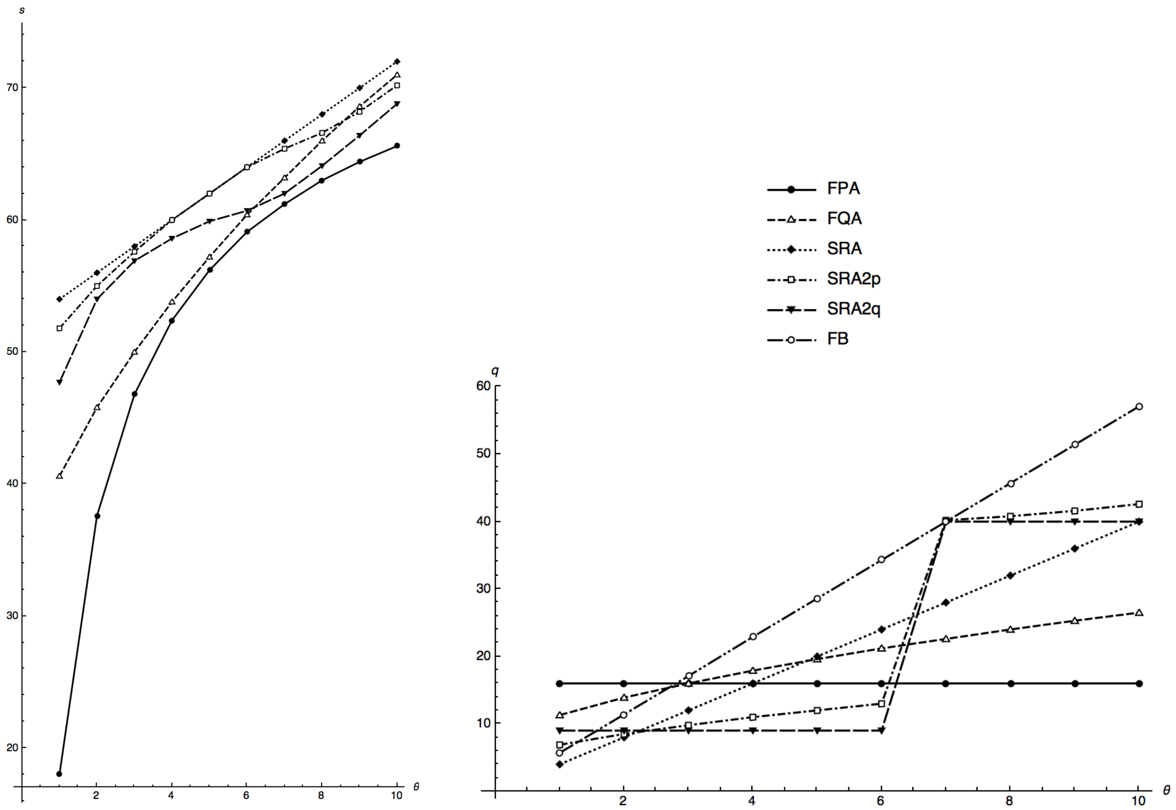
$$u_B(q,p) = \frac{20}{7}q - p. \tag{4}$$

[6] Our rematching protocol implies that, given the size of the sub-groups (six subjects), subjects interacted with the same opponent an average of once every five periods. Although this approach is not perfect stranger protocol, it leaves little room for developing punishment-reward strategies over multiple periods. The rematching protocol was intended to increase the number of independent observations, and non-parametric tests were performed to check the robustness of the main parametric results.

[7] All the results related to this section are derived in the Appendix. To obtain these results, we assumed that types and bids (price and/or quantity) are continuous variables (whereas in the experiment only integers were allowed): Hence, we assumed that θ is drawn from a continuous uniform distribution with support $[1, 10]$, and that p and q can be any positive number (up to the maximum admissible). This approach allowed us to use calculus and to avoid the complications associated with ties.

The exogenous parameters of the five treatments outlined in the previous section are derived from this benchmark model by applying an optimality criterion. Specifically, the weights attached to quality and price in (1) are those that maximize the ex-ante expected utility of a buyer with objective function (4), conditional on the sellers' playing their equilibrium bidding strategies in a scoring auction with linear scoring rule.[8] It is important to stress that the buyer's utility (4) differs from the optimal scoring rule (1): In particular, relative to the utility of the buyer, the optimal scoring rule under-rewards quality, a result that is consistent with what already shown by Che (1993). Likewise, the two admissible values for quality in *SRA2q* (price in *SRA2p*) are those that maximize the buyer's ex-ante expected utility, conditional on sellers' bidding their equilibrium strategies in an auction game like *SRA2q* (*SRA2p*) that uses (1) as an awarding rule. Finally, the exogenous value of quality in *FPA* (price in *FQA*) is set to maximize the buyer's ex-ante expected utility, assuming that sellers' bid according to equilibrium in an auction game like *FPA* (*FQA*).

Figure (1a) displays the equilibrium scores as a function of θ in the five treatments. Observe that, in all treatments, the equilibrium score is strictly increasing in the seller's type:



(a) Score **(b)** Quality

Figure 1 – Equilibrium score and quality as a function of θ

hence, theoretically, the auction should always be won by the seller with the highest type. Notice also that the equilibrium score is uniformly (i.e., for all types) highest in *SRA* and lowest in *FPA*. The remaining three treatments–*FQA*, *SRA2q*, and *SRA2p*–lie in between,

[8]We confined ourselves to linear scoring rules for the sake of simplicity and to mimic what usually happens in real-world multidimensional auctions.

but the ranking among them is ambiguous: For relatively low types, the equilibrium score of *FQA* is well below that of *SRA2q* and *SRA2p*, but the first is steeper and eventually overtakes the latter two, almost reaching *SRA*. Overall, the equilibrium scores in the five treatments become more concentrated as θ increases.

Figure (1b) looks at the equilibrium quality bid (remind that quality is fixed in *FPA*). In those treatments where it can be set freely – *FQA*, *SRA2p* and *SRA* – the submitted quality is strictly increasing in type, but it increases more quickly in *SRA* than in *FQA*. In *SRA2p*, the submitted quality is rather flat for $\theta \leq 6$ and $\theta \geq 7$, but it jumps between $\theta = 6$ and $\theta = 7$. This pattern closely tracks what happens in *SRA2q* (where only $q_L = 9$ and $q_H = 40$ are admissible).

Moving to the analysis of welfare, we assume that utility is transferable, and thus we measure social welfare simply as the sum of the buyer's utility (4) and the winning seller's monetary payoff (2). Therefore, social welfare is given by:

$$W(q^w, \theta^w) = \frac{20}{7} q^w - C(q^w; \theta^w), \tag{5}$$

where q^w is the quality submitted by the winner of the auction and θ^w is her type. As equation (5) suggests, there are two dimensions that jointly affect efficiency:

1. Cost Efficiency: whatever level of quality is delivered, the object should be produced at the lowest possible cost;

2. Quality Efficiency: whatever seller produces it, the quality of the object should be such that the marginal benefit of quality (to the buyer) is equal to the seller's marginal cost of providing that level of quality. In particular, since the marginal benefit of quality is constant and equal to 20/7, and the marginal cost is $q/(2\theta)$, the efficient level of quality when the object is delivered by a type-θ seller is $q^{\text{EFF}}(\theta) = (40/7)\theta$.[9]

It is immediate to see that, in equilibrium, all treatments are cost-efficient: In fact, since scores are strictly increasing in all treatments, the object is always assigned to the low-cost seller. On the other hand, as Figure (1b) shows, in equilibrium no treatment is efficient in terms of quality: therefore, any efficiency loss is attributable to sellers' offering an inefficient level of quality. In particular, with some exceptions for $\theta < 3$, the submitted quality falls short of its efficient level. To establish the ranking across treatments in terms of welfare, we computed the expected social welfare generated in equilibrium in each treatment. The most efficient treatment is *SRA*, closely followed by *SRA2p*, and *SRA2q*. *FQA* is ranked fourth, whereas *FPA* is by far the least efficient treatment.

To assess how welfare is distributed between buyer and sellers, we also computed their expected payoffs in equilibrium. The ranking in terms of buyer's utility essentially reproduces that in terms of social welfare: *SRA* and *SRA2p* generate the highest buyer's utility – actually, the buyer's utility is slightly larger in *SRA2p*, but the difference is negligible – followed by *SRA2q*, *FQA*, and *FPA*. On the other hand, the ranking largely reverses when looking at sellers, as the expected sellers' payoff is higher in the treatments with one-dimensional bids than it is in those with two-dimensional bids.

The welfare rankings across treatments can be understood in light of the differences in the strategy spaces. Intuitively, in *SRA* and, to a lesser extent, in *SRA2q* and *SRA2p*, sellers

[9] The efficient level for quality is denoted *FB* – which stands for first best – in Figure (1b).

have more flexible strategies at their disposal, as they can leverage on both quality and price to compete in the auction. In particular, a seller whose cost for quality is high can still be competitive by pairing a low-quality bid with a low-price. Clearly, this choice is not possible in treatments with one-dimensional bids, where sellers can rely on price or quality only as a competitive instrument. As a result, competitive pressure is stronger in the treatments with bi-dimensional bids (and, within these, it is stronger in *SRA* than in *SRA2q* and *SRA2p*): this increases efficiency and favors the buyer to the detriment of sellers. Notice, finally, that the shape of the cost function (3) is at the origin of the poor performance of *FPA* in terms of welfare: In fact, with quality fixed at some exogenous level, the convexity of costs generate cost differences that get larger and larger as the type increases. As a consequence, the competitive pressure from low to high types is extremely weak, negatively affecting both the social and the buyer's welfare.

We summarize the main predictions associated with our benchmark model in the following statements.

Theoretical Predictions. *In equilibrium:*

(a) In terms of expected social welfare, the ranking across treatments is as follows:

$$SRA \succ SRA2p \succ SRA2q \succ FQA \succ FPA.$$

(b) In terms of expected buyer's utility, the ranking across treatments is as follows:

$$SRA2p \succ SRA \succ SRA2q \succ FQA \succ FPA.$$

(c) In terms of expected sellers' payoff, the ranking across treatments is as follows:

$$FQA \succ SRA2q \succ FPA \succ SRA \succ SRA2p.$$

(d) In all treatments, the score functions are strictly increasing in types.

(e) For all types, the score is maximal in SRA and minimal in FPA. FQA, SRA2q and SRA2p lie in between.

(f) The score function in FQA is steeper than the score function in SRA.

(g) The score in FQA is lower (higher) than the score in SRA2q for types $\theta \leq 6$ $(\theta > 6)$*.*

(h) The score in FQA is lower (higher) than the score in SRA2p for types $\theta \leq 8$ $(\theta > 8)$*.*

(i) With some exceptions for low types, in all treatments the submitted quality is below the efficient level.

5 Experimental results

The experimental results are presented in two steps. First, we focus on the level of social welfare generated in our treatments and how it is distributed between buyer and sellers. Next, we analyze the bidding behavior. To have a comparable measure across all treatments, we look at the observed score as defined by (1). For the scoring auctions, we also look into the two components of the score, that is, the price and quality bids.

The non-parametric tests presented here are based on twelve independent observations (at the rematching group level) per treatment. Similarly, in the parametric analysis, we properly account for dependency of observations over repetitions by either clustering standard errors, or introducing random effects at the rematching group level. All regressions pool data from the five treatments and use *FQA* as a baseline.

5.1 Welfare

We consider three measures of welfare: the overall social welfare (SW), the buyer's utility (BU) and the supplier's payoff (SP). These measures are constructed as follows: For each pair and in each period, we divide the realized social welfare (as defined by (5)), the buyer's utility (as defined by (4)) and the monetary payoff of the winning sellers (as defined by (2)) by the level of welfare associated with the efficient allocation, that is, the level of overall surplus that would have been generated if the good had been awarded to the low-cost seller and this seller had provided the most efficient quality level. Then, to control for potential (statistical) dependency, we average these measures by period and rematching group.

Table 1 shows the descriptive statistics for these welfare measures in the five treatments. The table also reports the corresponding theoretical predictions and the results from a (two-sided) Mann-Whitney rank-sum test for the null hypothesis of equality between observed and predicted levels. The rank-sum test is computed for rematching groups only, averaging observations over all periods.

Table 1 – Welfare: descriptive statistics

	SW			*BU*			*SP*		
	Avg.	Pred.	$p(Avg. = Pred.)$	Avg.	Pred.	$p(Avg. = Pred.)$	Avg.	Pred.	$p(Avg. = Pred.)$
FPA	0.609	0.613	1.000	0.430	0.402	0.182	0.179	0.212	0.034
	(0.195)			(0.200)			(0.092)		
FQA	0.834	0.788	0.002	0.665	0.557	0.002	0.169	0.231	0.002
	(0.066)			(0.086)			(0.052)		
SRA2q	0.801	0.855	0.006	0.597	0.643	0.136	0.204	0.212	0.530
	(0.133)			(0.181)			(0.148)		
SRA2p	0.805	0.889	0.003	0.620	0.704	0.008	0.185	0.186	0.814
	(0.143)			(0.177)			(0.153)		
SRA	0.823	0.910	0.002	0.625	0.702	0.015	0.198	0.208	0.182
	(0.452)			(0.225)			(0.613)		
Obs.		900			900			900	

Note. For each of the three relative efficiency measures, *SW*, *BU*, and *SP*, this table reports (i) mean and standard deviation (in parentheses), (ii) predicted values and (iii) p-value of a (two-sided) Mann–Whitney rank-sum test for the null hypothesis of equality between mean and predicted values. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

According to theory, *SRA* should be the most efficient mechanism since it can extract 91% of the potential surplus, followed by *SRA2p* (88.9%) and *SRA2q* (85.5%). Social welfare should be lower in the two treatments that have one-dimensional choice sets, with *FQA* and *FPA* generating 78.8% and 61.3% of the potential surplus, respectively.

Results differ in the lab, where *FQA*, which ranked fourth theoretically, is the most efficient mechanism (83.4% of the potential surplus), followed by *SRA* (82.3%), *SRA2p* (80.5%), and *SRA2q* (80.1%). This result derives from *FQA*'s significantly outperforming its theoretical prediction and the scoring auctions' significantly underperforming. Overall, across these four treatments, the observed differences are small, but *FPA* is by far the least efficient treatment also in the lab (with no significant difference between predicted and observed level).

Essentially all the welfare loss recorded in the scoring auctions is borne by the buyer: The difference between the predicted and the observed buyer's utility is negative and significant in *SRA* and *SRA2p* and is negative but not significant in *SRA2q*, while no significant difference is recorded with respect to sellers' payoffs.

On the other hand, the observed overperformance of *FQA* greatly benefits the buyer (10.8% with respect to theory), whereas the sellers' payoff is slightly, though significantly, lower than predicted (−3.3%).

To assess the statistical validity of these preliminary observations, Table 2 reports parametric results of the determinants of these welfare measures.

In terms of social welfare (columns (1) and (2)), *FPA* is the least efficient treatment, as all the pairwise differences between *FPA* and the other treatments are negative and highly significant (in all cases, $p < 0.001$). Moving to the comparisons of *FQA* with the scoring mechanisms, we detect no significant difference with respect to *SRA* ($p = 0.737$), positive and marginal significance with respect to *SRA2p* ($p = 0.081$), positive and significant difference with respect to *SRA2q* ($p = 0.039$), and no significant differences across the three scoring mechanisms ($p = 0.547$ between *SRA* and *SRA2q*; $p = 0.621$ between *SRA* and *SRA2p*; $p = 0.863$ between *SRA2q* and *SRA2p*). These results do not change qualitatively when a linear time trend is added, as in column (2). The coefficient of the linear trend is positive and significant in *SRA2p* ($p < 0.001$), *SRA2q* ($p = 0.028$), and *FQA* ($p = 0.025$), and the trend is largely absent in the remaining two treatments, being only marginally significant in *SRA* ($p = 0.097$), and not significant in *FPA* ($p = 0.520$).

Table 2 also allows us to compare the observed social welfare with its theoretical prediction: Observed social welfare in *FQA* is higher than what is theoretically predicted (0.046, $p < 0.001$). On the other hand, all of the scoring mechanisms generate a significantly lower-than-predicted level of SW: −0.087 in *SRA* ($p = 0.007$), −0.085 in *SRA2p* ($p < 0.001$), −0.054 in *SRA2q* ($p < 0.001$). Finally, we detect no significant difference between the observed and the predicted level of SW in *FPA* ($p = 0.796$).[10] We summarize the main results concerning social welfare below.

Result 1.1. Social welfare: ranking. No significant differences in the observed social welfare are detected among *FQA*, *SRA2q*, *SRA2p*, and *SRA*; social welfare is significantly lower in *FPA*.

Result 1.2. Social welfare: observed vs. predicted. Observed social welfare is above its predicted level in *FQA* and below its predicted level in the scoring auctions, while no significant difference is recorded for *FPA*.

The results regarding the buyer's utility closely resemble those concerning social welfare. (See Table 2, columns (3) and (4).) *FPA* yields the lowest utility to the buyer: Indeed, the pairwise differences between *FPA* and all other treatments are negative and highly significant

[10]The main results concerning social welfare do not change if we restrict the analysis to the last five periods of the experiment to account for the effects of subjects' experience (Table A1 in the Appendix). Specifically, we still do not detect any difference among *FQA*, *SRA*, *SRA2p*, and *SRA2q*, whereas SW is significantly lower in *FPA*. In all treatments except *FPA*, social welfare is higher in the last five periods than overall: As a result, the difference with respect to the theoretical level is larger in *FQA* (0.058, $p < 0.001$) and smaller, but still significant, in the scoring mechanisms (in *SRA*: −0.040, $p = 0.011$; in *SRA2p*: −0.067, $p < 0.001$; in *SRA2q*: −0.041, $p = 0.025$). In the last five periods, the time trend is no longer significant in any of the treatments except *FQA*, where it is positive ($p = 0.035$).

Table 2 – Welfare: parametric analysis

	SW		*BU*		*SP*	
	(1)	(2)	(3)	(4)	(5)	(6)
FPA	-0.224***	-0.189***	-0.235***	-0.216***	0.010	0.027
	(0.016)	(0.031)	(0.023)	(0.040)	(0.015)	(0.019)
SRA2q	-0.033**	-0.051*	-0.067***	-0.144***	0.035*	0.093**
	(0.016)	(0.029)	(0.025)	(0.044)	(0.020)	(0.042)
SRA2p	-0.029*	-0.065***	-0.045*	-0.096**	0.017	0.031
	(0.017)	(0.024)	(0.024)	(0.048)	(0.018)	(0.046)
SRA	-0.011	-0.063	-0.040	-0.098	0.029	0.035
	(0.033)	(0.076)	(0.026)	(0.061)	(0.046)	(0.122)
Trend		0.003**		0.002*		0.001
		(0.001)		(0.001)		(0.001)
FPA × *Trend*		-0.005		-0.003		-0.002
		(0.003)		(0.004)		(0.002)
SRA2q × *Trend*		0.003		0.011***		-0.008**
		(0.003)		(0.004)		(0.004)
SRA2p × *Trend*		0.005***		0.007*		-0.002
		(0.002)		(0.004)		(0.004)
SRA × *Trend*		0.007		0.008		-0.001
		(0.006)		(0.006)		(0.011)
Constant	0.834***	0.813***	0.665***	0.647***	0.169***	0.165***
	(0.006)	(0.012)	(0.010)	(0.016)	(0.006)	(0.009)
Obs.	900	900	900	900	900	900
Wald $-\chi^2$	188.67	284.86	100.87	147.76	4.06	16.54
$p > -\chi^2$	0.000	0.000	0.000	0.000	0.398	0.056

Note. Table 2 report estimates (robust standard errors in parentheses) from GLS random effects models accounting for dependency within rematching group. In all regressions, the dependent variable is defined at the rematching group level. Columns (1) and (2) focus on *SW*, Columns (3) and (4) focus on *BU*, Columns (5) and (6) focus on *SP*. Trend is a linear time trend that starts from 0 in the first period of the experiment. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

(in all cases, $p < 0.001$). The realized buyer's utility is similar in *FQA* and in the scoring auctions: We do not document significant differences between *FQA* and *SRA*, between *SRA* and *SRA2q* ($p = 0.407$), between *SRA* and *SRA2p* ($p = 0.864$), or between *SRA2q* and *SRA2p* ($p = 0.487$); we detect a positive and significant difference only between *FQA* and *SRA2q* ($p = 0.008$) and, marginally, between *FQA* and *SRA2p* ($p = 0.060$). Again, adding a linear time trend does not alter these results; the trend is positive and significant in *SRA* ($p = 0.046$), *SRA2q* ($p < 0.001$), *SRA2p* ($p = 0.01$), and, marginally, *FQA* ($p = 0.096$), while it is not significant in *FPA* ($p = 0.960$).

As is the case for social welfare, the observed level of buyer's utility is higher than theoretically predicted in *FQA* – the difference is 0.108, $p < 0.001$ – whereas the opposite occurs for the scoring auctions. (In *SRA*, the difference is -0.077, $p = 0.001$; in *SRA2p*, the difference is -0.084, $p < 0.001$; in *SRA2q*, the difference is -0.046, $p = 0.048$); and no significant difference is detected for *FPA* ($p = 0.178$).[11]

Columns (5) and (6) of Table 2 focus on sellers' payoff. We detect few differences across treatments: The only marginally significant (positive) difference is that between *SRA2q* and *FQA* ($p = 0.077$), while all the other pairwise comparisons are not statistically significant. With respect to the theoretical prediction, we do not observe significant differences in the scoring auctions (in *SRA*: $p = 0.821$; in *SRA*: $p = 0.971$; in *SRA2q*: $p = 0.661$), whereas both *FQA* and *FPA* underperform (in *FQA*: -0.062, $p < 0.001$; in *FPA*: -0.032, $p = 0.016$). Finally, we detect a negative trend in *SRA2q* and *FPA* —- significant in the former ($p = 0.025$) and marginally significant in the latter ($p = 0.090$) —- whereas it is not significant in *SRA* ($p = 0.970$), *SRA2p* ($p = 0.712$), or *FQA* ($p = 0.651$).[12]

We summarize the main results concerning buyer's utility and sellers' payoff below.

Result 2.1. Buyer's utility: ranking. Buyer's utility is higher in *FQA* than in *SRA2p* or *SRA2q*. No remarkable differences are detected among *SRA*, *SRA2p*, and *SRA2q*. Finally, *FPA* generates the lowest level of buyer's utility.

Result 2.2. Buyer's utility: observed vs. predicted. The observed buyer's utility is above its predicted level in *FQA* and below its predicted level in the scoring auctions, while no significant difference is recorded for *FPA*.

Result 2.3. Sellers' payoff: ranking. No remarkable differences are detected across treatments in terms of sellers' payoff.

Result 2.4. Sellers' payoff: observed vs. predicted. The observed sellers' payoff is aligned with its predicted level in the scoring auctions and below its predicted level in *FQA* and *FPA*.

[11] In the last five periods only (Table A1 in the Appendix), we find no differences between *SRA2q* and *FQA* ($p = 0.122$), between *SRA2p* and *FQA* ($p = 0.220$), between *SRA* and *FQA* ($p = 0.174$), or between *SRA2p* and *SRA2q* ($p = 0.717$). Buyer's utility in *SRA* is higher than it is in either *SRA2p* ($p = 0.039$) or *SRA2q* ($p = 0.022$). *FPA* is the treatment with the lowest *BU* (in all the pairwise differences with the other treatments, $p < 0.001$). In all treatments, the time trend is no longer significant. Finally, buyer's utility is higher than predicted in *FQA* – the difference is 0.114, $p < 0.001$ – while the difference is negative and significant in *SRA2p* (-0.058, $p = 0.003$), negative and marginally significant in *SRA* (-0.032, $p = 0.074$), and not significant in either *SRA2q* ($p = 0.998$) or *FPA* ($p = 0.711$).

[12] Results on pairwise differences are confirmed when the analysis is replicated on the last five periods (Table A1 in the Appendix). The only two positive differences in *SP* that reach marginal significance are those between *SRA* and *FQA* ($p = 0.078$) and between *SRA* and *SRA2q* ($p = 0.070$). Moreover, we find a negative and significant linear time trend only in *SRA2q* ($p = 0.036$).

Our prediction of the results concerning welfare shows that the two dimensions that jointly determine the efficiency of the final allocation generated in the auction are cost efficiency and quality efficiency (Section 4). Recalling that, theoretically, all treatments should be cost-efficient – that is, the high-type (low-cost) seller should always win – the (theoretical) ranking in terms of social welfare is fully determined by differences in terms of quality efficiency. We take the observed fraction of auctions that the seller with the highest θ wins (denoted by CE) to measure the observed degree of cost efficiency. On the other hand, we measure quality efficiency with the percentage distance between the quality bid submitted by the winner of the auction and the theoretically efficient quality level (denoted by QE). We average these measures by period and by rematching group.

Table 3 shows the descriptive statistics for CE and QE in the five treatments.

Table 3 – Cost and quality efficiency: descriptive statistics

	FPA	*FQA*	*SRA2q*	*SRA2p*	*SRA*
CE	0.837	0.919	0.843	0.809	0.843
	(0.224)	(0.160)	(0.221)	(0.223)	(0.198)
QE	0.542	0.362	0.317	0.287	0.249
	(0.130)	(0.084)	(0.144)	(0.135)	(0.140)
Obs.	180	180	180	180	180

Note. This table reports mean and standard deviation (in parentheses) for cost efficiency (CE) and quality efficiency (QE), overall periods and by treatment.

In terms of cost efficiency, the best treatment is *FQA*, which selects the most efficient seller 91.9% of the time, followed by *SRA* and *SRA2q*, which select the most efficient seller 84.3% of the time. In terms of quality efficiency, instead, the observed ranking fully obeys the theoretical one: *SRA* is the treatment in which, on average, the quality level provided by the winner is closer to the efficient level – the average distance is 24.9% – followed by *SRA2p* (28.7%), *SRA2q* (31.7%), *FQA* (36.2%), and *FPA* (54.2%, recall that, in this last treatment, quality was fixed).

Table 4 reports parametric results on the determinants of CE and QE. Neither trend, nor any trend-treatment interaction are significant. (See columns (2) and (4).) Therefore, we focus on the two baseline models with treatment-dummies only.

Results on cost efficiency confirm that *FQA* is the treatment in which the low-cost seller wins more often: In fact, all the pairwise comparisons between *FQA* and the other treatments are positive and highly significant (in all cases, $p < 0.001$). We find no significant difference across the scoring mechanisms ($p = 0.118$ between *SRA* and *SRA2p*; $p = 1.000$ between *SRA* and *SRA2q*; $p = 0.189$ between *SRA2q* and *SRA2p*) or between any of the bidimensional mechanisms and *FPA* ($p = 0.806$ between *FPA* and *SRA*; $p = 0.281$ between *FPA* and *SR2pA*; $p = 0.834$ between *FPA* and *SRA2q*).

The differences in QE across treatments resemble our theoretical predictions. In all scoring mechanisms, the quality submitted by the winning seller is significantly closer to the efficient level than it is in *FQA* ($p < 0.001$ between *SRA* and *FQA*, and between *SRA2p* and *FQA*; $p = 0.012$ between *SRA2q* and *FQA*). Among the scoring auctions, *SRA* is more efficient in terms of submitted quality than *SRA2q* ($p = 0.001$) and marginally more effficient than

Table 4 – Cost and quality efficiency: parametric analysis

	CE	*CE*	*QE*	*QE*
FPA	-0.082***	-0.088**	0.180***	0.152***
	(0.021)	(0.036)	(0.013)	(0.023)
SRA2q	-0.076***	-0.125***	-0.045**	-0.045
	(0.021)	(0.036)	(0.018)	(0.032)
SRA2p	-0.109***	-0.136***	-0.074***	-0.064**
	(0.020)	(0.030)	(0.020)	(0.031)
SRA	-0.076***	-0.111***	-0.112***	-0.132***
	(0.015)	(0.040)	(0.014)	(0.025)
Trend		0.004		-0.003
		(0.003)		(0.002)
FPA × *Trend*		0.001		0.004
		(0.005)		(0.003)
SRA2q × *Trend*		0.007*		$-1 \cdot 10^{-4}$
		(0.004)		(0.003)
SRA2p × *Trend*		0.004		-0.001
		(0.004)		(0.003)
SRA × *Trend*		0.005		0.003
		(0.005)		(0.003)
Constant	0.919***	0.889***	0.362***	0.382***
	(0.009)	(0.022)	(0.007)	(0.016)
Obs.	900	900	900	900
Wald $-\chi^2$	51.88	103.90	353.15	382.50
$p > -\chi^2$	0.000	0.000	0.000	0.000

Note. This table reports estimates (robust standard errors in parentheses) from one-way linear random effects model accounting for dependency withing rematching group. Unit of observation is at the rematching level. Columns (1) and (2) focus on Quality Efficiency (*CE*), Columns (3) and (4) focus on Cost Efficiency (*QE*). Trend is a linear time trend that starts from 0 in the first period of the experiment. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

SRA2p ($p = 0.090$). We do not detect any significant difference between *SRA2p* and *SRA2q* ($p = 0.246$). *FPA* is the least efficient treatment in terms of quality, as all the pairwise differences between *FPA* and the other treatments are positive and highly significant (in all cases, $p < 0.001$).

We summarize the main results concerning cost and quality efficiency below.

Result 3. Cost and quality efficiency. With respect to cost efficiency, *FQA* is the best treatment, while no significant differences are observed among the remaining treatments. With respect to quality efficiency, the ranking across treatments is the same as the theoretical ranking in terms of social welfare.

Result 3 is of particular interest, as it sheds more light on the reasons behind the discrepancy between observed and predicted ranking in terms of social welfare (results 1.1 and 1.2). In particular, given that the ranking in terms of quality efficiency is aligned with the theoretical ranking in terms of social welfare, the finding that *FQA* generates (at least) as much welfare as the scoring auctions must be entirely ascribed to the other determinant of efficiency, cost efficiency. In fact, *FQA* is superior to all other treatments in selecting the low-cost seller as the winner.

5.2 Bidding behavior

To get a deeper understanding of the determinants of the results for social welfare, we analyze the bids. To allow for an easy comparison across treatments, we first focus on the score associated with the bid(s) submitted by sellers. Then, in the second part of this subsection we look separately at the price and quality bids for the bidimensional treatments. To facilitate comparison with the theoretical predictions, we also look at the percentage distance between observed and theoretically predicted bids: The corresponding variables are denoted *score_dist* for the score, *quality_dist* for the quality bid, and *price_dist* for the price bid.[13] We define *overbidding* whenever the observed bids are more aggressive than the theoretical ones (i.e., when the observed score/quality is higher than predicted, or when the observed price is lower than predicted). We define *underbidding* if the opposite occurs.

5.2.1 Score

Table 5 reports descriptive statistics for scores, price, and quality bids in the five treatments.[14]

[13] Specifically, *score_dist* = (observed scores−predicted scores)/predicted scores. *quality_dist* and *price_dist* are similarly defined.

[14] Figures A1, A2, and A3 in the Appendix provide a graphic representation of the descriptive statistics by treatment and seller's cost parameter.

Table 5 – Bids: descriptive statistics

	FPA	*FQA*	*SRA2q*	*SRA2p*	*SRA*
Score	51.977	60.500	56.479	56.428	53.874
	(16.499)	(12.147)	(11.870)	(13.244)	(11.990)
Price	30.023	32.000	37.240	41.051	48.941
	(16.499)	(0.000)	(24.882)	(26.389)	(27.548)
Quality	16.000	21.250	21.859	23.740	26.407
	(0.000)	(6.074)	(15.280)	(14.947)	(15.401)
score_dist	0.008	0.044	-0.059	-0.095	-0.144
	(0.164)	(0.077)	(0.156)	(0.183)	(0.160)
price_dist	-0.002	0.000	0.595	0.819	1.203
	(0.361)	(0.000)	(1.789)	(1.765)	(2.177)
quality_dist	0.000	0.061	0.111	0.202	0.308
	(0.000)	(0.114)	(0.664)	(0.718)	(0.568)
Obs.	1080	1080	1080	1080	1080
$p(score\_dist = 0)$	0.530	0.002	0.004	0.002	0.002
$p(price\_dist = 0)$	1.000	-	0.002	0.002	0.002
$p(quality\_dist = 0)$	-	0.002	0.002	0.005	0.002

Note. This table reports, for each treatment, mean and standard deviation (in parentheses) of scores, prices, and qualities associated with suppliers' choices. Statistics are built for both observed measures and as percent distance from predicted (Nash) levels. The table also reports p-values of a (two-sided) rank-sum test for the null hypothesis that the observed percent distance from predicted (Nash) levels is equal to zero.

Descriptive statistics highlight three main facts concerning the scores. First, unlike what is theoretically predicted, the average submitted score is highest in *FQA*, followed by *SRA2p* and *SRA2q*. *SRA* is ranked fourth, closely followed by *FPA*. Second, we observe overbidding in the unidimensional treatments, with the average score in *FQA* 4.4% higher than the predicted level. Third, all the scoring auctions are characterized by underbidding, with scores in *SRA* well below their predicted level (-14.4%). A (two-sided) rank-sum test confirms that *score_dist* is significantly different from zero for all treatments except *FPA*.

Table 6 investigates parametrically the differences across the treatments and determinants of both score and *score_dist*.

The baseline model (column (1)) confirms that *FQA* is associated with the highest score (for all the coefficients of the treatment dummies: $p < 0.001$). We find a nonsignificant difference between *SRA2p* and *SRA2q* ($p = 0.965$), while both these treatments are associated with a higher score than in *SRA* (between *SRA2p* and *SRA*: $p = 0.027$; between *SRA2q* and *SRA*: $p = 0.024$). Finally, we do not detect any significant difference between *SRA* and *FPA* ($p = 0.101$).

Table 6 also include the seller's type θ as a determinant of the score. In line with the theoretical predictions, the seller's type significantly increases the score in all treatments (in all cases, $p < 0.001$). In column (2), *SRA*, *SRA2p*, and *SRA2q* have positive coefficients, whereas their interactions with θ is negative. Therefore, we can use the estimates to determine for which seller' type the score difference between *FQA* and any of the scoring auctions becomes significant. We find that the submitted score in *FQA* is: (i) above the score in *SRA* for $\theta \in [3, 10]$; (ii) above the score in *SRA2p* for $\theta \in [4, 10]$; and (iii) above the score in *SRA2q* for $\theta \in [5, 10]$.

Finally, we add – in column (3), Table 6 – treatment-specific linear time trends as determinants of the score. We detect a positive and significant time pattern in the scoring auctions (in all cases: $p < 0.001$) and in *FPA* ($p = 0.002$). Even after controlling for the type parameter and the linear trend, the score in *FQA* remains higher than it is in *SRA* and *FPA* (in both cases, $p < 0.001$). We find a nonsignificant difference between *SRA2p* and *SRA2q* ($p = 0.906$), while both of these treatments are associated with a higher score than in *SRA* ($p = 0.018$ between *SRA2p* and *SRA*; $p = 0.013$ between *SRA2q* and *SRA*). Finally, we find a significantly lower score in *FPA* than in *SRA* ($p < 0.01$).[15]

We now turn our attention to *score_dist*, the deviation of the observed score, in percentage, from its predicted level. Results are also reported in Table 6. In the baseline model – in column (4) – we find a positive deviation (overbidding) of 4.41% in *FQA* ($p < 0.001$) and significant underbidding in all the scoring auctions of −5.95% in *SRA2q*, −9.48% in *SRA2p*, and −14.43% in *SRA* (in all cases, $p < 0.001$). No significant difference is observed in *FPA*.

These results remain significant even after controlling – in column (5) – for sellers' type: In *SRA* and *SRA2p*, observed scores are significantly below their predicted levels for all type parameters (while, in *SRA2q*, this occurs for $\theta \in [1, 7]$). Instead, in *FQA*, the overbidding is significant for $\theta \in [4, 10]$, and the degree of overbidding increases with θ. For example, a supplier of type $\theta = 5$ is associated with a positive deviation of 3.96% in *FQA* ($p = 0.001$) and negative deviations of 14.85% in *SRA* ($p < 0.001$), 10.18% in *SRA2p* ($p < 0.001$), and 6.53% in *SRA2q* ($p < 0.001$), respectively.

Controlling for the linear trend – in column (6) – does not affect the results in *FPA* or *FQA*

[15]This significant time pattern suggests the possibility of some learning effects in the bidding strategies. As a robustness check, we replicate all of the regressions by focusing on the last five periods of the experiment (Table A2 in the Appendix): The trend coefficients – in column (3) – are no longer significant. While results on the baseline model – column (1) – might suggest that most of the differences between *FQA* and the other treatments have disappeared, those differences persist and are strongly significant, but they depend on the type parameter θ. According to our theoretical predictions, as θ increases, the equilibrium score in *FQA* approaches the equilibrium score in *SRA*. In the lab and in the last five periods, as θ increases, the score in *FQA* approaches and then exceeds the *SRA*'s score. In particular, we detect – in column (2) – that the score in *FQA* is (i) below the score in *SRA* for $\theta \in [1, 3]$, (ii) not significantly different for $\theta \in [4, 5]$, and (iii) above the score in *SRA* for $\theta \in [6, 10]$. Similar results are obtained in the comparison of the score in *FQA* and the scores in *SRA2p* and in *SRA2q*.

Table 6 – Score: parametric analysis

	Score			*score_dist*		
	(1)	(2)	(3)	(4)	(5)	(6)
FPA	-8.523***	-8.650***	-9.636***	-0.036**	0.046**	0.034
	(1.157)	(1.249)	(1.361)	(0.017)	(0.020)	(0.021)
SRA2q	-4.021***	2.988**	-1.976	-0.104***	-0.123***	-0.206***
	(1.157)	(1.252)	(1.363)	(0.017)	(0.020)	(0.022)
SRA2p	-4.072***	2.741**	-2.137	-0.139***	-0.166***	-0.247***
	(1.157)	(1.255)	(1.360)	(0.017)	(0.020)	(0.021)
SRA	-6.626***	1.482	-5.328***	-0.188***	-0.205***	-0.318***
	(1.157)	(1.246)	(1.355)	(0.017)	(0.020)	(0.022)
θ		3.870***	3.871***		0.009***	0.009***
		(0.093)	(0.088)		(0.001)	(0.001)
FPA $\times\theta$		0.166	0.174		-0.015***	-0.015***
		(0.132)	(0.124)		(0.002)	(0.002)
SRA2q $\times\theta$		-1.251***	-1.215***		0.004*	0.004**
		(0.131)	(0.124)		(0.002)	(0.002)
SRA2p $\times\theta$		-1.235***	-1.238***		0.005**	0.005**
		(0.131)	(0.124)		(0.002)	(0.002)
SRA $\times\theta$		-1.397***	-1.372***		0.003*	0.004**
		(0.130)	(0.123)		(0.002)	(0.002)
Trend			0.039			0.001
			(0.057)			(0.001)
FPA $\times$ *Trend*			0.134*			0.002
			(0.081)			(0.001)
SRA2q $\times$ *Trend*			0.681***			0.011***
			(0.081)			(0.001)
SRA2p $\times$ *Trend*			0.699***			0.012***
			(0.081)			(0.001)
SRA $\times$ *Trend*			0.954***			0.016***
			(0.081)			(0.001)
Constant	60.500***	39.188***	38.903***	0.044***	-0.005	-0.010
	(0.818)	(0.889)	(0.966)	(0.012)	(0.014)	(0.015)
Obs.	5400	5400	5400	5400	5400	5400
Wald $-\chi^2$	61.61	6031.89	7405.95	158.64	470.14	1237.72
$p > -\chi^2$	0.000	0.000	0.000	0.000	0.000	0.000

Note. This table reports estimates (clustered standard errors in parentheses) from two-way linear random effects models accounting for both potential individual dependency over repetitions and dependency within rematching group. The first three columns are based on regressions using the observed score as dependent variable, while the last three are based on *score_dist*. θ is the cost parameter randomly assigned to the supplier. Trend is a linear time trend that starts from 0 in the first period. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

but reduces the magnitude of the underbidding that characterizes the scoring auctions.[16]

We summarize the main results concerning the observed score below.

Result 4.1. Score: relationship with the seller's type. In all treatments, the observed score increases with the type parameter θ. The score increases with θ more quickly in *FQA* than in any of the scoring auctions.

Result 4.2. Score: ranking. The average score is highest in *FQA*, followed by *SRA2p* and *SRA2q*.

Result 4.3. Score: observed vs. predicted. We observe overbidding in *FQA* and underbidding in the scoring auctions.

Result 4.1 sheds light on the reasons behind the observed ranking in terms of social welfare. Unlike what was predicted theoretically, the social welfare generated in *FQA* is as high as it was in the scoring auctions (result 1.1). Result 3 suggests that this is due to differences in cost efficiency across treatments: The high type (i.e., low-cost) seller wins more frequently in *FQA* than it does in the scoring auctions. Result 4.1 tells us that, *on average*, in all treatments, higher type (lower-cost) sellers submit higher scores (and, in doing so, win), so it must be the case that deviations from average bidding behavior are larger and/or more frequent in the scoring auctions than they are in *FQA*. This conclusion will guide us in the structural part of the analysis.

Result 4.3 highlights that, as we move from treatments with one-dimensional bids to treatments with two-dimensional bids, bidding behavior (in terms of score) changes qualitatively from overbidding (in *FQA*) or bidding that is aligned with the theoretical bids (in *FPA*) to underbidding. We come back to this point in the structural analysis.

Finally, in *FQA*, overbidding in the score is equivalent to an overbidding in quality, recalling that sellers do not bid on price here. That the predicted quality bid in *FQA* is below the efficient level explains why we observe overperformance in terms of social welfare and buyer's utility in this treatment. On the other hand, to understand the underbidding in the score that is observed in the scoring auction, we take a deeper look at both price and quality bids.

5.2.2 Price and quality bids in the scoring auctions

Descriptive statistics of price and quality bids are reported, by treatment, in Table 5. A clear result stands out: In all scoring auctions, we observe overbidding in quality and underbidding in price,[17] as sellers tend to offer a higher-than-predicted level of quality, accompanied by a higher-than-predicted price bid.[18] A rank-sum test confirms that *quality_dist* and *price_dist* are significantly different from zero.

Tables 7 reports parametric results on the determinants of *quality_dist* and *price_dist* in *SRA* (columns 1 and 2), *SRA2p* (columns 3 and 4), and *SRA2q* (columns 5 and 6).

[16] In the last five periods (Table A2 in the Appendix), neither trend nor any specific trend-treatment interactions are significant. In the baseline model – in column 4 – we still find a significant underbidding in *SRA* (-7.54%; $\chi^2(1) = 52.02$, $p < 0.001$) and in *SRA2p* (-5.09%, test results: $\chi^2(1) = 23.74$, $p < 0.001$) and a significant overbidding in *FQA* ($+4.22\%$; $\chi^2(1) = 16.33$, $p < 0.001$). Adding the type parameter θ does not qualitatively alter these results.

[17] When we say "underbidding in price," we refer to a price *above* its predicted level.

[18] The results in Table 5 concerning *FPA* and *FQA* confirm what we already knew from the analysis on score: There is overbidding (in quality) in *FQA* and neither over- nor under-bidding (in price) in *FPA*.

Table 7 – Price and quality bids in the scoring auctions: parametric analysis

	SRA2q		*SRA2p*		*SRA*	
	(1)	(2)	(3)	(4)	(5)	(6)
price_dist						
θ		-0.043**		-0.121***		-0.321***
		(0.018)		(0.018)		(0.019)
Constant	0.595***	0.833***	0.819***	1.486***	1.203***	2.918***
	(0.084)	(0.131)	(0.122)	(0.154)	(0.122)	(0.148)
quality_dist						
θ		-0.001		-0.025***		-0.042***
		(0.007)		(0.007)		(0.005)
Constant	0.111***	0.117***	0.202***	0.339***	0.308***	0.531***
	(0.021)	(0.043)	(0.044)	(0.059)	(0.032)	(0.042)
Obs.	1080	1080	1080	1080	1080	1080
cov(e.price_dist, e.quality_dist)	1.065***	1.065***	1.124***	1.101***	0.836***	0.724***
	(0.048)	(0.048)	(0.051)	(0.049)	(0.043)	(0.038)

Note. This table reports estimates (clustered standard errors in parentheses) from Seemingly Unrelated Regression (SUR) models allowing the standard errors of the linear models to be correlated. Each regression is based on a two-way linear random effects model accounting for both potential individual dependency over repetitions and dependency within rematching group. The dependent variables are *quality_dist* and *price_dist*, respectively. Regressions are run separately by treatment: $SRA2q$ in columns (1) and (2), $SRA2p$ in columns (3) and (4) and SRA in columns (5) and (6). θ is the cost parameter randomly assigned to the supplier. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

In all treatments, we find a significant overbidding in quality and a significant underbidding in price. As expected, the covariance of the two linear models' residuals is always significant. The seller's type negatively affects both the price and the quality distance in all treatments.

We summarize below the main results concerning price and quality bids in the scoring auctions.

Result 5. Quality and price bids in the scoring auctions. In all the scoring auctions, we observe overbidding in quality and underbidding in price.

In all treatments in which quality can be chosen (*FQA*, *SRA2q*, *SRA2p* and *SRA*), sellers have a clear tendency to offer a comparatively high level of quality, one that is significantly above that predicted by our theoretical model. However, in the scoring auctions, where sellers can also choose the price, this overbidding in quality is more than offset by even stronger underbidding in the price, leading to underbidding in the score (result 4.3) and explaining why the buyer's utility is lower than predicted in these treatments (result 2.2).

6 Structural analysis

The previous analysis highlights contrasting results on to the relationship between theoretical and observed bidding behavior.

First, relative to what predicted by our benchmark model of risk neutral sellers and equilibrium play, we observe that sellers clearly overbid in *FQA*, but this overbidding behavior turns to underbidding in treatments with a two-dimensional choice: here the submitted score is below the equilibrium one. The results are less clear-cut for *FPA*, where observed bids are, on average, in line with the theoretical predictions. The experimental literature on ordinary auctions points out that, when the winner of the auction is called to pay what she bids, bidders typically tend to overbid relative to the risk neutral equilibrium, as if they attached a higher value to winning the auction than what a risk neutral bidder would do. Several explanations to this phenomenon have been put forward, though the most common and most natural is certainly risk aversion.[19] Our mixed evidence, however, can hardly be explained in terms of a departure from risk neutrality alone: if bidders attached extra-value to winning the auction, they should overbid in all of our treatments, whereas we observe overbidding in *FQA* but underbidding in *SRA*, *SRA2q* and *SRA2p*.

The second remarkable result of our experiment is that *FQA* performs (at least) as well as the scoring auctions in terms of social welfare, whereas theory predicts a clear superiority of the latter. The evidence on cost efficiency made it clear that this result is due to the fact that, in the scoring treatments, the auction is more frequently won by the high-cost seller. Finally, the analysis of the relation between submitted score and type parameter led us to conclude that this happens because bids are more noisy in the scoring auctions than in *FQA*

The observation that the transition from overbidding to underbidding occurs as we move from one-dimensional to two-dimensional treatments, and the evidence that, in the latter treatments, bids tend to be more noisy, lead us to suspect that subjects' behavior could be somewhat related to the degree of complexity of the auction. Intuition suggests that choosing price and quality simultaneously is a more complex task than choosing only one or the other. Besides, when a two-dimensional bid is to be made, the choice is arguably easier when, on one dimension, only two markedly different alternatives are available, as is the case in treatments *SRA2q* and *SRA2p*. According to this intuition, the five treatments considered in our experiment are characterized by three levels of complexity: Treatments with unidimensional choice (*FQA* and *FPA*) are the least complex, *SRA* is the most complex, and *SRA2q* and *SRA2p* – treatments with two-dimensional choice, one of which is binary – lie in between.

Table 8 – Response time: descriptive statistics

	FPA	*FQA*	*SRA2q*	*SRA2p*	*SRA*
Response time	26.847	34.684	60.118	70.620	96.780
	(20.759)	(23.884)	(30.427)	(32.132)	(39.893)
Obs.	1080	1080	1080	1080	1080

Note. This table reports, for each treatment, the mean and the standard deviation (in parentheses) of response time in seconds.

This intuition is corroborated by the observation of the subjects' response times in the experiment. Table 8 shows, for every treatment, the average time elapsed before a subject submitted her bid in a generic period of the experiment. The difference in the response time between one-dimensional and two-dimensional treatments is remarkable. Moreover, among the scoring auctions, *SRA* required more time to answer than *SRA2p* or *SRA2q*. A battery of

[19] For a survey on this literature, see Kagel (1995) and Kagel and Levin (2011).

pairwise comparison tests shows that response time differs significantly across treatments.[20]

6.1 The Quantal Response Equilibrium approach

Once one accepts that our auction games differ in the degree of complexity in the sellers' decisions, then the question that naturally follows is how the complexity of the task affects behavior. We consider the simplest answer to this question: individuals, when faced with more complex tasks, are simply more likely to make errors, i.e., to make suboptimal choices. Following this idea, we adopt a model that explicitly allows the possibility that individuals involved in strategic interaction can make errors: the Quantal Response Equilibrium (QRE) introduced by McKelvey and Palfrey (1995). The QRE has been repeatedly applied to model non-equilibrium behavior observed in experimental auctions (often combined with other biases). In particular, Goeree et al. (2002) show how this model, even in the presence of risk averse bidders, may generate both over- and underbidding in private value auctions, depending on the structure of bidders' payoffs. In this sense, it is a promising model for use in explaining our contrasting evidence.

In a QRE model, the assumption that a player always chooses the strategy that maximizes her payoff (i.e., her best response) is replaced by a probabilistic choice function tuned by an error parameter: The probability of playing a suboptimal strategy is strictly positive, but it depends on the (relative) payoff associated with it. In other words, an individual is more likely to make an error that determines a small loss (relative to the payoff-maximizing strategy) than an error that causes a big loss. The error parameter measures the sensitivity of choices to payoffs. Therefore, in a QRE model, the assumption of full rationality is relaxed but the equilibrium requirement is preserved: players have correct expectations for the other players' behavior; in particular, they take into consideration the noisiness embedded in their opponents' decisions.[21]

6.2 Baseline QRE model with risk aversion

We consider the following QRE model with logistic errors: a type-θ seller bids b with probability

$$\Pr(b;\theta) = \frac{\exp[U_S(b;\theta)/\mu]}{\sum_{b\in B}\exp[U_S(b;\theta)/\mu]}.$$

In this expression:

- B is the set of (admissible) individually rational bids, i.e., bids that give a non-negative payoff to the seller in case of winning;
- $U_S(b;\theta)$ is the expected payoff of a type-θ seller when she bids b and the other seller bids according to her own QRE strategy;

[20]Parametric analysis produces the same conclusion. When trend and trend-treatment interactions are included, response time declines over periods for all treatments, but the overall ranking does not change. The trend disappears in the last five periods. Results are available upon request.

[21]We do not investigate what cognitive process(es) may lead individuals to make suboptimal choices. Perhaps such choices are the result of the trade-off between cognitive effort and the quality of the decision: The individual optimally decides the amount of cognitive effort to devote to a task by weighting the extra-cost of additional effort with its extra-benefit in terms of (expected) improvement of the solution to the task. As a result, when a task is highly demanding in terms of cognitive costs, the decision maker may (optimally) decide to stop thinking about it when a satisfactory, but not necessarily the best, solution has been identified.

- $\mu \geq 0$ is the error parameter: the higher μ, the higher the probability that the seller makes a bid that yields a relatively low payoff. In the limiting case $\mu \to \infty$, the correlation between a seller's payoff and her bid disappears, and choices become essentially random; in contrast, when $\mu = 0$, the best response is played with certainty, and the model boils down to the standard Bayes-Nash equilibrium.

We also allow for possible departures from risk neutrality by considering a Constant Relative Risk Aversion (CRRA) utility function for sellers. In particular, the utility of a seller who wins the auction, is paid a price p, and delivers a quality q is equal to

$$u_S(p, q; \theta) = \frac{1}{1-r} \left[p - C(q; \theta)\right]^{1-r},$$

where $r \geq 0$ is the Arrow-Pratt coefficient of relative risk aversion, and $C(q; \theta)$ is given by (3).

Table 9 presents, for each treatment, the estimates of the two free parameters of this model: The error parameter μ and the coefficient of relative risk aversion r. For computational reasons, estimations were performed after grouping bids into bins: In particular, the space of admissible bids were divided into four-unit intervals, and, to each observation belonging to a certain interval, the central value of the interval was assigned.[22] All estimations were performed via maximum likelihood.

The results contained in Table 9 show that the estimates for the risk-aversion parameter r are similar across treatments but that there are significant differences in the error parameter μ. These differences are consistent with our intuition based on the complexity of the treatment: The simplest, unidimensional treatments – *FPA* and *FQA* – have smaller values of μ with respect to the two-dimensional treatments. Hence, submitted bids are closer to the best responses in the former than in the latter treatments. Moreover, among the scoring auctions, those in which one choice is simplified to a binary choice – *SRA2q* and *SRA2p* – have a lower value of μ than *SRA*, the most complex treatment.

Table 9 – Estimates from the baseline QRE model with risk aversion, full sample

	FPA	*FQA*	*SRA2q*	*SRA2p*	*SRA*
r	0.68	0.68	0.68	0.66	0.62
μ	0.78	0.42	0.91	1.04	1.16
η	0.389	0.422	0.995	1.277	1.535
	(0.029)	(0.138)	(0.104)	(0.197)	(0.061)
LL	−355.82	−261.97	−438.21	−311.57	−1376.21
ϕ_M	0.73	0.83	0.74	0.76	0.40

Note. This table reports, for each treatment, maximum likelihood estimates of the two parameters of the baseline QRE model with risk aversion: r (the Arrow-Pratt coefficient of relative risk aversion), and μ (the error parameter). η is the relative utility loss predicted by the estimated model (standard error in parenthesis), LL is the value of the log-likelihood function, ϕ_M is a comparable measure of the goodness of fit.

[22]The binning methodology adopted is described precisely in the Appendix. For *FPA* and *FQA*, we were able to estimate the model also without bins, obtaining very similar results.

One may object that a direct comparison of the estimates of the error parameter μ may be disputable, as these estimates come from different auction games with different strategy spaces. To facilitate comparability, we then construct a measure of departure from rationality, denoted by η, that is less sensitive to the details of the underlying game, being directly built on relative payoffs. Specifically, η is computed as the average quadratic deviation between the utility associated with the QRE strategy and the maximum utility achievable (the one obtainable by playing the best response strategy with probability one), normalized by the latter. In symbols,

$$\eta = \sum_{\theta=1}^{10} \sum_{b \in B} \left[\widehat{\Pr}_{\theta}(b) \cdot \left(\frac{U_S(b;\theta) - U_S(b^*(\theta);\theta)}{U_S(b^*(\theta);\theta)} \right)^2 \right],$$

where $\widehat{\Pr}_{\theta}(b)$ is the probability that a type-θ seller bids b, as predicted by the estimated QRE model, and $b^*(\theta)$ is her utility-maximizing bid (i.e., her best response).

Loosely speaking, η is a sort of "money-left-on-the-table" measure, as it captures how much utility the subject gives up, on average, by using a suboptimal strategy. Now, the ranking across treatments in terms of η supports our starting intuition even more cleanly than when we look at μ: the value of η in *FPA* and *FQA* is much lower than it is in the bidimensional treatments; moreover, it is significantly higher in *SRA* than it is in *SRA2q* and *SRA2p* (Table 9).

One may also wonder whether these differences in rationality disappear once subjects learn "how to play". To address potential learning dynamics, we re-estimated the above QRE model, restricting our attention to the last 5 periods, where no trend was parametrically observed. Results are reported in Table 10.

Table 10 – Estimates from the baseline QRE model with risk aversion, last 5 periods

	FPA	*FQA*	*SRA2q*	*SRA2p*	*SRA*
r	0.68	0.67	0.68	0.65	0.63
μ	0.71	0.40	0.59	0.72	0.76
η	0.353	0.399	0.738	0.997	1.239
	(0.027)	(0.135)	(0.092)	(0.175)	(0.056)
LL	−191.05	−110.32	−213.23	−144.84	−616.40
ϕ_M	0.64	0.81	0.71	0.77	0.44

Note. This table reports, for each treatment, maximum likelihood estimates of the two parameters of the baseline QRE model with risk aversion: r (the Arrow-Pratt coefficient of relative risk aversion), and μ (the error parameter). η is the relative utility loss predicted by the estimated model (standard error in parenthesis), LL is the value of the log-likelihood function, ϕ_M is a comparable measure of the goodness of fit.

With respect to the estimates obtained from all observations, we do not find any relevant variation in the risk-aversion parameter r. In contrast, the parameter μ decreases in all treatments: by gaining experience, subjects tend to play better strategies (i.e., closer to the Bayes-Nash equilibrium of the game). Interestingly, though not surprisingly, μ decreases only slightly in the simplest one-dimensional treatments (−8.6% for *FPA* and −5.6% for *FQA*), but much more sharply in the two-dimensional treatments (−34.9% for *SRA2q*, −30.9% for *SRA2p*, and −34.4% for *SRA*). Similar results can be obtained if one looks at η.

The change in the estimates of μ that are due to learning partially modifies the ranking across treatments: While the value of μ remains lowest in *FQA* and highest in *SRA*, *FPA* now has a higher estimate of μ than *SRA2q*, and essentially the same as *SRA2p*. However, the more comparable measure, η, confirms the same ranking as in the full sample: η decreases with the complexity of the treatment, being highest in *FPA* and *FQA*, lowest in *SRA*, and intermediate in *SRA2q* and *SRA2p*.

The results presented so far suggest that the QRE model is able to produce estimates that are consistent with the idea that the complexity of the auction mechanism is a crucial driver of bidding behavior. Still, however, we need to check how well this model fits our experimental data. Looking at the value of the log-likelihood function in the various tratments may be problematic, as different tratments involve different games with different strategy spaces. To overcome this problem, we follow Camerer et al. (2016) and adopt a normalized measure of relative fit that is invariant to the dimension of the strategy set. This measure, which is analogous to a *Pseudo-*R^2, compares the value of the log-likelihood in the estimated model with two extreme models: the first is an ideal "clairvoyant" model in which each (type-dependent) bid is played with a probability exactly equal to the observed relative frequency; the second is a purely random model in which, for every type θ, each (individually rational) strategy is played with equal probability. The measure of fit is then computed as:

$$\phi_M = 1 - \frac{\ln M - \ln M^*}{\ln Random - \ln M^*}$$

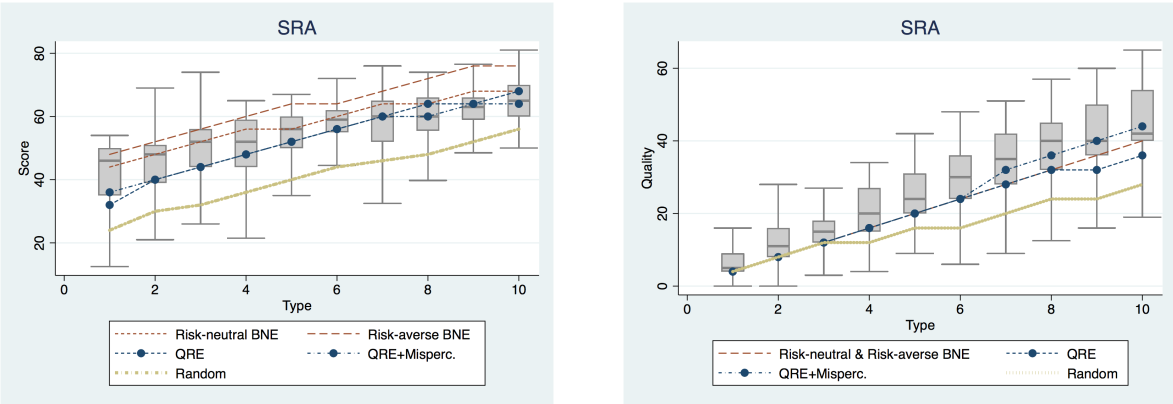
where $\ln M$ is the log-likelihood of the estimated model, $\ln M^*$ is the log-likelihood of the "clairvoyant" model, and $\ln Random$ is the log-likelihood of the random model. Table 9 reports, for each treatment, the estimated values of ϕ_M obtained from the full sample of observations.[23] Notice that ϕ_M is comparatively high in the first four treatments, *FPA*, *FQA*, *SRA2p* and *SRA2q*: With respect to a purely random choice, our model explains between 73% (in treatment *FPA*) and 83% (in treatment *FQA*) of the observed behavior. The value of ϕ_M reduces to 40% for the more complex *SRA* treatment. Although this latter value is broadly in line with the results obtained by Camerer et al. (2016) in a (richer than ours) QRE model applied to maximum value experimental auctions, we find this result worth further investigation.

Figure (2) displays, with reference to treatment *SRA* and for each type θ, the median score (2a) and the median quality (2b) predicted by our estimated baseline QRE with risk aversion and compares them with the observed bids. For ease of reference, the Bayes-Nash equilibrium (BNE) under risk neutrality and under risk aversion (with the coefficient of risk aversion set at the estimated value $r = 0.62$), and the median bid predicted by a purely random model are also reported.

Notice that the scores predicted by our baseline QRE model are always between the BNE under risk aversion and the random model. Moreover, they capture the observed behavior in a satisfactory way: In particular, the QRE model correctly predicts underbidding in the score, as observed in the experiment. Figure (2a) makes clear how the two elements of the model, risk aversion (synthesized by the parameter r) and bounded rationality (synthesized by the error parameter μ), affect bidding behavior in terms of the submitted score. As it is intuitive, relative to the BNE under risk neutrality, risk aversion leads sellers to bid more aggressively, submitting bids that produce higher scores (overbidding). On the other hand, their tendency

[23] The estimates of ϕ_M obtained for the last five periods produce similar values (see Table 10).

to make errors operates in the opposite direction, reducing the submitted scores. In fact, one can show that upward deviations from the equilibrium score generate higher payoff losses than downward deviations do: thus, if sellers make errors in a QRE fashion, scores below the equilibrium are more likely to be submitted than are scores above it. Since errors are relatively frequent – the estimated value of μ is comparatively high – the second effect prevails and the resulting behavior is underbidding.



(a) Score **(b)** Quality

Note. The gray boxes include bids within the second and the third quartile of the observed distribution; the dark-gray segment within each box is the median observation; the two vertical gray lines extend up to 1.5 times the interquartile range.

Figure 2 – Scores and quality bids in *SRA*: observed vs. predicted

The fit of our QRE model is less satisfactory when looking at the quality bid only: Figure (2b) shows that, while our data display a clear tendency by sellers to overbid in the quality component, the estimated QRE does not capture this tendency, predicting underbidding for the highest types and neither under- nor overbidding for the remaining types.[24]

6.3 QRE model augmented for a misperception of the quality trade-off

The previous subsection shows that a QRE model with risk aversion fits our experimental evidence well in four treatments over five. The fit is less satisfactory in *SRA*, where, in particular, the model does not predict the overbidding in the quality component that we observe in this treatment. A natural question is then what determines this overbidding and whether it is possible to improve our baseline QRE model in order to capture this behavior.

To investigate this question, it is useful to stress one remarkable theoretical result that holds for *SRA*: if a seller's ex-post utility depends only on her monetary earnings, then her utility-maximizing quality bid is the same regardless of the precise shape of the utility function and regardless of her belief on the opponent's play. To see this, let $v(p - C(q;\theta))$ be the utility of a seller in case of winning (notice that $v(\cdot)$ depends only on the monetary earnings), with v strictly increasing, and let $s(q,p) = g(q) - p$ be the (quasi-linear) scoring rule. The seller's

[24]In the Appendix, we report the analogues to Figure (2a) and (2b) for the remaining treatments. Differently from *SRA*, in these treatments we do not detect any systemic behavior in the data that is not predicted by the baseline QRE model.

expected utility (disregarding possible ties) is then

$$v\left(p - C(q;\theta)\right) \times \Pr\left[g(q) - p > \sigma(\cdot)\right],$$

where $\Pr\left[g(q) - p > \sigma(\cdot)\right]$ is the probability that this seller wins the auction with a bid (p, q) when the other seller's score function is $\sigma(\cdot)$. Using $s = g(q) - p$, the above expected utility can be written in terms of the choice of q and s as

$$v\left(g(q) - s - C(q;\theta)\right) \times \Pr\left[s > \sigma(\cdot)\right],$$

with first-order conditions

$$\begin{cases} v'\left(g(q) - s - C(q;\theta)\right) \times \left(g'(q) - C'(q;\theta)\right) \times \Pr\left[s > \sigma(\cdot)\right] = 0 \\ -v'\left(g(q) - s - C(q;\theta)\right) \times \Pr\left[s > \sigma(\cdot)\right] + v\left(g(q) - s - C(q;\theta)\right) \times \frac{d\Pr[s>\sigma(\cdot)]}{ds} = 0. \end{cases}$$

Since v' is strictly positive, and focusing on s such that $\Pr\left[s > \sigma(\cdot)\right] > 0$, the first first-order condition reduces

$$g'(q) = C'(q;\theta).$$

This condition simply states that the optimal quality bid is such that the marginal benefit to the score (which affects the probability of winning the auction) of a small increase in the quality bid is equal to its marginal monetary cost. This condition is unaffected by $\sigma(\cdot)$ and $v(\cdot)$: this means that, if one retains rationality (in the sense of utility-maximizing behavior), any departure from the optimal quality can be ascribed neither to the seller's possible incorrect beliefs about the opponent's play nor to a peculiar (but still dependent only on monetary earnings) type of utility function. Rather, it has to be the case that sellers have (or behave as if they had) an objective function that is not a (strictly increasing) function of their monetary earnings, but one that leads them to have a somewhat distorted perception of the trade-off between marginal benefit and marginal (monetary) cost of quality. In particular, for this distortion to lead to overbidding in quality, it has to be the case that, at the theoretically optimal quality bid, sellers must perceive that the marginal benefit (to the score) of a small increase in the quality bid exceeds its marginal cost.

In light of these considerations and in an attempt to improve our QRE model's fit to the observed behavior in *SRA*, characterized by overbidding in the quality bid (but underbidding in the score), we then amend the baseline QRE model to allow for the possibility that sellers have a distorted perception of the trade-off between marginal benefit and marginal cost of quality. In particular, we hypothesize that the perceived weight attached to quality in the scoring rule is $2 \times w_s$, and that the perceived cost of quality q for a type-θ seller is $w_c \times C(q;\theta)$. The full rationality hypothesis corresponds to $w_s = w_c = 1$. We introduce two more parameters, not just one, because, in equilibrium, one parameter only can generate overbidding in the quality bid but not, simultaneously, underbidding in the score, or vice versa.[25]

The estimates for the QRE model augmented with these two parameters are displayed in Table 11. Observe that the parameter w_s, which captures the perceived marginal contribution of the submitted quality to the score, is active only in those treatments in which there is indeed a score, namely *SRA*, *SRA2p*, and *SRA2q*. On the other hand, w_c, the parameter that measures the perceived cost of quality, is absent in *FPA*, where quality cannot be chosen, and in *SRA2q*, where no marginal change from the admissible qualities is possible.

[25]This simple result is shown in the Appendix.

Table 11 – Estimates from the augmented QRE model, full sample

	FPA	*FQA*	*SRA2q*	*SRA2p*	*SRA*
r	0.68	0.68	0.68	0.66	0.57
μ	0.78	0.42	0.91	1.05	0.94
w_s	-	-	1.00	0.98	0.77
w_c	-	1.00	-	0.96	0.67
η	0.389	0.422	0.995	1.249	1.266
	(0.029)	(0.138)	(0.104)	(0.194)	(0.053)
LL	−355.82	−261.97	−438.21	−309.10	−1284.66
ϕ_M	0.73	0.83	0.74	0.76	0.46

Note. This table reports, for each treatment, maximum likelihood estimates of the four parameters of the augmented QRE model: r (the Arrow-Pratt coefficient of relative risk aversion), μ (the error parameter), w_s (the perceived weight of quality in the score), and w_c (the perceived cost of quality). η is the relative utility loss predicted by the estimated model (standard error in parenthesis), LL is the value of the log-likelihood function, ϕ_M is a comparable measure of the goodness of fit.

Notice, first, that the four treatments for which the baseline QRE model was already satisfactory are almost unaffected by the introduction of these two additional parameters: The estimates of r, μ, and η are essentially unchanged, and the estimates of the additional parameters w_s and w_c are close (if not equal) to 1. Therefore, in these treatments, sellers seem to have full cognitive control over the effects of their quality choices on the score and on their costs.

Things are different for *SRA*, the most complex treatment and the one for which the baseline QRE model fitted the worst. For *SRA*, we obtain estimates for the two additional parameters that are significantly lower than one: Sellers seem to underestimate the impact of quality on both the submitted score and on their monetary costs. In particular, as expected, we find that the estimated value of w_s is greater than the estimated value of w_c: In the trade-off between marginal benefit and marginal cost of quality, the former is overestimated relative to the second, leading to the observed overbidding in the quality component. In fact, as Figure (2b) shows, the augmented model is now able to predict the observed overbidding in the quality bid while still producing underbidding in the score. That the introduction of the parameters w_s and w_c is essentially immaterial for *FPA*, *FQA*, *SRA2p* and *SRA2q*, but significantly improves the fit of the model for *SRA* is also confirmed by looking at the value of the log-likelihood (which increases by +91.55 for *SRA*) and the value of the statistic ϕ_M (which increases from 0.40 to 0.46 for *SRA*).[26]

Interestingly, as we pass from the baseline to the augmented model, for treatment *SRA* we obtain a significant reduction both in the estimated error parameter μ (from 1.16 to 0.94) and in the (more comparable) measure of departure from rationality η (from 1.535 to 1.266). This means that, conditional on their misperception, subjects make less errors than in the baseline model: part of what the baseline QRE model explained simply as an error turns out to be better described as an incorrect perception by subjects of the marginal effects of their

[26] To compare the relative fit of the baseline and augmented models, we also performed a Bayesian Information Criterion test (BIC, Schwarz (1978)) and a Vuong's closeness test (Vuong, 1989). Both tests confirm that the augmented QRE significantly improves fit in *SRA* but not in *FPA*, *FQA*, *SRA2p* or *SRA2q*.

quality choices on their payoffs.[27] Finally, the inclusion of w_s and w_c in the model does not alter the ranking of η across treatments:[28] the highest value is still that recorded by *SRA*, although the difference from that of *SRA2p* is small. Hence, our hypothesis of a positive relationship between the complexity of the auction mechanism and the subjects' likelihood of making suboptimal bids is still supported by data. However, the results from the augmented QRE model suggest that, when complexity is maximal (i.e., in *SRA*), a specific form of non-rationality is triggered: sellers seem to have a distorted perception of the trade-off between the marginal benefit and the marginal cost associated with their quality choice.

7 Concluding remarks

In this paper, we studied the problem of a buyer who wants to procure a good or service for which both the price and a non-price dimension (that we refer to as quality) matter. We implemented in the lab five auction mechanisms to award the contract. These mechanisms differ in their intrinsic trade-off between theoretical performance and bidding complexity. Specifically, in the simplest mechanisms, *FPA* and *FQA*, sellers bid on one dimension only (price or quality, respectively), while the other dimension was set (optimally) by the experimenter; in the most complex mechanism, *SRA*, sellers had to submit a price-quality offer and their bidimensional bids were (linearly) combined into scores that determined the winner; finally, we implemented two treatments of intermediate complexity, where sellers competed in a scoring auction like *SRA*, but where the choice set on one dimension (price in *SRA2p*, quality in *SRA2q*) was only binary. Theoretically, in equilibrium, *SRA* (the most complex mechanism) should yield the highest level of social welfare and buyer's surplus, while the unidimensional treatments should perform the worst, with *FPA* significantly worse than *FQA*.

Our experimental results showed that the theoretical ranking across treatments is partially upset in the lab: While *FPA* performs worst, as predicted, *FQA* overperforms and the three scoring auctions underperform. As a result, we found no significant differences among *FQA*, *SRA2p*, *SRA2q* and *SRA*, both in terms of social welfare and buyer's surplus (if anything, *FQA* performs better than the scoring auctions in terms of buyer's surplus).

The analysis of bidding behavior shed light on these findings: in all treatments (except *FPA* where quality was fixed), the submitted quality is higher than predicted and closer to the efficient level. This result – which, by itself, increases social and buyer's surplus – explains the overperformance of *FQA*. However, in the scoring auctions this positive effect is accompanied by two countervailing effects: on the one hand, a significant fraction of contracts is inefficiently allocated to the high cost bidder; on the other hand, a higher-than-predicted level of submitted quality tends to be accompanied by an even stronger upward adjustment

[27] Indeed, if we entirely shut down the stochastic choice component of the augmented model (setting $\mu = 0$), the overbidding in quality that the model predicts would be similar to the average choice observed in the experiment: With $w_s = 0.77$ and $w_c = 0.67$, the equilibrium quality offer is equal to $q(\theta) = 4.597 \cdot \theta$; regressing q on θ using the experimental data, we obtain $\widehat{q} = 2.825 + 4.417 \cdot \theta$. Both parameters of this regression are highly significant (p-value < 0.001).

[28] In the computation of η for the augmented model, we used the *perceived* expected utility of the subjects, that is,

$$\frac{1}{1-r}\left[p - w_c C(q;\theta)\right]^{1-r} \times \Pr\left[2w_s q - p > \sigma(\cdot)\right],$$

with w_s and w_c set at their estimated values. Hence, η has to be interpreted here as a measure of the departure from rationality, net of the distortion that is due to the subjects' misperception of the marginal effect of quality on their expected payoffs.

in prices, which eventually penalizes the buyer. We guessed that these negative effects may be somewhat related to the higher degree of complexity of the scoring auctions. In fact, by estimating a structural QRE model of bidding behavior, we found strong evidence in favor of a positive relationship between complexity of the mechanism and bidders' proneness to make suboptimal bids. This tendency of bidding away from the best response generates more noisy behavior which undermines the efficiency of the allocation and produces, on average, more conservative bidding (a lower than optimal quality-price ratio). Moreover, in *SRA*, where complexity is maximal, we also detected a distorted perception by bidders of the trade-off between marginal benefit and marginal cost of quality.

With the usual caveats regarding the applicability of experimental findings to real-world settings, we believe that our results may complement ,with clean and controlled evidence, what the empirical literature already showed regarding scoring auctions. In this respect, it is comforting that our results are consistent with Lewis and Bajary (2011)'s empirical findings. Their framework and our experimental setting have many similarities: only one quality attribute, a linear scoring rule in the scoring auction, a considerable weight attached to the quality component both in the scoring rule and in the buyer's objective function, convex cost functions for quality provision that do not cross for different bidders' types.[29] Our results regarding the comparison between *FPA* and *SRA* – the auction formats that are more popular in real world procurement – perfectly match theirs: Although the buyer pays a (slightly) lower price in *FPA*, the increase in quality obtained in *SRA* is such that the net effect on the buyer's surplus is positive and substantial. Moreover, the particularly noisy bidding behavior that we detect in *SRA* – which is at the heart of the underperformance of this format with respect to theory and that we attribute to the complexity of the bidding task – is in line with their findings: in fact, when we regress the quality bids in *SRA* on the cost parameter and on bidder fixed effects, we find that around 28% of the overall variance remains unexplained. This number is remarkably close to what we obtained by replicating, this time on Lewis and Bajari's dataset, a similar regression of the quality bid, including contract and bidder fixed effects to account for other time-invariant unobserved characteristics: with their data, 30% of the overall variance of quality choices remains unexplained.[30] The authors themselves recognize that "*bidder heterogeneity accounts for more of the variance than contract heterogeneity*" (Lewis and Bajary (2011), p. 1201), suggesting that some relevant behavioral effects may be at work.

Hence, our paper suggests that the market designer should carefully weigh the benefits that competition in multiple bidding dimensions is expected to yield, with the potential distortions triggered by the complexity of the mechanism adopted. This trade-off is clearly exemplified by our results regarding *SRA*: on the one hand, this mechanism is theoretically preferable as the scoring rule represents a detailed translation of the buyer's true preferences into an awarding rule; on the other hand, it places suppliers into a more complex strategic environment, and this leads them to make more frequent errors as if the strategic complexity generated some lost in translation of the buyer's preferences for quality and price into the suppliers' bids.

Clearly, the optimal solution to this trade-off should be evaluated case by case. In our experimental setting, the non-price attribute of the object was particularly important, both

[29]The buyer's objective function and the cost functions resulted from estimations in Lewis and Bajary (2011), were set exogenously in our experiment.

[30]The results of these regressions are reported and discussed in the Appendix. We are indebted to Gregory Lewis and Patrick Bajari for sharing their data and codes.

on the demand side (quality had a relevant weight in the buyer's objective function) and on the supply side (costs were very sensitive to the quality provided). In such circumstances, our paper suggests that letting bidders compete also on these non-price attributes is certainly preferable to using a price-only auctions, even though the latter is certainly a more straightforward mechanism. Nevertheless, our paper also suggests that the market designer should explore other less common awarding procedures to find the best solution to the trade-off between performance and complexity at the bidding stage. In this respect, our experiment shows that a simple mechanism like the quality-only auction (with an optimally predetermined price), despite its predicted inferiority in terms of social welfare and buyer's utility, actually performs at least as well as a genuine scoring auction. Further research is required to assess how the solution to this trade-off changes as the relative importance of the price and the non-price elements change.

Finally, note that in our experimental setting there are no reputational concerns. An interesting research question to be further investigated in the lab is how the mechanisms we have studied can perform differently when bidders have an incentive to supply high quality in order get future contracts.

References

Albano, G. L., A. Cipollone, R. Di Paolo, G. Ponti, and M. Sparro (2018). Scoring rules in experimental procurement. *CESIEG Working papers 10*(2018), 1–26.

Asker, J. and E. Cantillon (2008). Properties of scoring auctions. *RAND Journal of Economics 39*(1), 69–85.

Asker, J. and E. Cantillon (2010). Procurement when price and quality matter. *Rand Journal of Economics 41*(1), 1–34.

Bichler, M. (2000). An experimental analysis of multi-attribute auctions. *Decision Support Systems 29*(3), 249–268.

Camerer, C., S. Nunnari, and T. R. Palfrey (2016). Quantal response and nonequilibrium beliefs explain overbidding in maximum-value auctions. *Games and Economic Behavior 98*, 243–263.

Cameron, L. J. (2000). Limiting buyer discretion: Effects on performance and price in long-term contracts. *American Economic Review 90*(1), 265–281.

Che, Y. K. (1993). Design competition through multidimensional auctions. *RAND Journal of Economics 24*(4), 668–680.

Chen-Ritzo, C. H., T. P. Harrison, A. M. Kwasnica, and D. J. Thomas (2005). Better, faster, cheaper: An experimental analysis of a multiattribute reverse auction mechanism with restricted information feedback. *Management Science 51*(12), 1753–1762.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics 10*(2), 171–178.

Goeree, J. K., C. A. Holt, and T. R. Palfrey (2002). Quantal response equilibrium and overbidding in private-value auctions. *Journal of Economic Theory 104*(1), 247–272.

Gupta, D., E. M. Snir, and Y. Chen (2015). Contractors' and agency decisions and policy implications in a+b bidding. *Production and Operations Management 24*(1), 159–177.

Hyytinen, A., S. Lundeberg, and O. Toivanen (2018). Design of public procurement auctions: Evidence from cleaning contracts. *Rand Journal of Economics 49*(2), 398–426.

Kagel, J. H. (1995). Auctions: A survey of experimental research. *Handbook of Experimental Economics Princeton University Press*, 501–586.

Kagel, J. H. and D. Levin (2011). Auctions: A survey of experimental research, 1995–2010. *Handbook of Experimental Economics, Vol. 2 Princeton University Press*, 563–637.

Kagel, J. H., Y. Lien, and P. Milgrom (2010). Ascending prices and package bidding: A theoretical and experimental analysis. *American Economic Journal: Microeconomics 2*(3), 160–85.

Kwasnica, A., J. O. Ledyard, D. Porter, and C. DeMartini (2005). A new and improved design for multiobject iterative auctions. *Management Science 51*(3), 419–434.

Kwasnica, A. M. and K. Sherstyuk (2013). Multiunit auctions. *Journal of Economic Surveys 27*(3), 461–490.

Lewis, G. and P. Bajary (2011). Procurement contracting with time incentives: Theory and evidence. *Quarterly Journal of Economics 126*(3), 1173–1211.

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

Scheffel, T., G. Ziegler, and M. Bichler (2012). On the impact of package selection in combinatorial auctions: An experimental study in the context of spectrum auction design. *Experimental Economics 15*(4), 667–692.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics 6*(2), 461–464.

Strecker, S. (2010). Information revelation in multiattribute english auctions: A laboratory study. *Decision Support Systems 49*(3), 272–280.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*(2), 307–333.

A Experimental Instructions

Instructions were originally written in Italian. The following instructions refer to the SRA treatment. The only difference between these instructions and those used in FQA and FPA is that, the offer in the latter was made by choosing one dimension only, respectively quality and price, and therefore no reference to the scoring rule was required. The only difference between these instructions and those used in SRA2q and SRA2p is that, the domain of one dimension of the offer in the latter contained only two possible levels, 9 and 40 in in SRA2q, 12 and 65 for the price in SRA2p, respectively. Instructions used in FQA, FPA, SRA2q, and SRA2p are available upon request from the authors.

Instructions

Welcome! Thanks for participating in this experiment. By following these rules carefully, you can earn an amount of money that will be paid in cash at the end of the experiment.

There are 24 subjects participating in this experiment. Both the identities and the final payments of the subjects will remain anonymous throughout the experiment. During the experiment you are not allowed to communicate with the other participants. If you have questions, raise your hand and one of the assistants will come to your seat and assist you. The following instructions are the same for all the participants.

General rules

The experiment will consist of 15 periods and, in every period, you will face the same economic situation that is described in what follows.

At the beginning of each period, you will be randomly and anonymously assigned to a new group of two subjects. This means that the composition of the group will change in every period and you will never interact with the same participant in two consecutive periods.

During the experiment, your earnings will be expressed in tokens. At the beginning of the experiment you will receive an initial endowment of 20 tokens. The total number of tokens earned will be the sum of the total tokens earned in the 15 periods of the experiment, plus the initial endowment. At the end of the experiment, the total number of tokens will be exchanged in euro at the following exchange rate: 7 tokens = 1 euro.

Your experimental task

In each of the 15 periods of the experiment, you and the other subject in your group will play the role of two sellers who compete with each other in an auction for the sale of an object to a hypothetical buyer.

During each auction, you and the other seller in your group will anonymously and simultaneously choose your offers to submit to the buyer. This means that, when making your choice you will not receive any information about the offer submitted by the other seller in your group.

Your offer

Your offer consists of two choices: (1) the level of quality of the object and (2) the price of the object. As for the level of quality, you will choose an integer number included between 0 (the lowest quality level) and 70 (the highest quality level). As for the price, you will choose an integer number included between 0 and the buyer's maximum willingness to pay for the object.

The buyer's maximum willingness to pay for the object increases in the quality level of the object, according to the following expression:

$$maximum\ willingness\ to\ pay = level\ of\ quality + 50$$

Given your choices of price and quality, the computer will assign a score to your offer by using the following expression:

$$score = 50 + 2 * level\ of\ quality - price$$

The score assigned to your offer (a) increases in your quality choice, and (b) decreases in your price choice. Note that the score assigned to your offer takes a minimum value of 0 when you choose a quality level of 0 and a price of 50 (i.e. equal to the buyer's maximum willingness to pay for the object as determined by the quality level).

Auction result and earnings

Given your quality and price choices and those made by the other seller in your group, the computer will compare the scores assigned to the two offers. The seller whose offer has been assigned the highest score wins the auction. Ties will be randomly broken by the computer.

The loser of the auction does not obtain any earning in the period. Conversely, the earnings of the winner of the auction are given by the difference between his/her price and the cost of the object, that is:

$$Winner's\ earning = price - cost\ of\ the\ object$$

For the winner of the auction, the cost of the object depends on two elements: (a) the quality choice and (b) a random parameter. In particular, the computer will use the following expression to determine the costs incurred by the winner of the auction:

$$cost\ of\ the\ object = \frac{1}{4 * random\ parameter} * (level of quality)^2$$

where the random parameter is an integer number included between 1 and 10.

Note that: (a) the higher is the value of the random parameter, the lower is the cost incurred by the winner, for each level of quality; (b) the cost of the object increases more than proportionately with respect to the level of quality.

The value of your random parameter is selected by the computer in each period. In particular, the computer will randomly select an integer between 1 and 10 with equal probability. You will be informed about your random parameter before choosing your offer. The information about the random parameter is private, meaning that each seller will only observe her/his own parameter, while she/he will not receive any information about the parameter of the other seller in the group. This means that only you will know your random parameter, and that you will not receive any information about the random parameter extracted from the computer for the other seller. Finally, the random parameter assigned to you by the computer in one period does not depend on that of the other seller in your group, nor on those assigned in previous periods.

At the end of each auction, the computer will inform you and the other seller in your group about the outcome of the auction, the price and the quality level chosen by the winner, his/her score and your earnings, expressed in tokens.

For your convenience, the following table shows (a) the buyer's maximum willingness to pay, depending on the quality choice; (b) the costs incurred by the winner of the auction, depending on her/his random parameter and the quality choice. We have highlighted in grey the quality levels for which the costs for the sellers are higher than the buyer's maximum willingness to pay.

Quality	Max WTP	Costs as a function of the random parameter (r.p.) and of the quality choice									
		r.p. = 1	r.p. = 2	r.p. = 3	r.p. = 4	r.p. = 5	r.p. = 6	r.p. = 7	r.p. = 8	r.p. = 9	r.p. = 10
0	50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	51	0.25	0.13	0.08	0.06	0.05	0.04	0.04	0.03	0.03	0.03
2	52	1.00	0.50	0.33	0.25	0.20	0.17	0.14	0.13	0.11	0.10
3	53	2.25	1.13	0.75	0.56	0.45	0.38	0.32	0.28	0.25	0.23
4	54	4.00	2.00	1.33	1.00	0.80	0.67	0.57	0.50	0.44	0.40
5	55	6.25	3.13	2.08	1.56	1.25	1.04	0.89	0.78	0.69	0.63
6	56	9.00	4.50	3.00	2.25	1.80	1.50	1.29	1.13	1.00	0.90
7	57	12.25	6.13	4.08	3.06	2.45	2.04	1.75	1.53	1.36	1.23
8	58	16.00	8.00	5.33	4.00	3.20	2.67	2.29	2.00	1.78	1.60
9	59	20.25	10.13	6.75	5.06	4.05	3.38	2.89	2.53	2.25	2.03
10	60	25.00	12.50	8.33	6.25	5.00	4.17	3.57	3.13	2.78	2.50
11	61	30.25	15.13	10.08	7.56	6.05	5.04	4.32	3.78	3.36	3.03
12	62	36.00	18.00	12.00	9.00	7.20	6.00	5.14	4.50	4.00	3.60
13	63	42.25	21.13	14.08	10.56	8.45	7.04	6.04	5.28	4.69	4.23
14	64	49.00	24.50	16.33	12.25	9.80	8.17	7.00	6.13	5.44	4.90
15	65	56.25	28.13	18.75	14.06	11.25	9.38	8.04	7.03	6.25	5.63
16	66	64.00	32.00	21.33	16.00	12.80	10.67	9.14	8.00	7.11	6.40
17	67	72.25	36.13	24.08	18.06	14.45	12.04	10.32	9.03	8.03	7.23
18	68	81.00	40.50	27.00	20.25	16.20	13.50	11.57	10.13	9.00	8.10
19	69	90.25	45.13	30.08	22.56	18.05	15.04	12.89	11.28	10.03	9.03
20	70	100.00	50.00	33.33	25.00	20.00	16.67	14.29	12.50	11.11	10.00
21	71	110.25	55.13	36.75	27.56	22.05	18.38	15.75	13.78	12.25	11.03
22	72	121.00	60.50	40.33	30.25	24.20	20.17	17.29	15.13	13.44	12.10
23	73	132.25	66.13	44.08	33.06	26.45	22.04	18.89	16.53	14.69	13.23
24	74	144.00	72.00	48.00	36.00	28.80	24.00	20.57	18.00	16.00	14.40
25	75	156.25	78.13	52.08	39.06	31.25	26.04	22.32	19.53	17.36	15.63
26	76	169.00	84.50	56.33	42.25	33.80	28.17	24.14	21.13	18.78	16.90
27	77	182.25	91.13	60.75	45.56	36.45	30.38	26.04	22.78	20.25	18.23
28	78	196.00	98.00	65.33	49.00	39.20	32.67	28.00	24.50	21.78	19.60
29	79	210.25	105.13	70.08	52.56	42.05	35.04	30.04	26.28	23.36	21.03
30	80	225.00	112.50	75.00	56.25	45.00	37.50	32.14	28.13	25.00	22.50
31	81	240.25	120.13	80.08	60.06	48.05	40.04	34.32	30.03	26.69	24.03
32	82	256.00	128.00	85.33	64.00	51.20	42.67	36.57	32.00	28.44	25.60
33	83	272.25	136.13	90.75	68.06	54.45	45.38	38.89	34.03	30.25	27.23
34	84	289.00	144.50	96.33	72.25	57.80	48.17	41.29	36.13	32.11	28.90
35	85	306.25	153.13	102.08	76.56	61.25	51.04	43.75	38.28	34.03	30.63
36	86	324.00	162.00	108.00	81.00	64.80	54.00	46.29	40.50	36.00	32.40
37	87	342.25	171.13	114.08	85.56	68.45	57.04	48.89	42.78	38.03	34.23
38	88	361.00	180.50	120.33	90.25	72.20	60.17	51.57	45.13	40.11	36.10
39	89	380.25	190.13	126.75	95.06	76.05	63.38	54.32	47.53	42.25	38.03
40	90	400.00	200.00	133.33	100.00	80.00	66.67	57.14	50.00	44.44	40.00
41	91	420.25	210.13	140.08	105.06	84.05	70.04	60.04	52.53	46.69	42.03
42	92	441.00	220.50	147.00	110.25	88.20	73.50	63.00	55.13	49.00	44.10
43	93	462.25	231.13	154.08	115.56	92.45	77.04	66.04	57.78	51.36	46.23
44	94	484.00	242.00	161.33	121.00	96.80	80.67	69.14	60.50	53.78	48.40
45	95	506.25	253.13	168.75	126.56	101.25	84.38	72.32	63.28	56.25	50.63
46	96	529.00	264.50	176.33	132.25	105.80	88.17	75.57	66.13	58.78	52.90
47	97	552.25	276.13	184.08	138.06	110.45	92.04	78.89	69.03	61.36	55.23
48	98	576.00	288.00	192.00	144.00	115.20	96.00	82.29	72.00	64.00	57.60
49	99	600.25	300.13	200.08	150.06	120.05	100.04	85.75	75.03	66.69	60.03
50	100	625.00	312.50	208.33	156.25	125.00	104.17	89.29	78.13	69.44	62.50
51	101	650.25	325.13	216.75	162.56	130.05	108.38	92.89	81.28	72.25	65.03
52	102	676.00	338.00	225.33	169.00	135.20	112.67	96.57	84.50	75.11	67.60
53	103	702.25	351.13	234.08	175.56	140.45	117.04	100.32	87.78	78.03	70.23
54	104	729.00	364.50	243.00	182.25	145.80	121.50	104.14	91.13	81.00	72.90
55	105	756.25	378.13	252.08	189.06	151.25	126.04	108.04	94.53	84.03	75.63
56	106	784.00	392.00	261.33	196.00	156.80	130.67	112.00	98.00	87.11	78.40
57	107	812.25	406.13	270.75	203.06	162.45	135.38	116.04	101.53	90.25	81.23
58	108	841.00	420.50	280.33	210.25	168.20	140.17	120.14	105.13	93.44	84.10
59	109	870.25	435.13	290.08	217.56	174.05	145.04	124.32	108.78	96.69	87.03
60	110	900.00	450.00	300.00	225.00	180.00	150.00	128.57	112.50	100.00	90.00
61	111	930.25	465.13	310.08	232.56	186.05	155.04	132.89	116.28	103.36	93.03
62	112	961.00	480.50	320.33	240.25	192.20	160.17	137.29	120.13	106.78	96.10
63	113	992.25	496.13	330.75	248.06	198.45	165.38	141.75	124.03	110.25	99.23
64	114	1024.00	512.00	341.33	256.00	204.80	170.67	146.29	128.00	113.78	102.40
65	115	1056.25	528.13	352.08	264.06	211.25	176.04	150.89	132.03	117.36	105.63
66	116	1089.00	544.50	363.00	272.25	217.80	181.50	155.57	136.13	121.00	108.90
67	117	1122.25	561.13	374.08	280.56	224.45	187.04	160.32	140.28	124.69	112.23
68	118	1156.00	578.00	385.33	289.00	231.20	192.67	165.14	144.50	128.44	115.60
69	119	1190.25	595.13	396.75	297.56	238.05	198.38	170.04	148.78	132.25	119.03
70	120	1225.00	612.50	408.33	306.25	245.00	204.17	175.00	153.13	136.11	122.50

Summarizing...

The rules used in each period to determine your earnings are summarized below.

1. At the beginning of each period, you will be randomly and anonymously assigned to a new group of two subjects.

2. You and the other subject in your group will play the role of two sellers who compete with each other in an auction for the sale of an object to a hypothetical buyer.

3. You and the other seller in your group will anonymously and simultaneously choose your offers to submit to the buyer. In particular, you will have to make two choices: (1) the level of quality of the object and (2) the price of the object.

4. The maximum willingness to pay of the buyer for the object increases as the quality level of the object increases, according to the following expression:

$$maximum\, willingness\, to\, pay = level\, of\, quality + 50$$

5. According to the price and quality choice you make, the computer will assign a score to your sale proposal using the following expression:

$$score = 50 + 2 * level\, of\, quality - price$$

6. The winner of the auction is the seller whose sale proposal has obtained the highest score. In the event of a draw, the computer randomly, and with equal probability, selects the winner.

7. The seller who did not win the auction does not obtain any token in the period. Conversely, the winner of the auction makes an earning, in tokens, according to the following expression:

$$Winner's\, earning = price - cost\, of\, the\, object$$

8. The cost of the object, for the winner of the auction, depends on two elements: (a) the level of quality and, (b), a random parameter. In particular, the computer will use the following expression to determine the costs incurred by the winner of the auction:

$$cost\, of\, the\, object = \frac{1}{4 * random\, parameter} * (level of quality)^2$$

   where the random parameter is an integer number included between 1 and 10, randomly chosen by the computer at the beginning of each period and notified to the seller before he makes his sale offer. Note that: (a) the higher is the value of the random parameter, the lower is the cost incurred by the winner, for each level of quality; (b) the cost of the object increases more than proportionately with respect to the level of quality.

9. At the end of each auction, the computer will show the sellers the outcome of the auction, the price and the level of quality chosen by the winner, his score and the earnings, expressed in tokens.

B Theory and predictions: derivation of the results

The (symmetric) equilibrium bidding functions in the five treatments[31] are the following:

- *FPA*:

$$\alpha_1(\theta) = \begin{cases} 64 & \text{if} \quad \theta = 1 \\ 64(\theta-1)^{-1}\ln\theta & \text{if} \quad \theta \in (1, 10] \end{cases} ; \tag{6}$$

- *FQA*:

$$\beta_2(\theta) = 8\sqrt{\theta+1}; \tag{7}$$

- *SRA*:

$$[\alpha_3(\theta); \beta_3(\theta)] = [2(3\theta-1); 4\theta]\,; \tag{8}$$

- *SRA2p*:

$$[\alpha_4(\theta); \beta_4(\theta)] = \begin{cases} \left[12; 2\sqrt{6(\theta+1)}\right] & \text{if} \quad \theta \in [1, \hat{\theta}) \\ \left[65; \sqrt{\frac{130(\theta^2-\hat{\theta}^2)+(\hat{\theta}-1)\hat{b}^2}{\theta-1}}\right] & \text{if} \quad \theta \in [\hat{\theta}, 10] \end{cases} , \tag{9}$$

  where $\hat{\theta} = \left(65 + 30\sqrt{2}\right)/16$ and $\hat{b} = \left(59 + 15\sqrt{2}\right)/4$;

- *SRA2q*:

$$[\alpha_5(\theta); \beta_5(\theta)] = \begin{cases} [81/4; 9] & \text{if} \quad \theta = 1 \\ \left[81[4(\theta-1)]^{-1}\ln\theta; 9\right] & \text{if} \quad \theta \in (1, \hat{\theta}) \\ \left[\frac{400(\ln\theta - \ln\hat{\theta})+(\hat{\theta}-1)\hat{b}}{\theta-1}; 40\right] & \text{if} \quad \theta \in [\hat{\theta}, 10] \end{cases} , \tag{10}$$

  where $\hat{\theta} = 49/8$ and $\hat{b} = 62 + (162/41) \times \ln(49/8)$.

The corresponding proofs, treatment by treatment, are reported below. For each treatment, we also compute the expected buyer's utility and the expected total welfare.

B.1 *FPA*

Equilibrium. Let $\bar{q}$ denote the quality imposed by the buyer. With quality set exogenously, this is a standard (reverse) first-price auction with independent private values (costs). We then obtain the standard symmetric equilibrium bidding function[32]

$$\alpha_1(\theta) = \frac{1}{F(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} C(\bar{q}; \theta) f(\theta) d\theta,$$

where $F(\cdot)$ is the distribution of types and $f = F'$. With $C(q; \theta) = q^2/(4\theta)$, and under our assumptions on the type distribution $F(\cdot)$, it obtains:

$$\alpha_1(\theta) = \begin{cases} \frac{\bar{q}^2}{4} & \text{if} \quad \theta = 1 \\ \frac{\bar{q}^2}{4}\frac{\ln\theta}{\theta-1} & \text{if} \quad \theta \in (1, 10] \end{cases} .$$

[31] We use $\alpha_i(\theta)$ to denote a bidding strategy in the price dimension, $\beta_i(\theta)$ for the quality dimension; moreover, we use the index 1 for *FPA*, 2 for *FQA*, 3 for *SRA*, 4 for *SRA2p*, 5 for *SRA2q*. Clearly, $\sigma_i(\theta) = 50+2\beta_i(\theta)-\alpha_i(\theta)$ will be the corresponding score. Remind that the quality (price) is fixed in *FPA* (*FQA*).

[32] See, e.g., Krishna (2009).

Replacing $\bar{q}$ with 16 (the parameter used in the experiment), we get (6).

Buyer's utility. The expected utility of the buyer in equilibrium (for generic $\bar{q}$) is

$$U_B = \frac{20}{7}\bar{q} - E[p^w]$$

where $E[p^w]$ – the expected winning price – is

$$E[p^w] = \int_{\underline{\theta}}^{\overline{\theta}} \alpha_1(\theta) 2F(\theta) f(\theta) d\theta,$$

where $2F(\theta)f(\theta)$ is the density of the first-order statistic. It follows:

$$U_B = \frac{20}{7}\bar{q} - \frac{10\ln 10 - 9}{162}\bar{q}^2,$$

which is maximized for $\bar{q} = 1620/[7(10\ln 10 - 9)] \approx 16.5$. For $\bar{q} = 16$ – the (integer) parameter that we used in the experiment – it is $U_B(\bar{q} = 16) \approx 23.5$.

Total welfare. The expected total welfare in equilibrium is

$$W = \frac{20}{7}\bar{q} - E[C^w]$$

where $E[C^w]$ – the expected cost of the winner – is

$$E[C^w] = \int_{\underline{\theta}}^{\overline{\theta}} \frac{\bar{q}^2}{4\theta} 2F(\theta) f(\theta) d\theta.$$

It follows

$$W = \frac{20}{7}\bar{q} - \frac{\bar{q}^2}{162}(9 - \ln 10).$$

For $\bar{q} = 16$ – the parameter that we used in the experiment – it is $W(\bar{q} = 16) \approx 35.1$.

B.2 *FQA*

Equilibrium. Let $\bar{p}$ denote the price imposed by the buyer. The symmetric equilibrium bidding function is

$$\beta_2(\theta) = \sqrt{2\bar{p}(\theta + 1)},$$

which, for $\bar{p} = 32$ (the parameter used in the experiment) yields (7). To see that this is indeed an equilibrium, suppose that seller j bids according to $\beta_2(\cdot)$ and consider seller i, type θ. If this seller bids $\beta_2(\theta)$ (her equilibrium bid), her expected utility is

$$U_i(\beta_2(\theta); \theta) = \left[\bar{p} - \frac{\beta_2(\theta)^2}{4\theta}\right] \times F(\theta) = \left[\bar{p} - \frac{2\bar{p}(\theta + 1)}{4\theta}\right] \times \frac{\theta - 1}{9} = \frac{\bar{p}(\theta - 1)^2}{18\theta}.$$

If, on the other hand, this seller makes a different bid $\beta_2(z)$, $z \neq \theta$, her expected utility is

$$U_i(\beta_2(z); \theta) = \left[\bar{p} - \frac{\beta_2(z)^2}{4\theta}\right] \times F(z) = \left[\bar{p} - \frac{2\bar{p}(z + 1)}{4\theta}\right] \times \frac{z - 1}{9} = \frac{\bar{p}(z - 1)(2\theta - z - 1)}{18\theta}.$$

We have

$$U_i(\beta_2(\theta);\theta) - \pi_i(\beta_2(z);\theta) = \frac{\bar{p}}{18\theta}\left[(\theta-1)^2 - (z-1)(2\theta - z - 1)\right] = \frac{\bar{p}}{18\theta}\left[\theta - z\right]^2 > 0.$$

This shows that, for a type-θ seller, it is optimal to bid $\beta_2(\theta)$ when the other seller bids according to $\beta_2(\cdot)$, i.e. $\beta_2(\theta)$ is indeed a symmetric equilibrium bidding function.

Buyer's utility. The expected utility of the buyer in equilibrium (for generic $\bar{p}$) is

$$U_B = \frac{20}{7}E[q^w] - \bar{p}$$

where $E[q^w]$ – the expected winning quality – is

$$E[q^w] = \int_{\underline{\theta}}^{\bar{\theta}} \beta_2(\theta) 2F(\theta) f(\theta) d\theta.$$

It follows:

$$U_B = K\sqrt{2\bar{p}} - \bar{p},$$

where $K = \frac{80\,(253\sqrt{11}+8\sqrt{2})}{8505}$. The above expression is maximized for $\bar{p} = K^2/2 \approx 32.0$. For $\bar{p} = 32$ – the (integer) parameter that we used in the experiment – it is $U_B(\bar{p} = 32) \approx 32.0$.

Total welfare. The expected total welfare in equilibrium is

$$W = \frac{20}{7}E[q^w] - E[C^w]$$

where $E[C^w]$ – the expected cost of the winner – is

$$E[C^w] = \int_{\underline{\theta}}^{\bar{\theta}} \frac{\beta_2(\theta)^2}{4\theta} 2F(\theta) f(\theta) d\theta.$$

It follows:

$$W = K\sqrt{2\bar{p}} - \frac{(99 - 2\ln 10)}{162}\bar{p}.$$

For $\bar{p} = 32$ – the parameter that we used in the experiment – it is $W(\bar{p} = 32) \approx 45.3$.

B.3 *SRA*

Equilibrium. In *SRA*, sellers choose p and q that jointly determine the score according to $s = aq - p$ (we keep the quality weight a general for the moment and disregard any additive constant, which is immaterial). Notice that, for given q, there is a one-to-one relation between s and p. Therefore, we can equivalently think of a situation in which sellers submit s and q, the seller with the highest score wins and is paid $p = aq - s$. It is extremely useful to reformulate the problem in this way. In fact, in terms of the choice of s and q, the expected utility of a generic seller i, type θ, becomes

$$U_i(s, q; \theta) = [aq - s - C(q;\theta)] \times \mathrm{PW}(s),$$

where $\mathrm{PW}(s)$ is the probability of winning the auction with a score equal to s. The crucial point is that this probability depends only on s, not on q: this implies that the optimal choice of q is totally independent from the choice of s, as well as from the bidding strategy of the other seller.[33] In fact, it is immediate to verify that, regardless of the choice of s (as long as $\mathrm{PW}(s) \neq 0$), the optimal value of q is the one for which $a - C'(q;\theta) = 0$ (where $C'(q;\theta) = \partial C(q;\theta)/\partial q$). Let this optimal value, which in general depends on θ, be denoted by $q^*(\theta)$. It is easy to verify that, with $C(q;\theta) = q^2/(4\theta)$, it is $q^*(\theta) = 2a\theta$, which is then the quality component in the equilibrium bidding function.

Now, given the optimal choice of q just determined , our problem reduces to a one-variable optimization problem:

$$\max_s U_i(s;\theta) = [aq^*(\theta) - s - C(q^*(\theta);\theta)] \times \mathrm{PW}(s),$$

or, by setting $k(\theta) = aq^*(\theta) - s - C(q^*(\theta),$

$$\max_s U_i(s;\theta) = [k(\theta) - s)] \times \mathrm{PW}(s).$$

Notice that the above maximand is equivalent to the expected utility of a supplier who participates in a standard first-price auction having a valuation $k(\theta)$ for the object on sale. Hence, as long as $k(\theta)$ is strictly increasing (which is the case under our assumptions), we can apply the standard theory to obtain the symmetric equilibrium bidding function in the score dimension, which is

$$\sigma(\theta) = \frac{1}{F(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} k(\theta) f(\theta) d\theta.$$

After noticing that, under our assumptions, $k(\theta) = a^2\theta$, we get

$$\sigma(\theta) = \frac{a^2(\theta+1)}{2},$$

which immediately allows us to obtain the price component of the equilibrium bidding function, namely

$$\alpha_3(\theta) = \frac{a^2(3\theta-1)}{2}.$$

We conclude that the symmetric equilibrium is

$$[\alpha_3(\theta);\beta_3(\theta)] = \left[\frac{a^2(3\theta-1)}{2}; 2a\theta\right],$$

which, for $a = 2$ (the parameter used in the experiment) yields (8).

Buyer's utility. The expected utility of the buyer in equilibrium (for generic quality weight a) is

$$U_B = \frac{20}{7} E[q^w] - E[p^w]$$

where $E[q^w]$ – the expected quality of the winner – is

$$E[q^w] = \int_{\underline{\theta}}^{\overline{\theta}} \beta_3(\theta) 2F(\theta) f(\theta) d\theta = 14a,$$

[33] To be precise, this statement is true as long as s gives a strictly positive probability of winning to the seller.

and $E[p^w]$ – the expected price of the winner – is

$$E[p^w] = \int_{\underline{\theta}}^{\overline{\theta}} \alpha_3(\theta) 2F(\theta) f(\theta) d\theta = 10a^2.$$

It follows

$$U_B = \frac{20}{7} \times 14a - 10a^2 = 40a - 10a^2,$$

which is maximized for $a = 2$ (exactly the parameter we used in the experiment). We thus have $U_B(a = 2) = 40$.

Total welfare. The expected total welfare in equilibrium is

$$W = \frac{20}{7} E[q^w] - E[C^w]$$

where $E[q^w] = 14a$ and $E[C^w]$ – the expected cost of the winner – is

$$E[C^w] = \int_{\underline{\theta}}^{\overline{\theta}} \frac{\beta_3(\theta)^2}{4\theta} 2F(\theta) f(\theta) d\theta = 7a^2.$$

It follows:

$$W = 40a - 7a^2.$$

For $a = 2$ – the parameter that we used in the experiment – it is $W(a = 2) = 52$.

B.4 *SRA2p*

Equilibrium. *SRA2p* works like *SRA* (each seller submits a price p and a quality q, the winner is the seller whose score $s = 2q - p$ is the highest), with the difference that, while the choice of q is unconstrained (apart from being below the maximum admissible value), p must be chosen in a set of two values only, $\{p_L, p_H\}$, with $p_L < p_H$.

To derive the symmetric equilibrium bidding function, we guess that it will have the following shape

$$[\alpha_4(\theta); \beta_4(\theta)] = \begin{cases} [p_L, \beta_L(\theta)] & \text{if} \quad \theta \in [\underline{\theta}, \hat{\theta}) \\ [p_H, \beta_H(\theta)] & \text{if} \quad \theta \in [\hat{\theta}, \overline{\theta}] \end{cases},$$

with score

$$\sigma(\theta) = \begin{cases} 2\beta_L(\theta) - p_L & \text{if} \quad \theta \in [\underline{\theta}, \hat{\theta}) \\ 2\beta_H(\theta) - p_H & \text{if} \quad \theta \in [\hat{\theta}, \overline{\theta}] \end{cases},$$

and that $\sigma(\theta)$ is strictly increasing, i.e. that: (i) $\beta_L(\theta)$ and $\beta_H(\theta)$ are strictly increasing, and (ii) $2\beta_L(\hat{\theta}) - p_L \leq 2\beta_H(\hat{\theta}) - p_H$.

Now, suppose seller j bids according to $[\alpha_4(\cdot); \beta_4(\cdot)]$ as described above, and consider seller i, type θ. If this seller chooses price p_L and a quantity $q \in (\beta_L(\underline{\theta}), \beta_L(\hat{\theta}))$, her expected utility will be

$$U_i(p_L, \beta_L(\underline{\theta}) < q < \beta_L(\hat{\theta}); \theta) = [p_L - C(q; \theta)] \times F(\beta_L^{-1}(q)).$$

The first-order condition for a maximum is:

$$-C'(q; \theta) \times F(\beta_L^{-1}(q)) + [p_L - C(q; \theta)] \times \frac{f(\beta_L^{-1}(q))}{\beta_L'(\beta_L^{-1}(q))} = 0.$$

In a symmetric equilibrium, for types $\theta < \hat{\theta}$, it must be $q = \beta_L(\theta)$ (hence q is indeed lower than $\beta_L(\hat{\theta})$). Using this symmetry condition, we get:

$$-C'(\beta_L(\theta);\theta) \times F(\theta) \times \beta_L'(\theta) + [p_L - C(\beta_L(\theta);\theta)] \times f(\theta) = 0,$$

or

$$p_L f(\theta) = C'(\beta_L(\theta);\theta) \times F(\theta) \times \beta_L'(\theta) + C(\beta_L(\theta);\theta) f(\theta).$$

Under our assumptions, the equation above becomes

$$\frac{1}{9} p_L = \frac{\beta_L(\theta)}{2\theta} \frac{\theta - 1}{9} \beta_L'(\theta) + \frac{1}{9} \frac{\beta_L(\theta)^2}{4\theta},$$

or

$$4 p_L \theta = 2\beta_L(\theta)\beta_L'(\theta)(\theta - 1) + \beta_L(\theta)^2.$$

The above condition must hold for all $\underline{\theta} < \theta < \hat{\theta}$. After noticing that the RHS above is the derivative of $(\theta - 1)\beta_L(\theta)^2$, we get

$$\int_{\underline{\theta}}^{\theta} 4\, p_L\, x\, dx = (\theta - 1)\beta_L(\theta)^2,$$

or

$$\beta_L(\theta) = \sqrt{2p_L(\theta + 1)}.$$

Suppose now that seller i, type θ, chooses price p_H and a quantity $q \in \big(\beta_H(\bar{\theta}), \beta_H(\overline{\theta})\big)$. Her expected utility will be

$$U_i(p_H, \beta_H(\bar{\theta}) < q < \beta_H(\overline{\theta}); \theta) = [p_H - C(q;\theta)] \times F(\beta_H^{-1}(q)).$$

Proceeding exactly as before, we obtain that, for all $\bar{\theta} < \theta < \overline{\theta}$, it is necessarily

$$4 p_H \theta = 2\beta_H(\theta)\beta_H'(\theta)(\theta - 1) + \beta_H(\theta)^2.$$

Solving it, we get

$$\int_{\hat{\theta}}^{\theta} 4\, p_H\, x\, dx = (\theta - 1)\beta_H(\theta)^2 - (\hat{\theta} - 1)\beta_H(\hat{\theta})^2,$$

or

$$\beta_H(\theta) = \sqrt{\frac{2p_H(\theta^2 - \hat{\theta}^2) + (\hat{\theta} - 1)\beta_H(\hat{\theta})^2}{\theta - 1}}.$$

To pin down the value of $\hat{\theta}$ and $\beta_H(\hat{\theta})$, we use the following two conditions (that are necessary for an equilibrium):

1. type $\hat{\theta}$ must be indifferent between bidding $[p_L; \beta_L(\hat{\theta})]$ and bidding $[p_H; \beta_H(\hat{\theta})]$: if not, then a type $\hat{\theta} - \epsilon$ would rather deviate and bid $[p_H; \beta_H(\hat{t})]$. This condition is

   $$p_L - C(\beta_L(\hat{\theta}); \hat{\theta}) = p_H - C(\beta_H(\hat{\theta}); \hat{\theta}),$$

   which implies

   $$\beta_H(\hat{\theta})^2 - \beta_L(\hat{\theta})^2 = 4\hat{\theta}(p_H - p_L);$$

2. the score if one bids $[p_L; \beta_L(\hat{\theta})]$ must be equal to the score if one bids $[p_H; \beta_H(\hat{\theta})]$: if not, then type $\hat{\theta}$ would rather deviate and decrease quality below $\beta_H(\hat{\theta})$ but still keeping her score above the one associated with $[p_L; \beta_L(\hat{\theta})]$. This condition is

$$2\beta_L(\hat{\theta}) - p_L = 2\beta_H(\hat{\theta}) - p_H,$$

which implies

$$2\left[\beta_H(\hat{\theta}) - \beta_L(\hat{\theta})\right] = p_H - p_L.$$

Solving the system of the two conditions above, we obtain[34]

$$\hat{\theta} = \frac{p_H + \sqrt{2p_H p_L + 32 p_L - p_L^2}}{16}, \qquad \beta_H(\hat{\theta}) = \frac{p_H - p_L}{4} + 4\hat{\theta}.$$

For $p_L = 12$ and $p_H = 65$ (the parameters used in the experiment) we obtain (9).[35]

Buyer's utility. The expected utility of the buyer in equilibrium (for generic prices p_L and p_H) is

$$U_B = \frac{20}{7}E[q^w] - E[p^w]$$

where $E[q^w]$ – the expected quality of the winner – is

$$E[q^w] = \int_{\underline{\theta}}^{\hat{\theta}} \beta_L(\theta) 2F(\theta)f(\theta)d\theta + \int_{\hat{\theta}}^{\overline{\theta}} \beta_H(\theta)2F(\theta)f(\theta)d\theta,$$

and $E[p^w]$ – the expected price of the winner – is

$$E[p^w] = p_L F(\hat{\theta})^2 + p_H(1 - F(\hat{\theta})^2) = p_H - \frac{(p_H - p_L)(\hat{\theta} - 1)^2}{81}.$$

(In the above, $F(\cdot)^2$ is the cumulative distribution of the first order statistic.) To determine the values of p_L and p_H that maximize U_B, we resorted to numerical techniques, obtaining $p_L = 12$ and $p_H = 65$, the parameters used in the experiment. We then have $U_B(p_L = 12, p_H = 65) \approx 40.2$.

Total welfare. The expected total welfare in equilibrium is

$$W = \frac{20}{7}E[q^w] - E[C^w]$$

where $E[q^w]$ is as above and $E[C^w]$ – the expected cost of the winner – is

$$E[C^w] = \int_{\underline{\theta}}^{\hat{\theta}} \frac{\beta_L(\theta)^2}{4\theta} 2F(\theta)f(\theta)d\theta + \int_{\hat{\theta}}^{\overline{\theta}} \frac{\beta_H(\theta)^2}{4\theta} 2F(\theta)f(\theta)d\theta.$$

For $p_L = 12$ and $p_H = 65$, we obtain $W(p_L = 12, p_H = 65) \approx 50.6$.

[34] Notice that, if, for some (p_L, p_H) one gets $\hat{\theta} < \underline{\theta}$, this would mean that all types prefer bidding p_H, and we would be back in a *FQA* with fixed price p_H; similarly, if, for some (p_L, p_H) one gets $\hat{\theta} > \overline{\theta}$, this would mean that all types prefer bidding p_L, and we would be back in a *FQA* with fixed price p_L.

[35] Clearly, the one just presented is not a proof as it is based only on (local) necessary first-order conditions. Nevertheless, it can easily be shown that the bidding function obtained is indeed an equilibrium.

B.5 *SRA2q*

Equilibrium. *SRA2q* works like *SRA* (each seller submits a price p and a quality q, the winner is the seller whose score $s = 2q - p$ is the highest), with the difference that, while the choice of p is unconstrained (apart from being below the maximum admissible value), q must be chosen in a set of two values only, $\{q_L, q_H\}$, with $q_L < q_H$.

Following the same steps as those used for the *SRA2p* case, one obtains

$$[\alpha_5(\theta); \beta_5(\theta)] = \begin{cases} \left[q_L^2/4; 9\right] & \text{if} \quad \theta = 1 \\ \left[q_L^2[4(\theta-1)]^{-1}\ln\theta; 9\right] & \text{if} \quad \theta \in (1, \hat{\theta}) \\ \left[\frac{(q_H^2/4)(\ln\theta - \ln\hat{\theta}) + (\hat{\theta}-1)\hat{b}}{\theta-1}; 40\right] & \text{if} \quad \theta \in [\hat{\theta}, 10] \end{cases},$$

where

$$\hat{\theta} = \frac{q_L + q_H}{8}, \qquad \hat{b} = 2(q_H - q_L) + \frac{q_L^2}{4}\frac{\ln\hat{\theta}}{\hat{\theta} - 1}.$$

For $q_L = 9$ and $q_H = 40$ (the parameters used in the experiment) we obtain (10).

Buyer's utility. The expected utility of the buyer in equilibrium (for generic qualities q_L and q_H) is

$$U_B = \frac{20}{7}E[q^w] - E[p^w]$$

where $E[q^w]$ – the expected quality of the winner – is

$$E[q^w] = q_L F(\hat{\theta})^2 + q_H(1 - F(\hat{\theta})^2) = q_H - \frac{(q_H - q_L)(\hat{\theta} - 1)^2}{81}.$$

and $E[p^w]$ – the expected price of the winner – is

$$E[p^w] = \int_{\underline{\theta}}^{\hat{\theta}} \frac{q_L^2}{4(\theta - 1)} \ln\theta \cdot 2F(\theta)f(\theta)d\theta + \int_{\hat{\theta}}^{\overline{\theta}} \frac{(q_H^2/4)(\ln\theta - \ln\hat{\theta}) + (\hat{\theta} - 1)\hat{b}}{\theta - 1} 2F(\theta)f(\theta)d\theta.$$

To determine the values of q_L and q_H that maximize U_B, we resorted to numerical techniques, obtaining $q_L = 9.35$ and $q_H = 40.68$. For $q_L = 9$ and $q_H = 40$ – the parameters that we used in the experiment – it is $U_B(q_L = 9, q_H = 40) \approx 38.5$.

Total welfare. The expected total welfare in equilibrium is

$$W = \frac{20}{7}E[q^w] - E[C^w]$$

where $E[q^w]$ is as above and $E[C^w]$ – the expected cost of the winner – is

$$E[C^w] = \int_{\underline{\theta}}^{\hat{\theta}} \frac{q_L^2}{4\theta} 2F(\theta)f(\theta)d\theta + \int_{\hat{\theta}}^{\overline{\theta}} \frac{q_H^2}{4\theta} 2F(\theta)f(\theta)d\theta = q_L^2[\hat{\theta} - \ln\hat{\theta} - \underline{\theta} + \ln\underline{\theta}] + q_H^2[\overline{\theta} - \ln\overline{\theta} - \hat{\theta} + \ln\hat{\theta}].$$

For $q_L = 9$ and $q_H = 40$, we obtain $W(q_L = 9, q_H = 40) \approx 50.5$.

C Experimental results: figures

Figure (A1) displays the distribution of observed quality bids in *FQA* and price bids in *FPA* (gray box plot),[36] the mean observed bids (black connected line) and the risk-neutral Bayes-Nash equilibrium prediction (red dashed line).

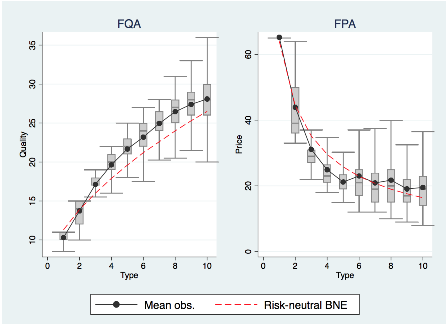
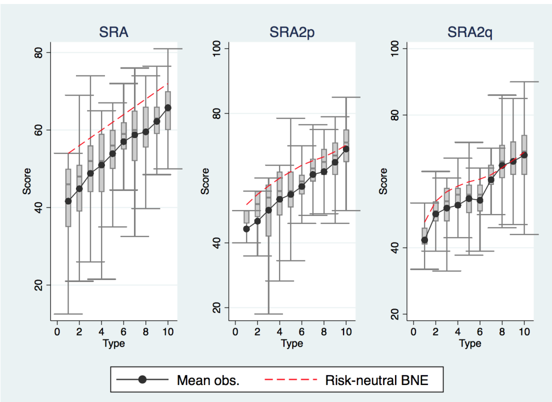
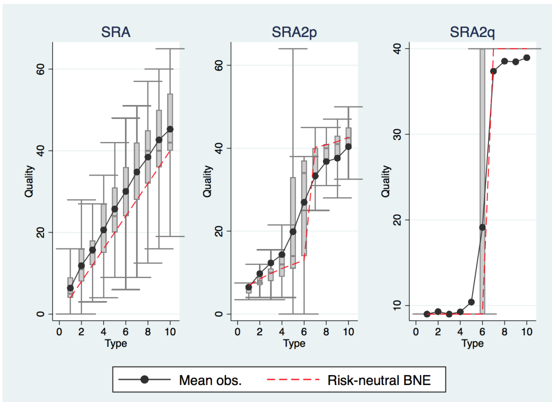


Figure A1 – Bids in *FQA* and *FPA*.

Figure (A2a) displays the distribution of scores (gray box plot) in *SRA*, *SRA2p* and *SRA2q*, the mean observed scores (gray connected line) and the risk-neutral Bayes-Nash equilibrium prediction (red dashed line). Figure (A2b) displays the distribution of the observed quality component of the bid (gray box plot) in SRA, SRA2p and SRA2q, the mean observed quality (gray connected line) and the risk-neutral Bayes-Nash prediction (red dashed line).



(a) Score



(b) Quality

Figure A2 – Bids in *SRA*, *SRA2p* and *SRA2q*

[36]In Figure (A1) and in Figure (A2a) and (A2b), the gray boxes include bids within the second and the third quartile of the observed distribution; the dark-gray segment within each box is the median observation; the two vertical gray lines extend up to 1.5 times the interquartile range

D Experimental results: parametric analysis, last 5 periods

Table A1 – Outcome of the auction, last 5 periods

	SW		BU		SP	
	(1)	(2)	(3)	(4)	(5)	(6)
FPA	-0.261***	-0.272***	-0.258***	-0.289***	-0.003	0.016
	(0.027)	(0.042)	(0.033)	(0.063)	(0.019)	(0.035)
SRA2q	-0.031	-0.020	-0.028	-0.047*	-0.003	0.026
	(0.020)	(0.020)	(0.025)	(0.027)	(0.015)	(0.017)
SRA2p	-0.023	-0.037	-0.025	-0.050	0.003	0.013
	(0.018)	(0.036)	(0.022)	(0.032)	(0.014)	(0.016)
SRA	-0.024	0.042	-0.001	0.008	0.026*	0.034
	(0.018)	(0.032)	(0.020)	(0.025)	(0.015)	(0.023)
Trend		0.007**		0.001		0.006*
		(0.003)		(0.003)		(0.003)
FPA∗Trend		0.006		0.015		0.010
		(0.018)		(0.022)		(0.010)
SRA2q∗Trend		-0.006		0.009		-0.015***
		(0.009)		(0.010)		(0.005)
SRA2p∗Trend		0.007		0.012		0.005
		(0.012)		(0.009)		(0.005)
SRA∗Trend		-0.009		-0.005		-0.004
		(0.011)		(0.008)		(0.007)
Constant	0.845***	0.831***	0.671***	0.669***	0.174***	0.162***
	(0.009)	(0.012)	(0.010)	(0.012)	(0.009)	(0.008)
Obs.	300	300	300	300	300	300
Wald $-\chi^2$	98.85	109.67	62.88	69.31	4.44	18.44
$p > -\chi^2$	0.000	0.000	0.000	0.000	0.350	0.030

Note. Table A1 report estimates (robust standard errors in parentheses) from GLS random effects models accounting for dependency within rematching group. In all regressions, the dependent variable is defined at the rematching group level. Only the last 5 periods are considered. Columns (1) and (2) focus on *SW*, Columns (3) and (4) focus on *BU*, Columns (5) and (6) focus on *SP*. Trend is a linear time trend that starts from 0 in the 11th period of the experiment. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Table A2 – Score, parametric analysis, last 5 periods

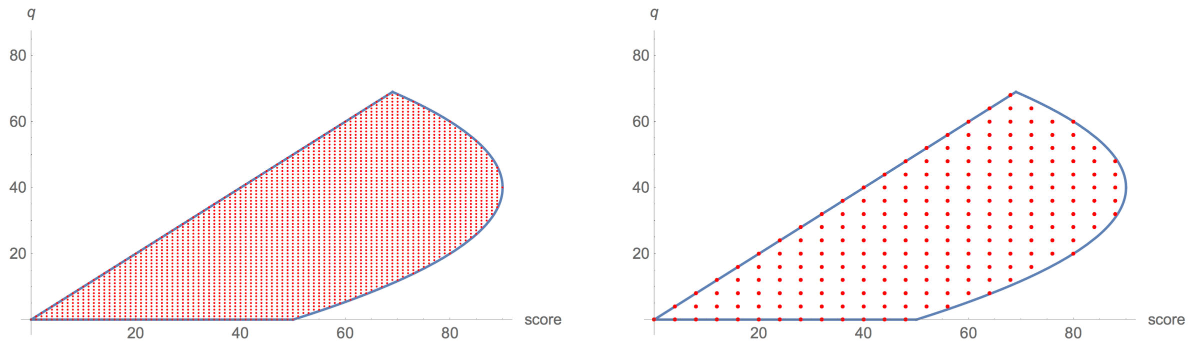
		Score			*Score_dist*	
	(1)	(2)	(3)	(4)	(5)	(6)
FPA	-8.136***	-9.987***	-11.148***	-0.029*	0.030	0.019
	(1.123)	(1.288)	(1.467)	(0.015)	(0.020)	(0.023)
SRA2q	-1.000	6.950***	6.541***	-0.057***	-0.050**	-0.060***
	(1.123)	(1.301)	(1.470)	(0.0148)	(0.020)	(0.023)
SRA2p	-0.908	5.984***	5.457***	-0.093***	-0.101***	-0.111***
	(1.123)	(1.299)	(1.471)	(0.015)	(0.020)	(0.023)
SRA	-2.025*	7.386***	6.741***	-0.118***	-0.096***	-0.107***
	(1.123)	(1.294)	(1.475)	(0.015)	(0.020)	(0.023)
θ		3.873***	3.874***		0.009***	0.009***
		(0.124)	(0.124)		(0.002)	(0.002)
FPA$*\theta$		0.493***	0.505***		-0.011***	-0.011***
		(0.177)	(0.176)		(0.003)	(0.003)
SRA2q$*\theta$		-1.463***	-1.462***		-0.001	-0.001
		(0.178)	(0.177)		(0.003)	(0.003)
SRA2p$*\theta$		-1.333***	-1.331***		0.001	0.001
		(0.175)	(0.174)		(0.003)	(0.003)
SRA$*\theta$		-1.705***	-1.697***		-0.004	-0.004
		(0.177)	(0.177)		(0.003)	(0.003)
Trend			0.046			$-9 \cdot 10^{-4}$
			(0.238)			(0.004)
FPA$*$Trend			0.548			0.005
			(0.337)			(0.005)
SRA2q$*$Trend			0.203			0.005
			(0.337)			(0.005)
SRA2p$*$Trend			0.257			0.005
			(0.337)			(0.005)
SRA$*$Trend			0.299			0.005
			(0.337)			(0.005)
Constant	59.956***	39.116***	38.565***	0.042***	-0.008	-0.008
	(0.794)	(0.915)	(3.019)	(0.010)	(0.014)	(0.016)
Obs.	1800	1800	1800	1800	1800	1800
Wald $-\chi^2$	68.20	3360.71	3394.81	82.55	172.44	181.06
$p > -\chi^2$	0.000	0.000	0.000	0.000	0.000	0.000

Note. This table reports estimates (clustered standard errors in parentheses) from two-way linear random effects models accounting for both potential individual dependency over repetitions and dependency within rematching group. Only the last 5 periods are considered. The first three columns are based on regressions using the observed score as dependent variable, while the last three are based on *score_dist*. θ is the cost parameter randomly assigned to the supplier. Trend is a linear time trend that starts from 0 in the 11th period of the experiment. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

E Structural analysis: binning procedure

In this section we present the binning methodology used to estimate the QRE model. Binning was unavoidable: *SRA*, the treatment with the largest strategy set, had 19,146 different combinations of score, quality and cost parameter and, to the best of our knowledge, a QRE model with such a huge number of strategies cannot be estimated on reasonable time. After binning, we ended up with 1,298 different combinations.

For illustration, the red points in Figure (A3a) represent all the available strategies in *SRA* for a seller of type $\theta = 10$. After binning, the relevant strategies for this seller were reduced to those depicted in Figure (A3b).



(a) All strategies **(b)** Binned strategies

Figure A3 – Available and binned strategies in *SRA* for a seller of type $\theta = 10$

To construct the binned strategies for *SRA*, we divided the space of score-quality pairs[37] into squares of side equal to 3. Then, all the observed score-quality pairs falling into a certain square were treated as identical, and assigned a value equal to the coordinates of the center of the square. In practice, the first square has vertices $(0,0)$, $(3,0)$, $(0,3)$, and $(3,3)$; then, moving horizontally, we have the square with vertices $(4,0)$, $(7,0)$, $(4,3)$, and $(7,3)$, while moving vertically, we have the square with vertices $(0,4)$, $(3,4)$, $(0,7)$, and $(3,7)$. And so on. As an example, the strategy pair $(1,3)$ belongs to the first square: as such, a bid $(1,3)$ has been re-assigned the value corresponding to the central point of that square, namely $(1.5,1.5)$. For squares crossing the boundaries of the set of admissible strategies and whose center was outside this set, we picked a boundary point.

A similar binning procedure was applied to the other treatments as well. Clearly, for the unidimensional treatments, bins were intervals (of length 3), not squares. For *SRA2q* and *SRA2p*, the binning procedure was applied along the non-binary dimension only.

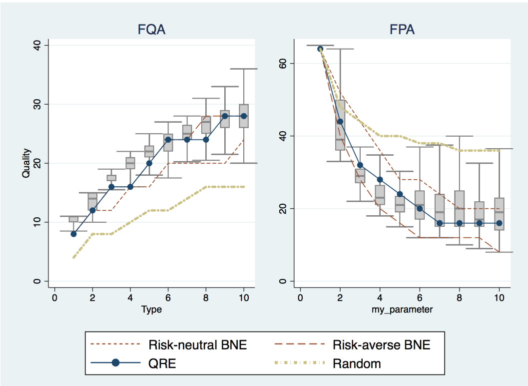
Our computer required around 4 hours to solve for the QRE model for *SRA*, for each set of parameters r and μ.

[37]We worked on score-quality pairs rather than on price-quality pairs because writing the codes in the former case was easier.

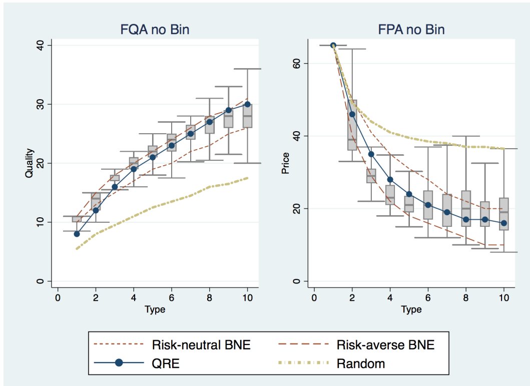
F Baseline QRE: results for *FPA*, *FQA*, *SRA2p* and *SRA2q*

In all the figures below, we report:

- the median bids of the fitted distribution of the QRE (qualities in FQA, prices in FPA and, in separate graphs, scores and qualities in SRA2p and SRA2q), using a blue line
- the distribution of the observed bids, using a gray box-plot[38]
- the Bayes-Nash risk-neutral equilibrium, using a red short-dashed line
- the Bayes-Nash risk-averse equilibrium,[39] using a red dashed line
- the median bids of the random model, using a yellow dotted line.

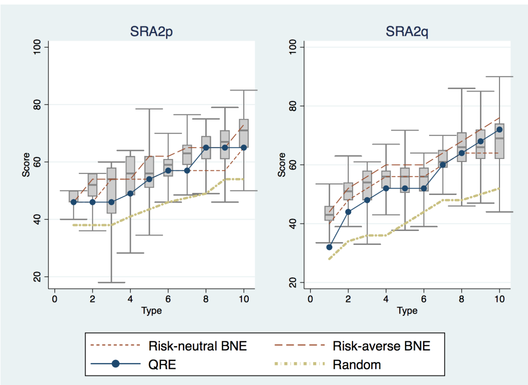


(a) With binning

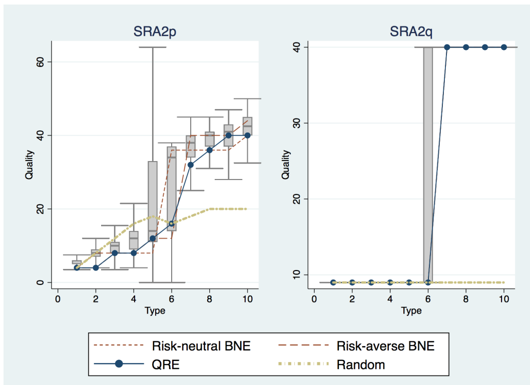


(b) Without binning

Figure A4 – Bids in *FQA* and *FPA*: observed vs. predicted



(a) Score



(b) Quality

Figure A5 – Scores and quality bids in *SRA2p* and *SRA2q*: observed vs. predicted

[38]The gray boxes include bids within the second and the third quartile of the observed distribution; the dark-gray segment within each box is the median observation; the two vertical gray lines extend up to 1.5 times the interquartile range

[39]Using a CRRA utility function with the same risk-aversion parameter r and the same bin (4 units in each dimension with a continouos choice set) as in the QRE.

G *SRA* with misperception

Recall the equilibrium quality and score for *SRA* (with generic quality weight a and disregarding the additive constant):

$$\begin{aligned}\beta_3(\theta) &= 2a\theta \\ \sigma_3(\theta) &= \frac{a^2(\theta+1)}{2}.\end{aligned}$$

Now, suppose that the (perceived) scoring rule is $\tilde{s}(q,p) = a \cdot w_s \cdot q - p$, and that the (perceived) cost function is $\tilde{C}(q;\theta) = w_c \cdot q^2/(4\theta)$ (with $w_s, w_c > 0$). The perceived expected utility of a generic seller i, type θ, is then:

$$\tilde{U}_i = \left[a \cdot w_s \cdot q - \tilde{s} - w_c \frac{q^2}{4\theta}\right] \times \text{PW}(\tilde{s}),$$

Using the same logic as in the derivation of the equilibrium in the standard *SRA*, it is straightforward to verify that the equilibrium quality and score are equal to:

$$\begin{aligned}\tilde{\beta}_3(\theta) &= 2a\theta \times \frac{w_s}{w_c} \\ \tilde{\sigma}_3(\theta) &= \frac{a^2(\theta+1)}{2} \times \frac{w_s^2}{w_c}.\end{aligned}$$

To have overbidding in the quality, i.e. $\tilde{\beta}_3(\theta) > \beta_3(\theta)$, it must be $w_s > w_c$. To have underbidding in the score, i.e. $\tilde{\sigma}_3(\theta) < \sigma_3(\theta)$, it must be $w_s^2 < w_c$. Both conditions are simultaneously satisfied only if $w_c < w_s < 1$.

H Quality bids in *SRA*: comparison with Lewis and Bajari

In this section, we refer to the data analyzed by Lewis and Bajary (2011) concerning scoring auctions adopted by the Californian Department of Transportation to award highway repair contracts. Our goal is to compare their real world data with our laboratory results. In particular, we focus on the quality bids only – not on price or on the overall score – for two reasons. First, the equilibrium quality is unaffected by the precise shape of the seller's utility function (as long as it only depends on the monetary earnings) and by the beliefs regarding the other seller's bidding strategy: this allows us to rule out several competing explanations to the observed variability in bidding behavior. Second, and differently from price, the quality bid in Lewis and Bajary (2011) is structurally modelled.

The main estimating equation for quality bids in Lewis and Bajary (2011) is as follows:[40]

$$\ln \widetilde{d}_{ij} = \frac{1}{\alpha}\left(\ln c_{Uj} - x_{ij}\beta - \zeta_j - \theta_{ij}^A\right), \tag{11}$$

where the index i refer to sellers and index j to contracts/auctions, α captures the curvature (convexity for $\alpha > 1$) of the cost function, c_U is the relative weight attached to quality in the scoring rule, $\widetilde{d}$ is the observed quality bid, x is a vector of observed seller- and contract-specific characteristics, ζ_j is an unobserved contract-specific dummy, and θ_{ij}^A is the type of seller i for contract j (essentially, the unobserved component affecting seller i's cost for project j).

[40] See Equation (5) on page 1196 in Lewis and Bajari (2011). For simplicity, in this section we keep their notation.

Table A3 – Quality decision: Log Days Accelerated, SRA auctions

	LB (1)	FE (2)	RE (3)	FE+RE (4)
Log Usercost	0.275**		0.342*	0.325*
	(0.136)		(0.175)	(0.178)
Log Engineer's Days	1.312***		1.417***	1.427***
	(0.228)		(0.227)	(0.232)
Log Engineer's Estimate	-0.314**		-0.452***	-0.500***
	(0.136)		(0.164)	(0.166)
Log Daily Traffic	0.002		0.055	0.067
	(0.067)		(0.084)	(0.085)
Plant Establishment	-0.292		-0.346*	-0.383**
	(0.205)		(0.193)	(0.194)
Lane Closure Fraction	0.836		0.956	1.115
	(0.686)		(0.722)	(0.742)
Reopening Penalty	-0.235		-0.355	-0.402*
	(0.216)		(0.233)	(0.233)
Firm Capacity > $50M	0.250		0.344	
	(0.202)		(0.250)	
Instate Contractor	-0.566***		-0.795	
	(0.200)		(0.515)	
Log Distance (miles)	-0.028	-0.105	-0.048	-0.068
	(0.084)	(0.076)	(0.071)	(0.071)
Participation Residual	-0.286	-0.103	-0.134	-0.176
	(0.725)	(0.608)	(0.652)	(0.606)
Constant	0.343	4.771***	0.629	1.069
	(1.522)	(0.542)	(1.739)	(1.742)
District/Work/Year FE	YES	NO	YES	YES
Contracts FE	NO	YES	NO	NO
suppliers FE	NO	YES	NO	YES
Std. Errors				
Contracts RE			0.338	0.382
suppliers RE			0.794	
Residuals	0.952	0.530	0.731	0.600
Mean Log. Days Accell.	4.250	4.250	4.250	4.250
Obs.	424	424	424	424
R^2	0.408	0.817	0.408	0.694
$adj-R^2$	0.369	0.693	0.369	0.577

Note. Column (1) reports the estimate of the linear regression model as in Table 4, Column (4) in Lewis and Bajari (2011) (clustered standard errors at the contract level in parentheses). Column 2 reports the estimate of a linear regression model with contract and supplier FEs (robust standard errors in parentheses). Column (3) reports a two-way crossed-effects model, accounting for both contract and supplier REs. Column 4 considers contracts REs and suppliers FEs. All REs standard errors have been reported, as well as the standard error of the residuals. Contract specific variables are: Log Usercost (log weight of quality into the scoring function), Log Engineering Days (the log engineering estimate on the number of days required to complete the project, i.e. the minimum admitted quality), Log Engineering Estimate (the log engineering estimate of the project's cost, i.e. the reserve price), Plant Establishment (1 if a plant has been established, 0 o/w), Log Daily Traffic (the log traffic volumes), Lane Closure Fraction (the fraction of lanes closed during the work) and Reopening Penalty (1 if the contract specifies a monetary penalty for late reopening, 0 o/w). Supplier specific variables are: Firm Capacity larger than $50M (1 if the firm has a capacity larger that $50M, 0 o/w) and Instate Contractor (1 if the bidder comes from the same State where the contract is supposed to be executed, 0 o/w). Supplier-contract variables are the Log Distance (the log distance between the firm's headquarter and where the work has to be executed) and Participation Residuals (used to correct for endogenous participation as described in Lewis and Bajary (2011)). Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

We replicate the estimation exercise of Lewis and Bajary (2011) by using a more parsimonious specification, where we include contract FEs (to control for unobserved contract characteristics) and supplier FEs (to control for unobserved and persistent sellers' characteristics, like productive efficiency, ability and psychological attitudes). We estimate this model in Column (2), Table A3. For comparison, we also report, in Column (1), the original results of Lewis and Bajary (2011). Our objctive is, first, to absorb all the fixed effects and, second, to estimate how much variance is left. Our regression model with 77 contract FEs, 94 supplier FEs and 2 supplier-contract interactions (in a dataset of 424 observations), leaves unexplained about 30% of the variability in the quality bids. This variability has a relevant magnitude, as one standard error of the residuals is equivalent to 12.5% of the mean dependent variable.

In order to check for robustness of the results, in Column (3) we keep all the independent variables as in Lewis and Bajari's model of Column (1), but we add both supplier and contract random effects. In Column (4), instead, we consider supplier fixed effects and contracts random effects. In both cases, moving from fixed to random effect models does not reduce the variance in the residuals.

We then repeat the same exercise on our experimental data. In the lab, we do observe the seller's costs and contracts are homogeneous. The predicted quality bid in *SRA* is a linear function of the seller's type (specifically, $q^* = 4\theta$). In the first column of Table A4, we estimate a panel regression model with individual FEs. Our specification controls are able to account for around 72% of the variability in the quality bid, thus leaving unexplained about 28% of the overall variance. Additionally, one standard error of the residuals is equivalent to 26.5% of the mean dependent variable. Results do not differ using a random effect model (see the second column).

Table A4 – Quality decision in the lab, SRA auctions

	FE	RE
	(1)	(2)
θ	4.479***	4.467***
	(0.140)	(0.077)
Constant	2.496***	2.558***
	(0.746)	(0.798)
Individuals FE	YES	NO
Std. Errors		
Individual RE		1.815
Subgroup RE		3.215
Residuals	7.016	7.259
Mean Quality	26.407	26.407
Obs.	1080	1080
R^2	0.720	0.720

Note. Column (1) reports estimates (robust standard errors clustered at the rematching group in parentheses) on a fixed-effect panel regression. FEs are at the individual level. Column (2) reports estimates (clustered standard errors in parentheses) from two-way linear random effects models accounting for both potential individual dependency over repetitions and dependency within rematching group. All REs standard errors have been reported, as well as the standard error of the residuals. In both regressions, the dependent variable is the quality chosen by suppliers in the SRA treatment. The mean dependent variable has been reported. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.