Journal of Behavioral and Experimental Economics 108 (2024) 102131

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Scoring rules in experimental procurement☆

Gian Luigi Albano [a], Angela Cipollone [b], Roberto Di Paolo [c,d], Giovanni Ponti [b,d,*], Marco Sparro [a]

[a] Consip S.p.A., Italy
[b] LUISS Guido Carli Roma, Italy
[c] IMT Lucca, Italy
[d] Universidad de Alicante, Spain

ARTICLE INFO

JEL Classification:
C91
D70
D81
D91

Keywords:
Scoring auctions
Mechanism design
Experimental economics

ABSTRACT

We report the results of an experiment where subjects compete for procurement contracts to be awarded by means of a scoring auction. Two experimental conditions are considered, depending on the relative weight of quality vs price in the scoring rule. We show that different quality-price weights dramatically alter the strategic environment and affect efficiency. Our evidence shows that each weighting better delivers against a matching objective function than using a scoring rule which misrepresents the buyer's objective function. Nonetheless, there are large deviations in how each performs, with the higher weight on quality delivering much greater efficiency evaluated against its own objective function than a low weight on quality evaluated against its own objective function, despite the higher quality weight inducing higher deviations from equilibrium. We propose a "mediation analysis" to show that the "direct effect" (due to the different strategic properties of the induced game-forms) outweighs the "indirect" one (how the different game-forms affect out-of-equilibrium behaviour). We also perform a structural estimation of the Quantal Response Equilibrium induced by subjects' behavior, where we find that subjects are risk averse and noisy play affects behavior in the direction of underbidding.

1. Introduction

During the last two decades, public procurement has undergone profound changes. Policy makers, academics and practitioners alike share the broad view that public procurement has evolved from a clerical signoff-ridden set of activities to a strategic tool to enhance efficiency in public organizations, to regulate markets and promote sustainable development. Thanks to a profound reformulation of public procurement regulations at a global level promoted by forward-looking policymakers as well as specialised procurement organizations, public procurement is being increasingly used to pursue objectives beyond the mere acquisition of works/products/services. Coherently with these objectives public organisations are urged to carry out competitive processes by evaluating a wide array of characteristics, comprising both financial and non-financial dimensions. For instance, the EU public procurement Directive 2014/24/EU foresees that "[…] *contracting authorities shall base the award of public contracts on the most economically advantageous tender*".[1] This implies that, under normal circumstances, public organisations shall consider both price and non-price dimensions in awarding public contracts, although the lowest-price award remains an admissible award criterion.[2]

Scoring (or multi-attribute) auctions are among the most widespread competitive mechanisms to evaluate heterogeneous tenders. In a scoring auction, the buyer commits to a scoring mechanism, which maps each

☆ We thank Todd Kaplan (Editor in charge) and two anonymous referees for their insightful comments that helped to improve our original submission considerably. Our gratitude also goes to Simona Berto, Federico Cesarini, Alessio Muscarnera, Laura Pacini, Annamaria Paolillo and Michele Rosi for their precious research assistance. The usual disclaimers apply. This paper was revised while Giovanni Ponti was visiting the Rady School of Business at the University of California San Diego. He would like to thank Uri Gneezy and all people at Rady for their kind hospitality. Financial support from the Spanish Ministry of Science and Innovation (Grant PID2022–142356NB-I00) and Generalitat Valenciana (Prometeo/2021/073 and Research Groups 3/086) is gratefully acknowledged.

* Corresponding author at: Deparamento de Fundamentos del Análisis Económico, Universidad de Alicante, 03071 Alicante, Spain.
E-mail address: giuba@ua.es (G. Ponti).

[1] Directive 2014/24/EU, art. 67(1).

[2] "Member States may provide that contracting authorities may not use price only or cost only as the sole award criterion or restrict their use to certain categories of contracting authorities or certain types of contracts." (Directive 2014/24/EU, art. 67(2)).

https://doi.org/10.1016/j.socec.2023.102131
Received 13 January 2022; Received in revised form 12 November 2023; Accepted 13 November 2023
Available online 22 November 2023
2214-8043/

tender's financial and non-financial attributes onto a one-dimensional score.[3] In a highest-score auction (Che, 1993) the tender awarded the highest score is deemed to be the winner and receives a financial payment equal to the submitted bid.[4]

Given the increasing relevance in private and public procurement markets of multi-attribute competitive mechanism, one may wonder to what extent bidders are able to cope with the arguably more sophisticated strategic environment of scoring auctions. This question becomes even more compelling as there exists substantial experimental evidence that -even in simple price-only auctions- actual behaviour may systematically differ from what theory predicts (see, for instance, Cason, 1995, Kagel & Levin, 2002, 2015). In this respect, the experimental evidence on scoring auctions is even more scant than the theoretical one, where -to the best of our knowledge- most experimental studies are concerned with the extent with which buyers may benefit from using a multi-attribute rather than a price-only auction. This strand of research has been initiated by Bichler (2000), who runs an experiment that mimics a financial market to assess the performance of multi-attribute mechanisms against single-attribute auctions. In his experiment, he finds that the buyer achieves a higher utility in the multi-attribute mechanisms, although efficiency -measured as the relative frequency of auctions where the higher value holder is awarded the contract- is similar. Chen-Ritzo et al. (2005) conduct an experiment that involves a multi-attribute auction where bidders have to set the price together with two other no-financial attributes: quality and time lead. 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. Along similar lines, Strecker (2010) studies the effect of revealing information in a multi-attribute reverse English auction with one buyer and five sellers in a setting where bids comprise one financial and two non-financial attributes. His-findings suggest that efficiency is greater when the scoring rule is revealed than when only limited information is provided to sellers. However, the buyer's surplus is not significantly affected by the specific information-revelation policy. Cason et al. (2011) report a laboratory experiment that investigates the role of revelation policies in a sequential procurement auction with private cost information. They compare two different information revelation policies: one in which all bids are revealed between auctions and one in which only the winning bid is revealed. Their results show that complete information revelation affects bidding behavior, reducing the buyer's surplus compared with the incomplete information case. On the agri-environment procurement auctions side, Hailu and Schilizzi (2004) simulate an agent-based model using repeated payment in ecosystem service (PES) auctions. PES are standard procurement auctions with a specific (environmental) good. Along their simulations, Hailu and Schilizzi (2004) results provide insight into long-term performance suggesting that, with repeated auctions, efficiency is lower than in one-shot auctions. Fooks et al. (2015) run an experiment in which they allow bidders to enter and submit offers at any time. Interestingly, they show that this dynamic mechanism is more efficient than the static, single-round, alternative.

In this paper, we present the results of a stylized procurement auction experiment where a simulated buyer has to select the contractor from a pool of five potential suppliers by means of a competitive mechanism. The buyer cares both about financial and non-financial aspects of the submitted tenders and announces, in advance, the maximum price she is willing to pay. Then, all potential sellers submit two-dimensional bids. These bids are comprised of a quality-price pair, in which the quality affects the seller's cost of production and the price is expressed as a rebate or discount to the buyer relative to the stated maximum price. Price and quality dimensions are then mapped into a unidimensional score and the contract is awarded to the highest-score bidder.

Despite the practical relevance in real procurement markets, scoring auctions have only attracted a limited theoretical investigation. Che's (1993) seminal paper provides the first comprehensive characterization of bidders' equilibrium strategies with *endogenous* quality choice. In his model, bidders privately observe their efficiency level (i.e., their quality production costs) and then, simultaneously, submit a quality-price pair. Within this framework, he can prove that the price/quality decision bidders face can be reduced to a single-dimensional problem by establishing that, as for the quality decision, rational bidders will always submit the socially efficient quality level, independent of their bidding behaviour. In this reduced one-dimensional problem, bidders can be ranked according to their "productive potential" -defined as *pseudotype*- that is, the highest level of social welfare they can produce. It also turns out that if Che's pseudotypes are monotonic in the efficiency levels then scoring auctions can be assimilated to first-price auctions and, therefore, well-known results in price-only auctions can be applied to derive bidders' equilibrium behavior.[5]

Unlike Che (1993), in our multi-period experiment quality is *exogenously* determined, in that each participant, at the beginning of each period, is endowed with a fixed quality level, an independent uniform draw (without replacement) from a finite grid. There are several reasons for designing such an adverse-selection framework. First, there are many procurement environments where quality choices are made before -or independent of- the design of the scoring auction. This is usually the case in the procurement of medical equipment, where firms' decisions about the quality characteristics of, say, an ultrasound or magnetic resonance imagining (MRI) machine are made by considering the impact on global sales rather than the competitive processes carried out by a single hospital in a specific country. This situation also applies to the procurement of IT equipment such as photocopiers or laptops. Second, a scoring auction with fixed quality levels gives rise to a less complex strategic environment for the participants in the experiment. Given that the scoring rule is known to participants beforehand, each bidder, endowed with a certain quality level, becomes immediately aware of his technical score. Hence his strategic problem boils down to computing the rebate so as to maximize expected profits, where the event of winning coincides with the event that the same bidder has the highest score. Last, but not least, by providing each bidder with a full range of possible qualities (without replacement) we are able to elicit a full bidding function -the unit of analysis for our empirical exercises- for each participant.[6]

The remainder of the paper is arranged as follows. Theory is presented in Section 2, where we model our competitive mechanism as a linear scoring auction with exogenous quality levels. Our two treatment conditions are especially designed so that pseudotypes may or may not monotonically increase with quality. This depends on the relative weight of the financial attribute in the scoring rule. In one treatment the weight of the quality is sufficiently high so that the strategic environment is compatible with Che's (1993) modeling assumption and the distribution of pseudotypes is monotonically increasing in the quality level. By contrast, in the other treatment, the weight of the rebate is sufficiently high so that the distribution of pseudotypes becomes reverse U-shaped, which, in turn, implies that the seller with the highest pseudotype lays in the interior of the support of the possible quality levels. Thus, when the scoring rule puts a relatively high weight on price, not

[3] A similar mechanism is the so-called *buyer-determined* procurement auction, which can be considered as a multi-dimensional auction in which the scoring rule is private information. In a buyer-determined procurement auction the buyer simply sets the reserve price and a list of conditions on the quality of the good/services. Once sellers have submitted their bid, the buyer is free to assign the contract at her wish (Santamaria, 2015).

[4] This is arguably the most widely used mechanism in the family of scoring auctions.

[5] Asker and Cantillon (2008) further generalize and extend Che's (1993) results by allowing for multidimensional type-space.

[6] See Grimm et al. (2008, 2009).

only are bidders provided with an incentive to bid more aggressively, but also the resulting non-monotonic distribution of pseudotypes dramatically alters the strategic problem they face. Proposition 1 collects the main characteristics of these two alternative equilibrium configurations, which depend on the relative weight of quality vs. rebate. Our theoretical analysis calls for an experimental design -described in detail in Section 3- which is built upon two (between-subject) conditions, depending on the relative weight of quality vs. price. Fixed groups of five bidders play repeatedly for 11 rounds, where each bidder is assigned each and every quality level within the grid. Subjects receive no feed-back about the outcome of the auctions they participate in until the end of the experiment, where a random draw selects the auction period relevant for payment. These design features make the auction environment closer to the -essentially static- features of the theoretical model upon which our experiment is based.

Section 4 reports our experimental results. In Section 4.1, where we report summary statistics of subjects' bidding behavior, we observe that our two treatment conditions yield a stark difference in behavior: when the relative weight on the rebate is high subjects bid more aggressively and closer to equilibrium. This is because the score/rebate elasticity is higher in the treatment in which the weight of the rebate is high. We also detect a stark difference in terms of allocative efficiency between the two treatments. In Table 1, for each treatment, we compute the relative frequency with which the auction has been awarded to each group member, ranked according to her relative efficiency, with RANK1 (RANK5, respectively) indicating the bidder with the highest (lowest) pseudotype within the group.

As Table 1 shows, when quality has a higher weight than price, 95 % of the auctions are awarded to the most efficient player (RANK1); when the rebate has a higher weight, this percentage drops to 43 %. In sum, our descriptive statistics point towards a 51.52 % higher probability of getting the most efficient outcome when the weight of quality in the scoring mechanism is high rather than low. And this difference is observed *despite* the higher noise detected in treatment which favours quality over price (see Fig. 4 below).

While this "allocative" efficiency measure captures the relative frequency with which the most efficient pseudotype wins the auction, it does not show how much relative welfare is gained (or lost), depending on whether (or not) the winner (compared with the second highest bid, or even participants with a lower score) is awarded the contract. This is the reason why in Section 4.2.1 we put forward an alternative proxy, measured as the relative share of efficiency, i.e., the difference between the score of the winner and the score of random participant in the auction relative to the highest possible difference (that is, if the winner were the highest pseudotype) characterizing any specific play. As Fig. 5 shows, also using this alternative efficiency proxy does not alter our main finding: The treatment which puts higher weight on quality is far more efficient at delivering against its own objective function than is the lower quality weighting against its own objective. Section 4.2.1 also considers an alternative approach to welfare analysis, tracking the buyer's surplus across treatments. Here, we find that despite a lower efficiency being delivered against its own objective function, the buyer still enjoys greater efficiency using the lower-weight-on-quality scoring rule to deliver against a lower-weight on-quality objective function than

Table 1
Distribution of winners by efficiency and treatment.

Auction winner	Relative frequencies		
	High weight on quality	High weight on rebate	Total
RANK1	94.95	43.43	69.19
RANK2	5.05	42.42	23.74
RANK3	0.00	11.11	5.56
RANK4	0.00	3.03	1.52
RANK 5	0.00	0.00	0.00
Total	100.00	100.00	100.00

by misreporting their objective function and using the higher-quality scoring rule. In other words, a truthful representation of preferences in the scoring rule definitively benefits the buyer.

The striking difference in efficiency is probably due to multiple factors, which may include -among others- auctions features and the impact of the latter on bidding behavior, as well as behavioral effects due to individual-specific characteristics. This suggests a more sophisticated econometric exercise whose aim is to disentangle the "direct" efficiency effect of a treatment change (i.e., the one which is due to the difference in the strategic characteristics of the two alternative mechanisms) from the "indirect" effect (i.e., the one that depends upon the level of the deviations from equilibrium that may be also influenced by the treatment). The "mediation analysis" (Imai et al., 2011) of Section 4.2.3 is carried out precisely to identify direct from indirect effects in the determination of efficiency and yields two main conclusions. First, the direct and indirect effects point in opposite directions, favouring (hampering, respectively) efficiency in the high (low, respectively) weight on quality treatment. Second, the direct effect outweighs the indirect one, which justifies the overall difference in efficiency in favor of the high-quality treatment.

The analysis of Section 4.2.3 ascribes most of the difference in efficiency between the two treatments to the "direct" effect, without pointing to any specific behavioral content for such a difference. On the other hand a standard behavioural approach would point out two "usual suspects", often invoked in behavioural auction theory: risk aversion and "noise", here defined as perturbation -for whatever reason- from equilibrium play. This is the reason why, in Section 4.3, we conclude our empirical analysis by estimating the maximum-likelihood quantal response equilibrium (QRE, McKelvey & Palfrey, 1995) induced by our data removing the assumption of risk neutrality we impose -as it is standard in auction theory- in Section 2. Three alternative models are estimated: *i)* one in which risk aversion is allowed to vary across treatments, *ii)* one in which we impose the same risk aversion parameter, independent of the treatment and *iii)* one in which we impose risk neutrality, exactly as in Section 2. The results of our structural estimations confirm that (subjects are risk averse and) noise is higher in the treatment that primes quality over price and for bidders with a higher pseudotype, although differences by treatment -exactly like in the mediation analysis- are not always significant.

Finally, Section 5 concludes, followed by appendices containing the proof of Proposition 1 (Appendix A), a more detailed account of our econometric strategy (Appendix B), supplementary statistical evidence (Appendix C) and the experimental instructions (Appendix D).

2. Theory

We consider a highest-score (procurement) auction whereby a risk-neutral bidder i submits a quality-rebate pair, (q, r), which is ranked according to the following linear scoring rule:

$$S_i(q,r) = (1-\gamma)q + \gamma\, r, \tag{1}$$

where $\gamma \in \{1/3; 2/3\}$ in our experimental implementation. Normalizing the reserve price to one, player i gets a payoff of

$$\pi_i(q,r) = \begin{cases} \dfrac{1-r-c(q))}{n^*} \ if \ S_i(\cdot) = \max_j(S_j(\cdot)), \\ 0 \ otherwise, \end{cases} \tag{2}$$

where $n^* \geq 1$ identifies the number of winners (in case of ties). By analogy with our experimental conditions, this section parametrizes the cost function as $c(q) = \frac{1}{4} + \frac{3}{4}q^2$.

A strategy for bidder i is a function $r: [0,1] \rightarrow [0,1]$ that maps each bidder's privately observed quality into a rebate. A symmetric Bayes-Nash equilibrium (BNE) is a vector of identical strategies, $(r(q))$, such that each bidder maximizes her expected payoff under the constraint that $0 \leq r(q) \leq 1 - c(q)$. In other words, by design, bidders can neither bid above the reserve price nor get negative profit.

In a standard lowest-price auction -where bidders privately receive iid signals about their production costs and only submit a price for the procurement contract- a symmetric equilibrium can be characterized by assuming that the bidding function is strictly increasing in production costs (that is, in bidders' types). Consequently, in equilibrium, winning probabilities coincide with the probability that any bidder has drawn the lowest cost. This is not the case of our scoring auction where, to derive a BNE, we mimic the approach pioneered by Che (1993), whereby bidders are characterized by "pseudotypes", which allows to rank bidders according to their winning probability.

Our definition of pseudotype has, though, to be adapted to the adverse-selection nature of our problem. Define type-q bidder's *pseudotype* or *potential score* as $s_\gamma(q) \equiv \gamma(1 - c(q)) + (1 - \gamma)q$, which corresponds to the score obtained by submitting $r_{\max}(q) = 1 - c(q)$, which yields the bidder zero profit.[7] The potential score coincides with the one that type-q bidder would get under *perfect and symmetric information*. That is, had the buyer perfect information about bidders' quality vector she would be able to make each bidder a financial offer to leave him with zero profit. The buyer would award the contract to the bidder with the *highest* potential score, ensuring the first-best allocation of the contract. The bidder with the highest potential score will represent our efficiency benchmark against whom we measure the winner's "observed" score in each repetition of the experiment.

In what follows, we will show how to use the potential score to derive a Bayes-Nash equilibrium (BNE) of the scoring auction. For the time being, let us assume, by analogy with Che (1993), that the higher the pseudotype, $s_\gamma(q)$, the higher the probability for a player with type q to win the auction when the financial weight parameter in the scoring rule is γ.

As shown in Fig. 1, depending on the value of γ, $s_\gamma(q)$ may or may not be monotonically increasing in q. More precisely, $s_\gamma(q)$ is strictly increasing in q if and only if $\gamma \leq \frac{2}{5}$, that is, when the weight associated to the financial score is sufficiently low, which is true in our experiment only when $\gamma = \frac{1}{3}$. In this case, the weight of quality evaluation in the scoring function is sufficiently high so as to make the bidder with the highest q to be the most likely winner. When $\gamma > \frac{2}{5}$, $s_\gamma(q)$ has an interior maximum, $q^* = \frac{2(1-\gamma)}{3\gamma}$. In particular, $q^* = \frac{1}{3}$ when $\gamma = \frac{2}{3}$ (our alternative treatment).

Proposition 1. If $r^*_\gamma(q)$ denotes the symmetric BNE of our scoring auction with weight equal to γ, then

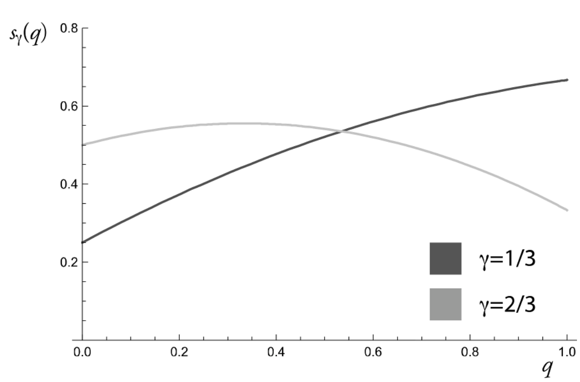


Fig. 1. Potential score function $s_\gamma(q)$ in the two treatments.

$$r^*_\gamma(q) = \max\left\{\frac{1}{\gamma}\left[s^*_\gamma(s_\gamma(q)) - (1-\gamma)q\right], 0\right\} \text{if } \gamma = 1\Big/3 \text{ and} \tag{3}$$

$$r^*_\gamma(q) = \frac{1}{g}\left[s^*_\gamma(s_\gamma(q)) - (1-\gamma)q\right] \text{if } \gamma = 2\Big/3, \tag{4}$$

where $s^*_g(s_\gamma(q)) = \frac{1}{H_g(s_\gamma(q))}\int_{\underline{s}_\gamma}^{s_\gamma(q)} y h_g(y)dy$, with $H_\gamma(s) = G^4_\gamma(s)$, $h_\gamma = H_\gamma'(s)$ and $G_\gamma(s)$ is the c.d.f. of the random variable s and $\underline{s}_\gamma = \min_{q\in[0,1]}[s_\gamma(q)]$ is the lower bound of the potential score distribution.

Proof. See Appendix A.∎

While relegating the proof of Proposition 1 to Appendix A, it may be instructive, at this point, to sketch the intuition behind our result. Following Che (1993), this is obtained by showing that our scoring auction is strategically equivalent to a first-price *selling* auction in which bidder i observes a signal s (his pseudotype) and submits a score, $\sigma_\gamma{}^*(s)$. At equilibrium, the submitted score $\sigma_\gamma{}^*(s) \leq s$ as rational bidders get positive profit by reducing the value of the rebate below its maximum level, that is, $r \leq r_{\max}(q)$. The bidding functions $\sigma_\gamma{}^*(s)$ associated with our treatments are reported in Fig. 2. Notice that, coherently with the results in a "standard" first-price auction, the bidding function $\sigma_\gamma{}^*(s)$ lays below the 45-degree (dotted) line, as each bidder shades his bid below his value (that is, his pseudotype).

The explicit forms of either $\sigma_\gamma{}^*(q)$ or its strategic equivalent rebate function, $r_\gamma{}^*(q)$, are complex and uninstructive, but we plot them in Fig. 3 for both values of γ (1/3 and 2/3) used in the experiment.

Given that the equilibrium bidding function $r_\gamma{}^*(q)$ is derived from the equilibrium of an "equivalent" first-price auction, $\sigma_\gamma{}^*(s)$, it is immediate to realize that, in equilibrium, (i) bidders with the same *potential* score *(s)* are expected to submit the same score $\sigma_\gamma{}^*(s)$ and (ii) the winner is the bidder with the highest signal, $s_\gamma(q)$.

Consider the graphs depicted in Fig. 3. First, notice that the closer the equilibrium bids (solid line) to the zero-profit bids (dotted line) the lower the expected profit in case of winning. Consistently with intuition, when the weight of the rebate in the scoring rule is high ($\gamma = 2/3$), the submitted rebates are higher than in the case of $\gamma = 1/3$ for almost any q (precisely, for any $q > 0.05$). Second, when $\gamma = 1/3$ the most likely winner is the type with the highest q, because the scoring rule greatly rewards quality. It takes quite high a difference between two bidders' submitted rebates to more than compensate the score gap induced by different quality levels. Hence, in equilibrium, the types with high quality can "safely" increase their expected profit (by lowering the rebate) without considerably reduce their winning chances. In other words, the gap in the potential score among bidders with different quality levels makes it harder (relatively to the case of $\gamma = 2/3$) for less

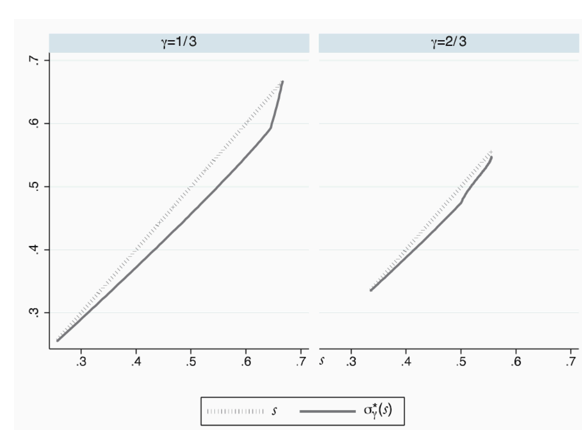


Fig. 2. Private signals (pseudotypes) s and equilibrium bids $\sigma_\gamma{}^*(s)$.

[7] In the remainder of the paper, "pseudotype" and "potential score" will be used interchangeably.

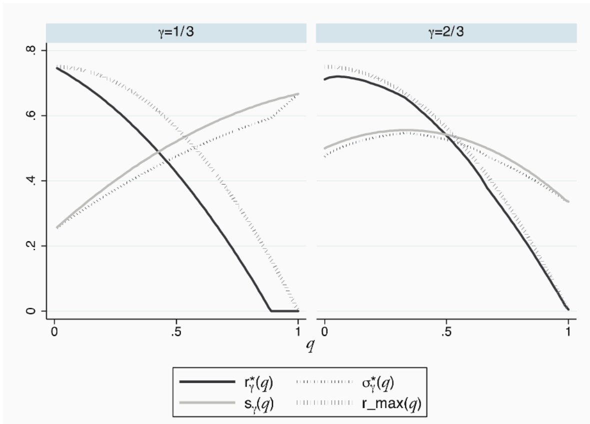


Fig. 3. Equilibrium Analysis. The bidding functions are plotted both in terms of the submitted rebate r and the obtained score σ and compared with the maximum potential rebate/score (dotted lines).

efficient bidders to overbid more efficient competitors. This also helps us understand why $r_{1/3}{}^*(q)$ becomes flat above a certain threshold (approx. 0.88 with our parametrization): bidders with sufficiently high quality anticipate to be awarded a high score for quality and would then submit a discount below 0 (i.e., a price higher than the reserve price), which is not allowed by the rules of the game.

The opposite is true when $\gamma = 2/3$. As shown in Fig. 1, $s_{2/3}(q)$ is not monotonic, which shortens the length of the support of the random variable $s_{2/3}(q)$. This makes bidders closer in terms of efficiency, thus increasing their incentive to compete more aggressively and submit higher rebates. In fact, the higher weight of the rebate in the scoring rule allows bidders with lower quality to compensate their gap in quality by increasing their financial score, which is made possible by lower production costs.

3. Experimental design

3.1. Sessions

Four experimental sessions were conducted at the centro di economia sperimentale a roma est (CESARE), at LUISS Guido Carli Roma. A total of 90 students were recruited among the undergraduate population of LUISS Guido Carli using the ORSEE recruiting system (Greiner, 2004), with no particular bias in favour of students from the Departments of Economics and Finance or Business Administration and Management. All sessions were "gender balanced", with approximately the same number of male/female subjects.

Experimental sessions were computerized. Instructions were read aloud, and we let subjects ask about any doubt they may have had.[8] At the end of each session, subjects were asked to compile an extensive debriefing questionnaire (see Section 3.4 below), before receiving –in cash and privately- their monetary winnings.

3.2. Matching

In each session, subjects are randomly sorted into matching groups (cohorts) of 5 participants, with subjects from different cohorts never interacting with each other throughout the experiment.[9] Matching groups remain constant throughout the experiment, with no feedback until the very end, where the period relevant for payment is publicly drawn and monetary payoffs are determined.[10]

For each treatment $\gamma \in \left\{\frac{1}{3}, \frac{2}{3}\right\}$, subjects play 11 rounds of a procurement auction characterized in which an iid random draw without replacement determines the value of $q \in \left\{\frac{k}{10}\right\}, k = 0, 1, \ldots, 10$, each player's idiosyncratic quality, randomized across periods to make sure that every bidder faces each and every feasible quality level during the experiment. This permits to elicit the entire bidding function, $r(q)$, of each participant.

3.3. Financial rewards

Subjects receive € 10 just to show up. The value of the contract for the winner has been set at € 10. The experimental design prevents participants to make losses in the auction: they cannot set a rebate bid high enough so that their monetary profits, net of the quality cost, are below zero. While constraining subjects' bidding space is quite common in auction experiments, what makes our design peculiar is that players' action space is endogenous (and varying, across players and rounds) in that it depends on the random realization of the individual private cost. While this design choice avoids the influence of classic behavioral biases (such as, loss aversion) and can be considered as a "minimal rationality requirement" imposed by the experimenter, it has the drawback to create a possible confound when it comes to a give a quantitative assessment of the impact of out-of-equilibrium play (as we do in Sections 4.2 and 4.3).[11]

For payment, we use a random lottery incentive protocol by which we draw one round at random and add to all participants their monetary payoffs in that selected round. Average monetary winnings were € 12, for a 60′ experiment, including debriefing and payment.

3.4. Debriefing

At the end of each session, subjects are asked to answer a detailed questionnaire from which we elicit proxies of their observable heterogeneity. As it turns out, one of the key variables used in Section 4.2.3 for our regression analysis is derived from the well-known Cognitive Reflection Test (CRT, Frederick, 2005). The CRT is a simple test of a quantitative nature especially designed to elicit the "predominant cognitive system at work" in respondents' reasoning:

1. A bat and a ball cost 1.10 dollars. The bat costs 1.00 dollars more than the ball. How much does the ball cost? (Correct answer: 5 cents).
2. If it takes 5 machines 5 min to make 5 widgets, how long would it take 100 machines to make 100 widgets? (Correct answer: 5 min).
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? (Correct answer: 47 days).

The CRT provides not only a measure of cognitive ability, but also of impulsiveness and, possibly, other individuals' unobservable characteristics. In this test, the "impulsive" answer (10, 100 and 24, respectively) is shown to be the modal answer (Frederick, 2005). These answers, although incorrect, may have been selected by those subjects who do not think carefully enough. Following Cueva et al. (2016), we partition individuals into three groups. *Impulsive* subjects answer the erroneous intuitive value at least in two questions, *reflective* ones answer

[8] The experiment was programmed and conducted with the software *z-Tree* (Fischbacher, 2007). A copy of the experimental instructions can be found in the Appendix.

[9] Due to the absence of some participants, we lost one cohort in session 2 and 3.

[10] Given this design feature, we shall read the data under the assumption that the history of each individual subject corresponds to an independent observation.

[11] We thank an anonymous referee to have raised this issue and discuss it more in detail at the end of Section 4.3.

correctly at least two questions, and others are the *residual group*. CRT group identifiers have been used as instruments in the two-step regression analysis of Section 4.2.3 (see Appendix B for details).

4. Results

4.1. Bidding behavior

Fig. 4 tracks average and equilibrium bidding functions by treatment, together with the treatment score pseudotypes. As expected, when the scoring rule puts more weight on quality (that is, when $\gamma = 1/3$), players submit, on average, lower rebates. This simple evidence lets us conclude that submitted bids correctly follow the incentives induced by the two treatments and, for all quality levels, players bid less aggressively when the scoring rule favours quality with respect to price. We also notice that the dispersion of bids around the average is higher at low quality levels since, for higher quality levels, bids are constrained by the rule that prevents losses.[12]

From Fig. 4 where $\gamma = 2/3$ and $\gamma = 1/3$ are compared, two distinct empirical evidences result: when $\gamma = 2/3$, (i) a smaller average difference between predicted and observed bids; (ii) a smaller dispersion of the bids. In Sections 4.2 and 4.3 we shall look at both these stylized facts in more detail, associating deviation from equilibrium to two classic behavioural phenomena: risk aversion and noise.

4.2. Welfare analysis

The evidence provided in Fig. 4 -that individuals playing auctions with higher weight on the rebate play closer to equilibrium- *could* support the conclusion that auctions with $\gamma = 2/3$ may be characterized by higher efficiency. However, we already know that this is not the case: overall, the descriptive statistics of Table 1 deliver a ballpark estimate of a 52 % higher allocative efficiency when the weight of quality in the scoring rule is high ($\gamma = 1/3$). As discussed in the introduction, the allocative efficiency measure shown in Table 1 does not consider the amount of the utility achieved by the buyer and does not compare the realized welfare with that of second-best (or even less efficient) participants. To this purpose, in this section we refine our welfare analysis by looking at the buyer's surplus and then argue for an alternative proxy for efficiency.

4.2.1. Buyer's surplus

We first consider the *loss* of surplus accruing to the buyer compared (a) to the first best (i.e., bidders submitting the highest allowed rebate) and (b) to the BNE. More precisely, for each auction round, we compute the difference between the buyer's utility – that is, the winning bidder's score – (a) when the most efficient bidder submits a zero-profit rebate or (b) when bidders submit BNE bids and the observed score. Fig. 5 shows the distribution of both our measures of loss of the buyer's surplus under both treatments. The first stark, albeit unsurprising, evidence is that when $\gamma = 1/3$ in most cases there is no loss for the buyer with respect to the first best (Fig. 5a), that is, the buyer achieves the highest possible utility. This merely reflects our previous result on allocative efficiency: when the weight on quality is high, the highest-type bidders easily win the auction with no rebate (as previously emphasized, they are constrained to make no rebate by their high production cost and cannot bid a higher price given the ceiling set by the reserve price). Furthermore, many auction rounds yield the buyer a higher surplus than under the BNE behaviour (a negative loss is shown in Fig. 5b), which implies that the winning bidder bids more aggressively than when submitting the BNE bid.

When $\gamma = 2/3$, instead, we observe high density of cases where the buyer's realized surplus is very close to the equilibrium surplus as well as many cases of surplus loss (and lower probability of sizeable surplus gain) with respect to BNE compared to the $\gamma = 1/3$ case. If we put together this evidence with Figs. 4 (and 6), we can confirm that, on the one hand, more bidders underbid in the case $\gamma = 2/3$; but, on the other hand, winning bidders seem to underbid to a lesser extent.

Before moving to a more formal analysis of the buyer's achieved efficiency under the two different scoring rules, we carry out a further exercise to better understand how effective the two scoring rules are in pursuing the buyer's objective. In our experimental setting, we have assumed that the buyer sets a scoring rule that "truthfully" reveals her own utility function. In other words, she sets a value of γ that perfectly represents her relative price/quality preference. Therefore, we abstract away from the (mechanism design) problem of setting the "optimal" scoring rule given the buyer's utility function. In spite of the "naïf-buyer" assumption it is instructive to look at how much utility a buyer with utility characterized by $\gamma = 1/3$ (resp. 2/3) would achieve if she were to set a scoring rule with $\gamma = 2/3$ (resp. 1/3). In other words, we compute the buyer's utility under the assumption of a misrepresentation

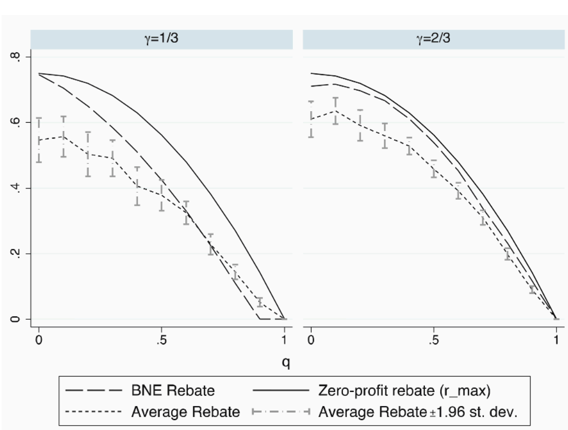
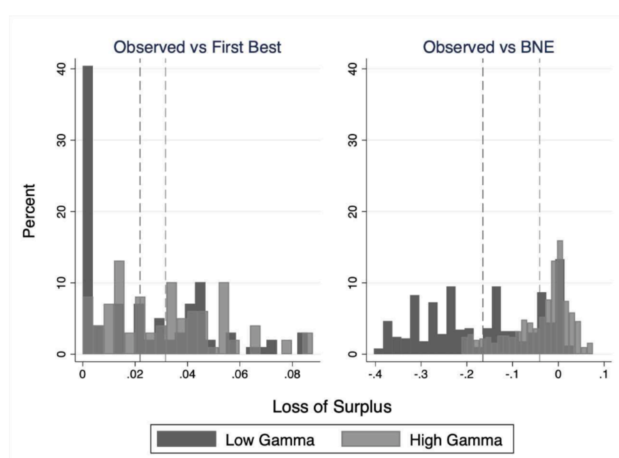


Fig. 4. Equilibrium and empirical bidding functions by treatment.



a) *b)*

Fig. 5. Distribution of the loss of realized buyer's surplus with respect to (a) the first best, i.e. the case where the most efficient bidder of each round bids zero-profit rebate and (b) the case where all bidders submit BNE rebate. Black (gray) vertical dashed lines correspond to the average surplus loss for the low (high) gamma treatment, respectively.

[12] As a result, when $q = 1$, $c(q) = 1$, i.e., players are forced to bid a rebate equal to zero.

of her utility function. Fig. 6 shows the result, by providing a boxplot representation of the distribution of the surplus obtained by two different buyers (with a different utility function) under two alternative scoring rules. In other words, for each "alternative" buyer, we compute the achieved surplus in the auction rounds carried out under both treatments (i.e., under both scoring rules). More specifically, Fig. 6a shows that the surplus of a buyer with objective function $\gamma = 1/3$ is considerably higher when she adopts a truthful scoring rule (i.e., with $\gamma = 1/3$) rather than the "alternative" rule (i.e., with $\gamma = 2/3$). The same holds for a buyer with utility function $\gamma = 2/3$, as shown in Fig. 6b. This result, albeit unsurprising, is still relevant under two different perspectives. First, it confirms the robustness of our experimental results, in the sense that it makes it evident that the participants' behavior under alternative mechanisms does affect the buyer's utility sizeably. Second, it shows how a "incorrect" choice of the scoring rule may hamper the buyer's objective. This is relevant for the procurement practice, insofar as in some legal environments the procurement law or regulation set constraints to the price/quality weighting that can be used by procuring entities – let alone that many procurers may find difficult to express their own preferences with respect to price and quality in terms of relative weightings in a scoring rule.

Finally, the comparison between the two sides of the figure also starkly enlightens a relevant implication of our setting: When the buyer primes quality ($\gamma = 1/3$ in her utility function), but sets the "wrong" scoring rule ($\gamma = 2/3$), she undergoes a huge variability in her achieved utility. This goes back again to the one of main findings of our experiment, that is, when the scoring rules weights more price bids, more participants with different types have concrete chance to win, because a difference in q-types leads to a (relatively) smaller difference in production cost. This means that the auction outcome depends more heavily on bidding behaviour (submitted bids) than on drawn bidders' types (quality). Yet, the higher dispersion in the winner's quality (which characterizes the $\gamma = 2/3$ scoring rule) heavily affects the realized surplus of a buyer that primes quality.

4.2.2. *Efficiency*

The welfare analysis sketched so far is based on the buyer's surplus and it focuses on the winning bids only. In order to investigate more in depth the empirical evidence arising from all participants' behavior, we provide an alternative definition of efficiency, which measures the extent to which the buyer achieves higher efficiency by assigning the contract to the auction winner, *compared with a randomly selected bidder* within the current matching group. Our alternative efficiency measure, $\boldsymbol{\eta}$, is defined as

$$\boldsymbol{\eta} = \frac{S(q_W, r_W) - S(q_A, r_A))}{\bar{s} - S(q_A, r_A))} = \frac{S(q_W, r_W) - \frac{1}{5}\sum_{i=1}^{5} S(q_i, r_i))}{\bar{s} - \frac{1}{5}\sum_{i=1}^{5} S(q_i, r_i))} \in [0, 1], \tag{5}$$

where, within each matching group and period, $S(q_W, r_W)$ is the observed score obtained by the auction winner, $\bar{s}$ is the optimal score, i.e. the score of the highest pseudotype among bidders and $S(q_A, r_A)$ is the average score of that specific matching group and round.

Fig. 7 reports the distribution of $\boldsymbol{\eta}$ by treatment. Again, we detect a stark contrast between treatments where, when γ is low, $\boldsymbol{\eta}$ is about 1 for more than 40 % of all observations.

Searching for reasons for such a strong treatment welfare effect, in Fig. 8 we refine the evidence of Fig. 4 by plotting average and equilibrium bidding functions by treatment isolating, for each cohort and period, the auction winners (panel *a*) from the rest (panel *b)*. Here we observe a stark difference in behaviour across treatments: while winners tend to overbid (with the respect to the BNE benchmark) when quality

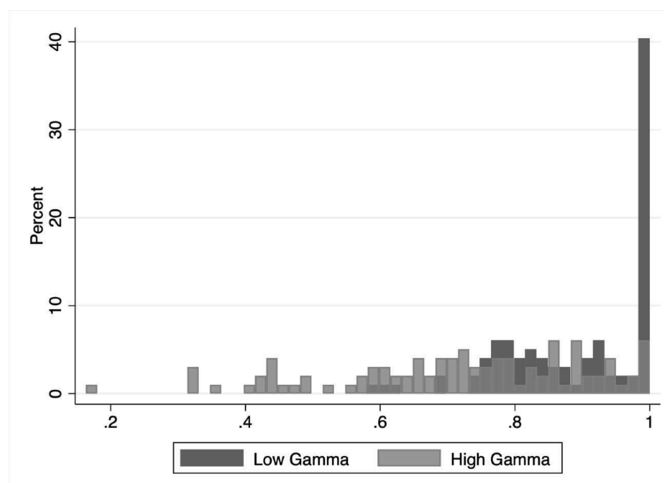


Fig. 7. Distributions of $\boldsymbol{\eta}$ by treatment.

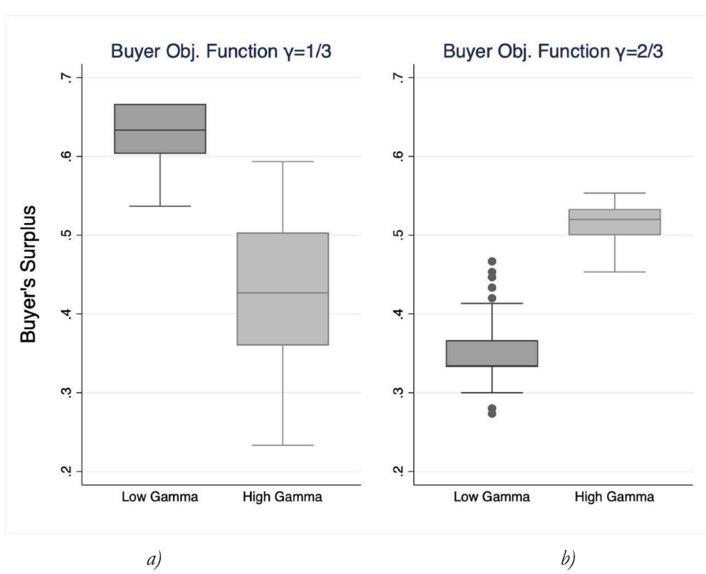
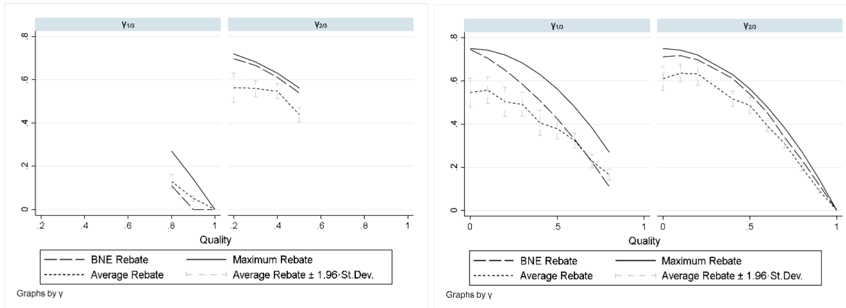


Fig. 6. Distribution of the observed buyer's surplus when her utility function is characterized by $\gamma = 1/3$ (panel a) and $\gamma = 2/3$ (panel b). In both cases, the distribution is provided considering the outcome of the auction rounds where the scoring rule was characterized by $\gamma = 1/3$ (left side of each panel) and $\gamma = 2/3$ (right side).



a) *b)*

Fig. 8. Equilibrium and empirical bidding functions for the most efficient pseudotypes (panel *a)* compared to the rest (panel *b)*.

weighs more than price, they underbid in the alternative scenario. Similarly, auction "losers" underbid more heavily in the treatment that primes quality over rebate.

Notice that, by (5), $\frac{d\eta}{dr_W} = \frac{\gamma}{\bar{s}-S(q_A,r_A)} > 0$ and $\frac{d\eta}{dr_A} = -\frac{\gamma(\bar{s}-S(q_W,r_W))}{(\bar{s}-S(q_A,r_A))^2} < 0$. In other words, an increase in the rebate of the winner (resp., a random participant), fosters (lowers) efficiency. As Fig. 8 shows, both these effects favor efficiency in the treatment that primes quality over rebate.

This intriguing evidence suggests further empirical analysis. First of all, while separating auction winners from the rest makes perfect sense when measuring the relative efficiency of the auction result (to the extent that it is the winner that sets the auction's efficiency, however measured), it also creates obvious endogeneity problems in that some bidders win the auction *exactly because* they behave differently from the others. On the other hand, it may well be that the fact of being exogenously assigned the most efficient pseudotype (which predicts winning the auction almost perfectly when γ is low, see Table 1) may induce differences in behavior at the treatment level. In Sections 4.2.3 and 4.3 we follow this conjecture by way of two complementary approaches. In Section 4.2.3 we perform a "mediation analysis" (Imai et al., 2011) with the aim of separating the "direct" treatment effect on efficiency due to the difference in the allocation mechanism, from the "indirect" effect, due to out-of-equilibrium behavior. Finally, in Section 4.3 we structurally estimate the QRE induced by our data, with the aim of identifying how the two structural parameters of the model, -ρ measuring risk aversion and μ measuring behavioural noise- depend on the treatment and the random allocation of pseudotypes.

4.2.3. Mediation analysis

In what follows, we disentangle the "direct" treatment effect on efficiency –which is due to the different strategic characteristics of the two treatments- from the "indirect" one –which is due to out-of-equilibrium behavior under the two treatments. To this aim, our estimation strategy is based on the following claims:

1. *for any given* deviation from equilibrium, γ has a *direct* effect on η through the shape of the potential score function (i.e., the strategic properties of the different treatment conditions);
2. γ also exerts an *indirect* effect by affecting the magnitude of bidders' "trembles" around equilibrium, which may also depend on auction and matching group specific characteristics (such as the 5 bidders' realized quality and individual heterogeneity).

Fig. 9 illustrates these two effects upon which we design our estimation strategy.

If players were always to play the equilibrium, we would always observe an efficient allocation. Hence, the question concerning to what extent γ determines -either directly or indirectly- relative efficiency in the auction outcome -proxied by η- makes only sense out of equilibrium. In this respect, for *any given* deviation from equilibrium, the direct effect explains to which extent γ - that is, the characteristics of the underlying game- affects the actual realization of η; by contrast, the indirect effect captures the impact of γ on efficiency via the level of noise which can be ascribed to a change in the treatment conditions.

Identifying the direct from the indirect effect is also relevant for the auction designer. If, say, the direct effect did outweigh the indirect one, then the auction designer would be in the position to select which game is more likely to generate her preferred outcome by simply looking at the equilibrium properties of alternative game-forms, which is the standard practice of mechanism design. Conversely, if the indirect effect turned out to be stronger, the auction designer should also take into account behavioural and context-specific factors, which may substantially complicate his task.

With these premises in mind, we adopt a two-stage least-squares random-effect estimator to quantify the direct and indirect effects of γ on efficiency. Our estimation strategy (see Appendix B for details) relies on the following 2-step procedure:

- Step 1. We regress the difference between observed and equilibrium bids of RANK1 and non-RANK1 players on: i) our treatment variable, γ, by way of a binary index, positive when $\gamma = 2/3$; ii) proxies of the auction-specific randomized quality levels and iii) identifiers of the CRT partition (see Section 3.4). Step 1 allows us to quantify the value "A" in Fig. 7 as the marginal impact of γ on the observed "trembles" around equilibrium for two groups: RANK1 and non-RANK1 players.
- Step 2. We regress η on i) the predicted deviations from equilibrium of RANK1 and non-RANK1 players estimated in Step 1; ii) our treatment variable, γ, and iii) the same proxies ii) used in Step 1. Step 2 allows us to disentangle the value "C" (as the marginal impact of γ on efficiency) from the value "B" of Fig. 7 (as the marginal impact on efficiency of the predicted bidders' trembles around equilibrium).

The value "C" from Step 2 represents the *direct* effect of our treatment variable on efficiency, that is, how the potential score function characteristics would affect η if players made identical mistakes under both treatments. The product of values "A" (from Step 1) and "B" (from Step 2) represents, instead, the *indirect* effect of γ on efficiency. Detailed results from the estimation strategy are reported in Appendix B. Table 2 reports only the estimated coefficients of the direct/indirect effects, together with their sum.

As Table 2 shows, we find an overall negative and significant treatment effect on efficiency, that is the average efficiency measure of auctions characterized by higher weight on price is significantly lower than that of auctions characterized by higher weight on quality. This result is mainly due the estimated "direct" effect, once the effect on players' out-of-equilibrium behavior (the "indirect" effect) has been controlled for. In other words, when $\gamma = 1/3$, the score component determined by bidders' quality endowment plays a more significant role than bidders' submitted discount in determining the auction outcome

In particular, the "direct" effect is found to be significantly negative: when $\gamma = 2/3$, the shape of the score function does not monotonically

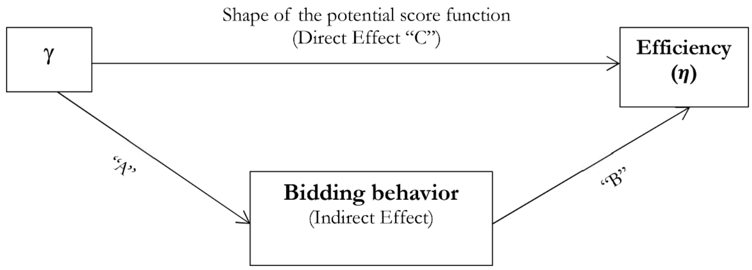


Fig. 9. Direct and indirect effect of γ on efficiency.

Table 2
Estimation of the direct/indirect effects.

	Marginal impact	*p*-value
Direct effect	−0.289	0.085
Indirect effect	+0.150	0.338
Total	−0.139	0.022

increase in quality. Hence, even small "trembles" of bids around equilibrium are potentially more able to reduce the likelihood that RANK1 plyer wins the auction and, in general, the auction's efficiency, compared to a $\gamma = 1/3$ situation. Instead, the "indirect" effect of $\gamma = 2/3$ on efficiency turns out to be positive. Indeed, the results from step 1 suggest that, when $\gamma = 2/3$, bids' deviations from the equilibrium of RANK1 players are (on average) significantly lower (since the zero-profit and the BNE rebates are much closer than when $\gamma = 2/3$). This negative effect of $\gamma = 2/3$ on the observed deviations from the equilibrium, combined with its negative "direct" effect via the shape of the potential score function, makes the overall "indirect" effect of $\gamma = 2/3$ positive, although not significant ($p = .338$, see Table 2).

4.3. *Structural estimation*

The aim of this section is to measure the impact on behavior of two natural ingredients of models of choice under risk and uncertainty. *risk aversion* and *behavioral noise*. Risk aversion is a common finding in experimental auctions and its presence may yield more overbidding (compared to the risk-neutral equilibrium prediction) in that risk aversion increases the opportunity cost of not winning the auction (see Krishna, 2009). Since (as it is standard in applied auction theory) we neglect to consider risk aversion in theoretical discussion of Section of Section 2, we are still interested in assessing what we miss using our first-order approximation. On the other hand, noise is another key ingredient of behavioral models, and we have already discussed the identification of a treatment effect in the observed deviation from equilibrium in Section 4.3. In this respect, compared with risk aversion, the impact of noise on efficiency is more subtle, in that it clearly depends on whether noise leads to overbidding rather than underbidding, and how it is related with the play of auction winners (or, higher pseudotypes), rather than losers (or, lower pseudotypes).

In our structural estimate we follow closely Goeree et al. (2002) by estimating the QRE induced by our data under the assumption that bidders' utility is a Constant Relative Risk Averse (CRRA) transformation of their monetary payoff, $u(y) = \frac{y^{1-\rho}}{1-\rho}$ and the probability of submitting any specific rebate, $\hat{r} \in \left\{0, \frac{1}{10}, \dots, 1 - c(q)\right\}$, is as follows:

$$P_i(\hat{r}|q) = \frac{\exp[\mu\ u(1 - c(q) - \hat{r})w(\hat{r}|q)]}{\sum_{r=0}^{1-c(q)} \exp[\mu\ u(1 - c(q) - r)w(r|q)]}, \tag{6}$$

where the error parameter, $\mu > 0$, determines the sensitivity of choice probabilities with respect to payoffs: when $\mu \to \infty$, the option with the highest expected utility is chosen for sure, while in the limiting case of $\mu \to 0$, behavior becomes essentially random. We identify with $w(\hat{r}|q)$ strategy $\hat{r}$'s winning probability, that is, the probability that the associated score is the highest, given the other group members' QRE play.

Table 3 reports the estimates of ρ and μ for three nested models: in Model i) we condition the estimates of both parameters to depend upon γ and on a binary index, we call it RANK_1, which is positive if the bidder's pseudotype is the highest within her matching group. As we discussed earlier, this index is meant to capture structural changes in behavior due to the perception of being in an advantageous position in the auction. While changes in the values of μ may well depend on treatment effects (due, for example, to the fact that, players' pseudotype is not increasing in the quality when γ is high, this increasing the strategic complexity of the mechanism) it is less plausible to think of risk aversion -usually modeled as an intrinsic characteristic of each individual- may depend on the treatment, or contingencies along the sequence of play. This is why Model ii) constrains the estimation of ρ to a single constant, independent of the treatment and relative ranking within each individual auction. Finally, Model iii) -by analogy with Section 2, imposes Risk Neutrality (i.e., ρ equal to 0) to all subjects.

The structural estimates of Table 3 deliver a qualitative message which is robust across the various specifications of the model: *i)* subjects are risk averse (and risk aversion does not depend on treatment, or the relative advantage in the auction) and *ii)* γ (but, especially, RANK_1) reduces (enhances, resp.) noise. In this respect, the qualitative differences in bidding behavior we observe in Fig. 6 find a quantitative confirmation in the differences in the estimated noise across treatments and relative positions.

Fig. 10 tracks average bids conditional on quality comparing actual vs predicted behavior under model ii) of our QRE estimations by treatment. First, we notice that, when γ is low, the QRE prediction is able to capture the tendency to underbid -in comparison with the BNE prediction- in our data. In this case, this is mainly due to the presence of noise rather than risk aversion, to the extent to which the latter pushes behavior in the opposite direction. As for the QRE prediction when γ is high, this goes again in the direction of underbidding for high level of quality, although this behavior is only partially supported by the data.

Before we conclude, we acknowledge that the quantitative assessments of noise performed in this paper may be underestimated by the fact that, as we mentioned earlier, subjects are prevented by design to push their rebate above the level that, conditional on their individual cost, would lead to negative payoffs in case they win the auction. Although such behavior is rarely observed in auction experiments, we test the robustness of our results by estimating our model (relying on bootstrapping to make standard errors comparable across estimations) over reduced datasets: i) one in which we set $q = 0$ (and bids can vary over the entire feasible interval and ii) only considering $q \leq .5$ (where the effect of the constrained on rebates is reduced since costs ae comparatively low). Results are reported in Table C.3 and confirm our main finding: noise is higher in the treatment with high weight on quality, but treatment effects are never significant.

Table 3
Structural estimation.

		i) 3 rho		ii) 1 rho		iii) RN	
		Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
ρ	gamma_66	−0.067	(0.351)	0	N/A	0	N/A
	RANK_1	−0.071	(0.122)	0	N/A	0	N/A
	rho_cons	0.487	(0.333)	0.414***	(0.092)	0	N/A
μ	gamma_66	13.522	(13.135)	10.902	(11.775)	78.969***	(21.326)
	RANK_1	−1.759	(5.617)	−5.349**	(2.753)	−14.862***	(5.637)
	mu_cons	10.527	(8.259)	13.927**	(6.411)	35.437***	(11.014)
	Obs.	990		990		990	
	Log-lik.	−1491.4597		−1491.7206		−1507.8205	

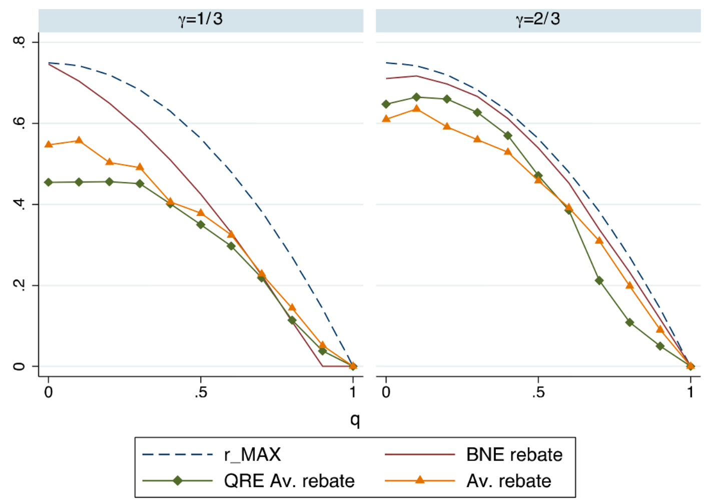


Fig. 10. Average QRE bidding by treatment.

5. Conclusions

Our experiment provides the mechanism designer with two complementary pieces of information -although confined within the very specific bounds of our parametric settings (a linear scoring rule, quadratic costs, just two weights, etc.…). First, more weight on rebate reduces noise, as out-of-equilibrium deviations are more costly (in terms of score) for the bidders. Second, more weight on quality yields higher efficiency, in spite of the higher level of the associated noise. It should be noticed, though, that the (quite natural, from a viewpoint of mechanism design) search for an "optimal γ" is well beyond the scope of this paper. This is because the latter is usually influenced by contextual factors specific of each tender and by the constrains put in place by the legislators. For instance, in Italy the national Law for Public Contracts used to make it mandatory to use *at least* a 70 %-weight on quality when public buyers wish to carry out a procurement procedure by using a scoring auction.[13] These considerations notwithstanding, our analysis allows us to conclude that i) the level of deviation from equilibrium (the *indirect effect*) varies with the weight associated with each dimension composing the score, and that ii) in the choice of the optimal weights the designer should take into account the differences in efficiency due to both –direct and indirect- effects. Nonetheless, our results suggest that large misrepresentations of the buyer's objective function in the scoring rule are unlikely to be optimal.

The most natural extension to this paper would be to look at a procurement environment in which -by analogy with Che (1993)- participants have to decide *both* the level of quality and the rebate. This could be implemented by considering bidders with heterogeneous (and privately observed) productivities who have to determine -simultaneously and independently- the quality and the price of their tender. Along these lines, Camboni et al. (2023) study an experiment based on scoring rule procurement auctions in which subjects simultaneously submit both quality and price, and a linear scoring rule similar to ours awards the contract to the highest score bid. They compare five different experimental treatments in which sellers *i)* bid only on one dimension (price or quality), ii) bid on both dimensions but the choice of one is constrained to be binary, or iii) they can bid, unconstrained, on both dimensions. In this sense, they can naturally rank their treatments in terms of their relative complexity. They also perform a structural analysis and find results that are consistent with ours (in particular, risk aversion does not depend on the treatment). They also find that noise increases with complexity, although their definition of complexity -upon which their design is built- is rather different than ours and results cannot be compared in a straightforward way.

Data availability

Happy to share the data.if the paper is accepted.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.socec.2023.102131.

References

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

[13] This provision was suppressed in the newly adopted Public Procurement Regulation, which entered into force on the 1st July 2023.

Bichler, M. (2000). An experimental analysis of multi-attribute auctions. *Decision Support Systems, 29*(3), 249–268.
Cason, T. N. (1995). An experimental investigation of the seller incentives in the EPA's emission trading auction. *The American Economic Review*, 905–922.
Cason, T. N., KN, Kannan, & Siebert, R. (2011). An experimental study of information revelation policies in sequential auctions. *Management Science, 57*(4), 667–688.
Che, Y. K. (1993). Design competition through multi-dimensional auctions. *RAND Journal of Economics, 24*, 668–680.
Chen-Ritzo, C. H., Harrison, T. P., Kwasnica, A. M., & Thomas, D. J. (2005). Better, faster, cheaper: An experimental analysis of a multiattribute reverse auction mechanism with restricted information feedback. *Management Science, 51*(12), 1753–1762.
Cueva, C., Iturbe-Ormaetxe, I., Mata-Pérez, E., Ponti, G., Sartarelli, M., Yu, H., et al. (2016). Cognitive (Ir)Reflection: New experimental evidence. *Journal of Behavioral and Experimental Economics, 64*, 81–93.
Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 24–42.
Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.
Fooks, Jacob R., Messer, Kent D., & Duke, Joshua M. (2015). Dynamic entry, reverse auctions, and the purchase of environmental services. *Land Economics, 91.1*, 57–75.
Camboni, R., Corazzini, L., Galavotti, S., & Valbonesi, P. (2023). *Bidding on price and quality: An experiment on the complexity of scoring auctions*. The Review of Economics and Statistics. in press.
Greiner B. (2004). *The online recruitment system ORSEE 2.0 - A guide for the organization of experiments in economics*, University of Cologne WP Series in Economics 10.
Goeree, J. K., Holt, C. A., & Palfrey, T. R. (2002). Quantal response equilibrium and overbidding in private-value auctions. *Journal of Economic Theory, 104*(1), 247–272.
Grimm, V., Kovarik, J., & Ponti, G. (2008). Fixed price plus rationing: An experiment. *Experimental Economics, 11*, 402–422.
Grimm, V., Mengel, F., Ponti, G., & Viianto, L. A. (2009). Investment incentives in procurement auctions: An experiment. In A Hinlopen, & T Norman (Eds.), *Experiments and competition policy* (pp. 267–300). Cambridge University Press.
Hailu, A., & Schilizzi, S. (2004). Are auctions more efficient than fixed price schemes when bidders learn? *Australian Journal of Management, 29*(2), 147–168.
Imai, I., Keele, L., Tingley, D., & Tamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review, 105*(4), 765–789.
Kagel, J. H., & Levin, D. (2002). Bidding in common-value auctions: A survey of experimental research. *Common Value Auctions and the Winner's Curse, 1*, 1–84.
Kagel J Levin, D. (2015). Auctions: A survey of experimental research. In JH Kagel, & AE Roth (Eds.), *The handbook of experimental economics – vol. 2*. Princeton NJ: Princeton University Press.
Krishna, V. (2009). *Auction theory* (2nd edition). Academic Press.
McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior, 10*(1), 6–38.
Santamaria, N. (2015). An analysis of scoring and buyer-determined procurement auctions. *Production and Operations Management, 24*(1), 147–158.
Strecker, S. (2010). Information revelation in multiattribute English auctions: A laboratory study. *Decision Support Systems, 49*(3), 272–280.