WP-AD 2018-02

Scoring Rules in Experimental Procurement

G.L. Albano, A. Cipollone, R. Di Paolo, G. Ponti and M. Sparro

Ivie

Working papers

Working papers

Working papers

Los documentos de trabajo del Ivie ofrecen un avance de los resultados de las investigaciones económicas en curso, con objeto de generar un proceso de discusión previo a su remisión a las revistas científicas. Al publicar este documento de trabajo, el Ivie no asume responsabilidad sobre su contenido.

Ivie working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication. Ivie's decision to publish this working paper does not imply any responsibility for its content.

La edición y difusión de los documentos de trabajo del Ivie es una actividad subvencionada por la Generalitat Valenciana, Conselleria de Hacienda y Modelo Económico, en el marco del convenio de colaboración para la promoción y consolidación de las actividades de investigación económica básica y aplicada del Ivie.

The editing and dissemination process of Ivie working papers is funded by the Valencian Regional Government's Ministry for Finance and the Economic Model, through the cooperation agreement signed between both institutions to promote and consolidate the Ivie's basic and applied economic research activities.

La Serie AD es continuadora de la labor iniciada por el Departamento de Fundamentos de Análisis Económico de la Universidad de Alicante en su colección "A DISCUSIÓN" y difunde trabajos de marcado contenido teórico. Esta serie es coordinada por Carmen Herrero.

The AD series, coordinated by Carmen Herrero, is a continuation of the work initiated by the Department of Economic Analysis of the Universidad de Alicante in its collection "A DISCUSIÓN", providing and distributing papers marked by their theoretical content.

Todos los documentos de trabajo están disponibles de forma gratuita en la web del Ivie http://www.ivie.es, así como las instrucciones para los autores que desean publicar en nuestras series.

Working papers can be downloaded free of charge from the Ivie website http://www.ivie.es, as well as the instructions for authors who are interested in publishing in our series.

Versión: febrero 2018 / Version: February 2018

Edita / Published by:
Instituto Valenciano de Investigaciones Económicas, S.A.
C/ Guardia Civil, 22 esc. 2 1º - 46020 Valencia (Spain)

DOI: http://dx.medra.org/10.12842/WPAD-2018-02

WP-AD 2018-02

Scoring Rules in Experimental Procurement*

Gian Luigi Albano, Angela Cipollone, Roberto Di Paolo, Giovanni Ponti and Marco Sparro**

Abstract

We report the results of a procurement experiments where subjects compete for procurement contracts to be awarded by means of a scoring auction. Two experimental conditions are considered, depending on the relative quality-price weight in the scoring rule. We show that different quality-price weights in the scoring rule dramatically alter the strategic environment and affect the extent to which the competitive mechanism leads to an efficient allocation of the contract. Our evidence suggests that, in spite of inducing significantly higher deviations from equilibrium, the scoring rule that gives more weight to quality over price is far more efficient (52% overall). We propose a "mediation analysis" to explain how the quality-price ratio determines the likelihood that an efficient allocation is realized, disentangling a "direct effect" (due to the equilibrium different properties of the induced game-forms) from an "indirect" one (how the different game-forms affect out-of-equilibrium behaviour).

Keywords: Scoring Auctions, Mechanism Design, Experimental Economics.

JEL classification numbers: C91, D70, D81, D91.

---

* We thank Federico Cesarini, Alessio Muscarnera, Laura Pacini, Annamaria Paolillo, and Michele Rosi for their precious research assistance. The usual disclaimers apply. Financial support from the Spanish Ministry of Economic Competitiveness (ECO2015-65820-P), Generalitat Valenciana (Research Projects Grupos 3/086 and PROMETEO/2013/037) and Instituto Valenciano de Investigaciones Económicas (Ivie) is gratefully acknowledged.

** G. L. Albano and A. Cipollone: Consip S.p.A. and LUISS Guido Carli Roma. R. Di Paolo: Universidad de Alicante. G. Ponti: Universidad de Alicante, The University of Chicago and LUISS Guido Carli Roma. M. Sparro: Consip S.p.A. Corresponding autor: Giovanni Ponti, Department of Economics, The University of Chicago. 1126 East 59th Street, Chicago IL 60637. E-mail: giuba92@uchicago.edu.

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, and to the emergence of more qualified procurement workforce, 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 tender's financial and non-financial attributes onto a one-dimensional score.[3] In a highest-score auction the tender awarded the highest score is deemed to be the winner and receives a financial payment equal to the submitted bid.[4] In spite of 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' optimal 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 is able to 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, independently on 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 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' optimal behaviour.[5]

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

[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))

[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.

strategic environment of scoring auctions. This question becomes even more compelling as there exists a substantial experimental evidence that -even in simple price-only auctions- actual behaviour may systematically differ from what theory predicts (see, for instance, Kagel and 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- the only experimental studies are concerned with extent with which buyers may benefit from using a multi-attribute rather than a price-only auction. This strand of research, initiated by Bichler (2000), has been further refined by Strecker (2003) and Chen-Ritzo *et al.* (2005).[6]

In this paper, we present the results of a stylized procurement auction experiment where a simulated buyer has to select the contractor out of 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. More specifically, the buyer solicits two-dimensional bids comprising a *quality* (where the latter affects production costs) and a *financial offer*, a rebate with respect to a publicly announced reserve (base) price. Price and quality dimensions are then mapped into a unidimensional score and the contract is awarded to the highest-score bidder.

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 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 independently 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 before bidding, each bidder, endowed with a certain quality level, becomes immediately aware of his technical score. Hence his strategic problem boils down to computing the optimal rebate 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 for each participant (see Grimm *et al.,* 2008).

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 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

[6] See also Chang et al. (2014), (2016).

possible quality levels. Thus, when the scoring rule puts a relatively high weight on price, not 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 bidders face. Proposition 1 collects the main characteristics of these two 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. Participants receive no feed-back until the end of the experiment, where a random draw selects the auction relevant for payment.

Section 4 reports our experimental results. We first notice that our two conditions yield a stark difference in behaviour: 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 efficiency between the two treatments, where efficiency is measured by the likelihood with which the subject with the highest pseudotype within the matching group wins the auction. 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 his relative efficiency, with RANK1 to (RANK5) indicating the bidder with the highest (lowest) pseudotype, respectively.

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
Total	100.00	100.00	100.00

Table 1. Distribution of winners by efficiency and treatment.

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 an efficient outcome when the weight of quality in the scoring mechanism is high rather than low. And this difference in efficiency is observed *despite* the higher noise detected in treatment which favours quality over price (see Figure 4 below).

This 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 behaviour, as well as behavioural 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 only 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). Our "mediation analysis" (Imai *et al.*, 2011) yields two main conclusions. First, the direct and indirect effects are both significant and 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 favour of the high-quality treatment.

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:

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

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

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

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]→[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 follow the approach pioneered by Che (1993), whereby bidders are characterized by "pseudotypes", which allows to rank bidders according to their winning probability.

To this aim, we first define type-q bidder's *pseudotype* or *potential score*[7], $s_\gamma(q) \equiv \gamma(1-c(q)) + (1-\gamma)q$, which corresponds to the score when submitting a rebate $r = r_{\max}(q) = 1 - c(q)$ and, by doing so, reducing to 0 the profits in case of winning. Since the scoring rule -basically- reflects the buyer's preferences with respect to the

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

trade-off between quality and price, we can consider the bidder with the highest potential score to be the *most efficient* in serving the contract.

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.

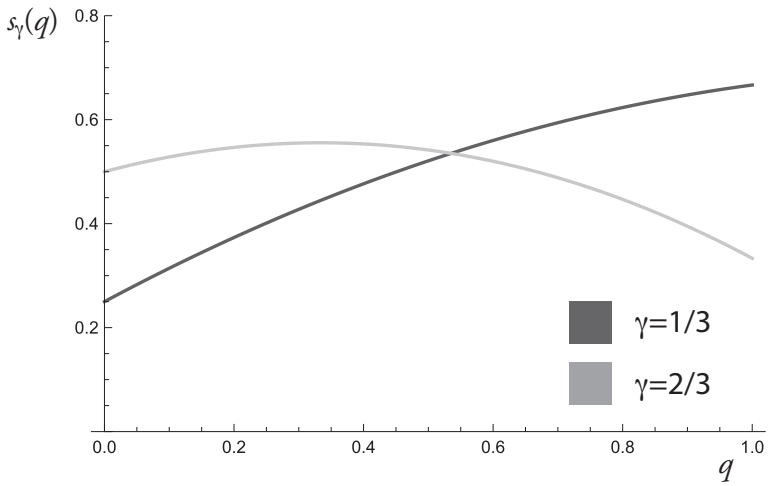


Fig. 1. Potential score function $s_\gamma(q)$ in the two treatments.

As shown in Figure 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

$$r_\gamma^*(q) = \max\left\{\frac{1}{\gamma}\left[\sigma_\gamma^*\left(s_\gamma(q)\right) - (1-\gamma)q\right], 0\right\} \text{ if } \gamma=1/3 \text{ and}$$

$$r_\gamma^*(q) = \frac{1}{\gamma}\left[\sigma_\gamma^*\left(s_\gamma(q)\right) - (1-\gamma)q\right] \text{ if } \gamma=2/3,$$

where $\sigma_\gamma^*\left(s_\gamma(q)\right) = \frac{1}{H_\gamma(s_\gamma(q))}\int_{\underline{s}_\gamma}^{s_\gamma(q)} y h_\gamma(y)dy$, with $H_\gamma(s)=G_\gamma^4(s)$, $h_\gamma=H_\gamma'(s)$ 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 Figure 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 optimally shades his bid below his value (that is, his pseudotype).

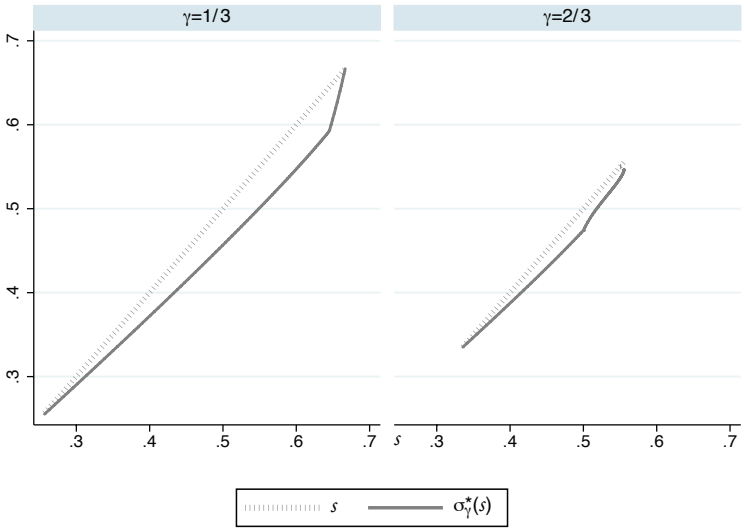


Figure 2. Private signals (pseudotypes) s and optimal bids $\sigma_\gamma^*(s)$.

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 Figure 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 Figure 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 (γ=2/3), the submitted rebates are higher than in the case of γ=1/3 for almost any q (precisely, for any $q > 0.05$). Second, when γ=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 γ=2/3) for less 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 optimally submit a discount below 0 (i.e., a price higher than the reserve price), which is not allowed by the rules of the game.

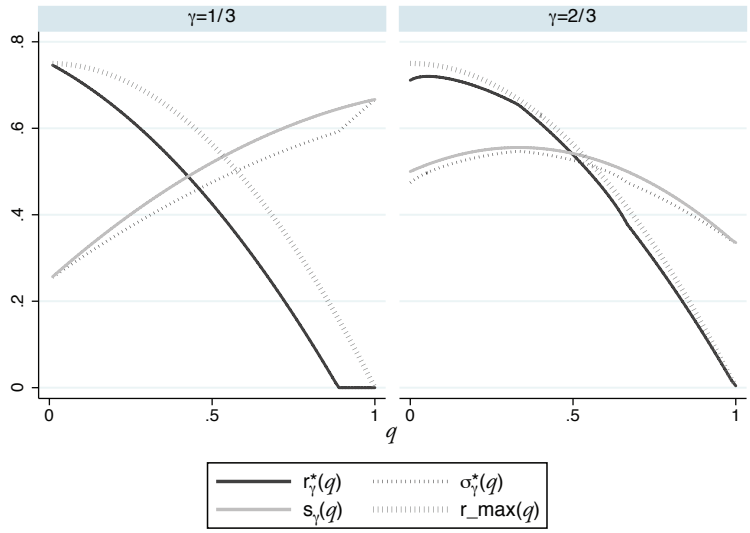


Figure 3. Equilibrium Analysis. The bidding functions are plotted both in terms of the submitted rebate *r* and the obtained score σ and compared with the maximum potential rebate/score (dotted lines).

The opposite is true when $\gamma = 2/3$. As shown in Figure 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 5 matching groups (cohorts) of 5, with subjects from different cohorts never interacting with each other throughout the experiment. 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.[9]

For each treatment $\gamma \in \{\frac{1}{3}, \frac{2}{3}\}$, subjects play 11 rounds of a procurement auction characterized in which an iid random draw without replacement determines the value of $q \in \left\{\frac{k}{10}\right\}, k = 0, 1, \dots, 10$, 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 (common) value was set at another € 10. 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 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 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (Correct answer: 5 minutes).
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).

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

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 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 (see Appendix B for details).

4. Results

4.1. Descriptive stats

Figure 4 tracks average and equilibrium bidding functions by treatment, together with the treatment 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 significantly higher at low quality levels since, for higher quality levels, bids are constrained by the rule that prevents losses.[10]

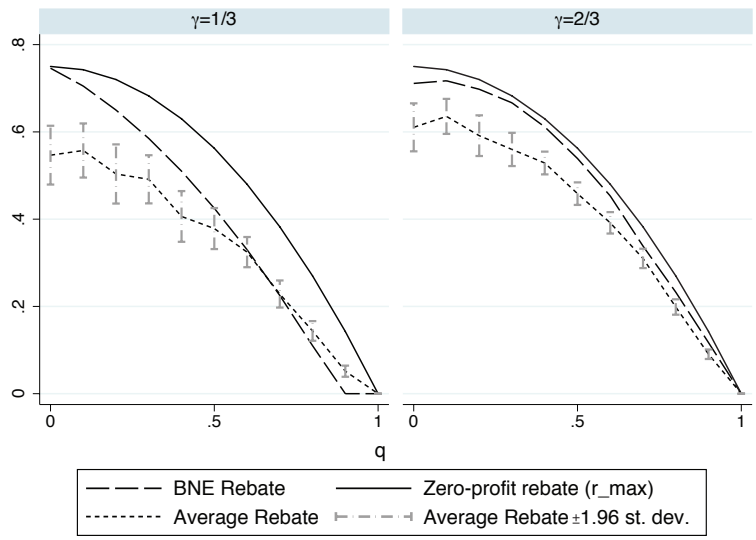


Figure 4. Equilibrium and empirical bidding functions by treatment

More importantly, when γ=2/3, players bid closer to equilibrium. In other words, the level of noise is *endogenous* and depends on the treatment conditions. One possible explanation for this phenomenon relies on the fact that the same deviation from equilibrium, call it Δ, yields a variation of the overall score equal to $\Delta/3$ ($2\Delta/3$)

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

if γ=1/3 (γ=2/3), respectively. Consistently with classic models of equilibrium with endogenous noise -take, for example, McKelvey and Palfrey's (1995) Quantal Response Equilibrium- we should then expect less noise in the treatment in which the impact of the latter on the overall score is higher, as it happens when γ=2/3. This evidence is of extreme importance for us since bidders' noise around equilibrium may be responsible for inefficient allocations (this is what we define as the "indirect effect" in Section 4.2). As a consequence, the evidence provided in Figure 4 -that individuals playing auctions with higher weight on the rebate play closer to equilibrium- *could* support the conclusion that auctions with γ=2/3 may be characterized by higher efficiency. However, as anticipated in Table 1, this is not the case: overall, our descriptive statistics deliver a ballpark estimate of a 52% higher efficiency when the weight of quality in the scoring rule is high (γ=1/3). Section 4.2 aims at rationalizing this apparent contradiction.

4.2. Welfare analysis: A two-stage approach

In what follows, we disentangle the "direct" treatment effect on efficiency –which is only due to the different strategic characteristics of the two treatments- from the "indirect" one –which is due to out-of-equilibrium behaviour. 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 the efficient allocation 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 on efficiency 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).

Figure 4 illustrates these two effects upon which we design our estimation strategy.

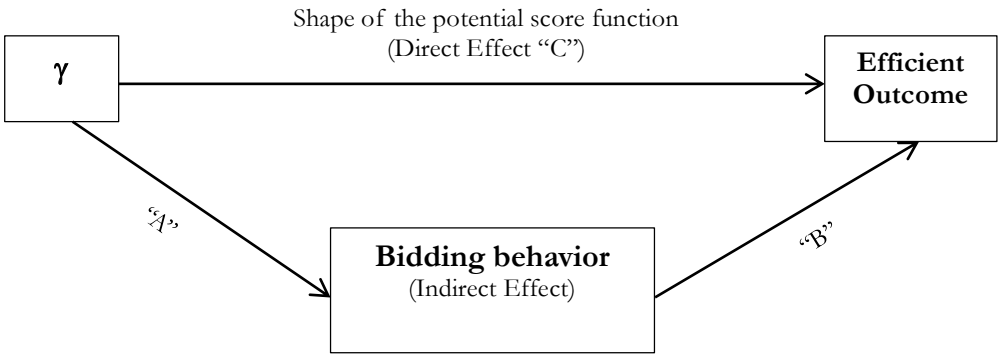


Figure 5. Direct and indirect effect of γ on efficiency

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- the likelihood of an efficient outcome 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 likelihood that an

efficient outcome is realized. By contrast, the indirect effect captures the impact of γ on efficiency via the level of noise which can be ascribed to a treatment change.

Identifying the direct and the indirect effect is relevant for the auction designer. If, say, the direct effect outweighs 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 must also take into account behavioural and context-specific factors, which may substantially complicate his task.

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

- **Step 1.** We regress the difference between observed and equilibrium bids on: i) our treatment variable, γ, by way of a binary index, positive when γ=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 Figure 4 as the marginal impact of γ on the observed "trembles" around equilibrium.
- **Step 2.** We regress the likelihood of an efficient outcome on i) the predicted deviations from equilibrium 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 Figure 4 (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 the probability of the efficient player to win the auction 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 3 reports only the estimated coefficients of the direct/indirect effects, together with their sum.

	Marginal impact	***p*-value**
Direct effect	-0.708	0.003
Indirect effect	+0.198	0.096
Total	-0.510	0.008

Table 3. Estimation of the direct/indirect effects

As Table 3 shows, we find an overall negative and significant treatment effect on efficiency in that auctions characterized by higher weight on price are 51% less likely to be awarded to the most efficient players. 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. This direct effect of γ is indeed negative and significant, suggesting that the same deviation from equilibrium is over 70% less likely to yield efficiency when $\gamma=2/3$. Moreover, our results also suggest a positive and significant indirect effect (+19.8%), which is not enough to outweigh this result.

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 a 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 Contract makes it mandatory to use *at least* a weight of 0.7 on quality when public buyers wish to carry out a procurement procedure by using a scoring auction. 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.

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.

References

Asker J, Cantillon E (2008) "Properties of Scoring Auctions", *RAND Journal of Economics*, 39, 69-85.

Bichler M (2000). "An experimental analysis of multi-attribute auctions", *Decision Support Systems,* 29(3), 249-68.

Chang WS, Salmon TC, Saral KJ (2016). "Procurement Auctions with Renegotiation and Wealth Constraints". *Economic Enquiry*, 54: 1684-704.

Chang WS, Chen B, Salmon TC (2014). "An Investigation of the Average Bid Mechanism for Procurement Auctions". *Management Science*, 61(6)6: 1237-54.

Che YK (1993). "Design Competition through Multi-Dimensional Auctions", *RAND Journal of Economics*, 24, 668-80.

Chen-Ritzo CH, Harrison TP, Kwasnica AM, Thomas DJ (2005). "Better, Faster, Cheaper: An Experimental Analysis of a Multiattribute Reverse Auction Mechanism with Restricted Information Feedback". *Management Science* 51(12), 1753-62.

Cueva C, Iturbe-Ormaetxe I, Mata-Pérez E, Ponti G, Sartarelli M, Yu H, Zhukova V (2016). "Cognitive (Ir)Reflection: New Experimental Evidence". "Cognitive (Ir)Reflection: New Experimental Evidence". *Journal of Behavioral and Experimental Economics*, 64, 81-93.

Fischbacher U (1999). "z-Tree: Zurich Toolbox for Ready-made Economic Experiments", *Experimental Economics*, 10(2), 171-178.

Frederick S (2005). "Cognitive reflection and decision making", *Journal of Economic Perspectives*, 19, 24-42.

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.

Grimm V, Kovarik J, Ponti G (2008). "Fixed Price plus Rationing: an Experiment", *Experimental Economics*, 11, 402-22.

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), 765-89.

Kagel JH, 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 Kagel JH, Roth AE (eds), *The Handbook of Experimental Economics – vol. 2*, Princeton NJ, Princeton University Press.

Krishna V (2009). *Auction Theory*, Academic Press (2nd edition).

McKelvey RD, Palfrey TR (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), 147-58.

Strecker S (2003). *Preference Revelation in Multi-Attribute Reverse English Auctions: A Laboratory Study* ICIS 2003 Proceedings. 23.

Vorob'ev NN (1993). *Foundations of Game Theory - Noncooperative Games*, Springer Basel AG

APPENDIX A

Proof of Proposition 1

Let $s_i = s(q) \equiv 1 - c(q)$ define player i's pseudotype. Consider the modified game in which each bidder privately observes a "value" s and submits a bid (that is, announces a score) $\sigma(s_i) \leq s_i$. Bidder i's expected payoff $\tilde{\pi}$ writes

$$E[\tilde{\pi}_i(s_i, \sigma(s_i))] = (s_i - \sigma(s_i))Prob[\sigma(s_i) > \max_{j \neq i} \sigma(s_j)]$$

where $max_{j \neq i}\sigma \ (s_j)$ indicates the highest score among bidder i's competitors. This modified game is then a first-price (selling) auction where, upon observing s_i, bidder i submits a bid $\sigma_i(s_i)$. If bidder i's bid is the highest submitted bid then bidder i gets profit equal to $(s_i - \sigma_i(s_i))$, and zero otherwise.

It is easy to show that the modified game is strategically equivalent to the original one. Loosely speaking, two games, A and B, are strategically equivalent when the two games have both the same set of agents and strategies, and game B's payoff function(s) can be obtained through an affine transformation of game A's payoff function(s).[11] Since $s_i = s(q_i)$ is defined as the maximum score bidder i with type q can obtain, $\sigma(s_i)$ coincides with the score obtained type q submitting a rebate $r(q)$. That is, $S(q, r(q)) = \sigma(s(q))$. Moreover, bidders' profit in the "original" game, (π_i) equal that of the modified game $(\tilde{\pi}_i)$ except for a positive constant factor $(1/\gamma)$:

$$\pi_i(q, r(q)) = 1 - r(q) - c(q) = \frac{1}{\gamma}[s(q) - \sigma(s(q))] = \frac{1}{\gamma}\tilde{\pi}_i(s(q), \sigma(s(q))),$$

while the winning probability of winning is exactly the same, as $\sigma(s_i) = \sigma(s(q)) = S(q_i, r(q))$. It results that the payoff functions in the two payoff functions differ by a multiplicative positive constant only, $\frac{1}{\gamma}$. Consequently, the two games have the same equilibria. This implies that the BNE of our original problem – that is, the equilibrium bidding function $\tilde{r}_\gamma(q)$ of the scoring auction – can be derived by deriving the equilibrium of the modified game:

$$\tilde{r}_\gamma(q) = \frac{1}{\gamma}\left[\sigma^*_\gamma(s(q)) - (1 - \gamma)q\right],$$

where $\sigma^*(s)$ is the BNE of a standard first-price auction.[12] Thus:

$$\sigma^*_\gamma(s(q)) = \frac{1}{H_\gamma(s)} \int_{\underline{s}(\gamma)}^{s(q)} y h_\gamma(y) dy$$

with $H_\gamma(s) = G_\gamma^4(s)$, $h_\gamma = H'_\gamma(s)$ and $G_\gamma(s)$ is the distribution function of the random variable s.

The only caveat is that we have not imposed any condition to ensure that, for each of the relevant parametric cases $\gamma \in \{1/3; 2/3\}$, $\tilde{r}_\gamma \ (q) \geq 0$. We proceed by computing first the "tentative" equilibrium rebate function $\tilde{r}_\gamma(\cdot)$, then we check that the non-negativity constraint is fulfilled. Deriving the explicit form of the equilibrium rebate function turns out to be quite cumbersome and uninstructive, so it is not provided in this proof.

[11] For a more formal definition see, for instance, Vorob'ev (1994), p. 43.
[12] This is an immediate application of standard auction theory. See Krishna (2009).

Let $\tilde{r}_\gamma(q) = \left[\sigma_\gamma^*\big(s(q)\big) - (1-\gamma)q\right]/\gamma$. Explicit computations show that $\tilde{r}_\gamma(q)$ assumes feasible values only for $q \in [0,1]$ when $\gamma = 2/3$ and thus it is actually a BNE, then $r_{2/3}^*(q) = \tilde{r}_{2/3}(q)$. When $\gamma = 1/3$, instead, it becomes negative for all values of $q > q_0 \approx .8884$. Intuitively, this occurs because when the weight of quality in the scoring rule is sufficiently high, types with the high values of q enjoy such a large probability of winning that they would be willing to lower their rebate below 0, that is, to submit a bid above 1, which is forbidden by the rules of the game. We then conjecture that $r_{\frac{1}{3}}^*(q) = \max\left\{\tilde{r}_{\frac{1}{3}}(q), 0\right\}$.

In order to show that this is indeed an equilibrium let $q^\circ = sup\left\{q: \tilde{r}_{\frac{1}{3}}(q) \geq 0\right\}$, that is, q° is the highest unconstrained type. Notice that if $r_{\frac{1}{3}}^*(q) = 0$ is part of an equilibrium for all $q > q^\circ$, then the probability of winning is still monotonic in the range $(q^\circ,1]$.

Consider any type $q \leq q^\circ$. The equilibrium bidding function is not affected by the constrain operating on types $q > q^\circ$. Indeed, due to the monotonicity of $s(q)$, the latter set of types would still submit a higher score than the set of types with $q \leq q^\circ$, thus leaving their probability of winning of all types $q \leq q^\circ$ unaffected. Hence, bidders with such types would have no incentives to deviate from bidding $r^*_{\frac{1}{3}}(q)$ if bidders with $q > q^\circ$ bid $r^*_{\frac{1}{3}}(q) = 0$. Consider now type $q' > q^\circ$ and suppose it envisages to submit a feasible $r_{\frac{1}{3}}(q'): 0 \leq r_{\frac{1}{3}}(q') \leq 1 - c(q')$. Then there must exist a type q'' such that

$$\sigma_{\frac{1}{3}}^*\left(q'', r_{\frac{1}{3}}^*(q'')\right) = \sigma_{\frac{1}{3}}^*\left(q', r_{\frac{1}{3}}(q')\right) \Leftrightarrow (1-\gamma)q'' = (1-\gamma)q' + \gamma \cdot r_{\frac{1}{3}}(q') \Rightarrow q'' = q' + \frac{\gamma}{1-\gamma} r_{\frac{1}{3}}(q').$$

Thus

$$r_{\frac{1}{3}}(q') = \frac{1-\gamma}{\gamma}\Delta(q') = 2\Delta(q'),$$

where $\Delta(q') = (q'' - q')$. We then need to prove that

$$[1 - c(q')](q')^4 \geq [1 - c(q') - 2\Delta(q')](q'')^4, \forall q' > q^\circ, \forall \Delta(q'): r_{\frac{1}{3}}(q') \leq 1 - c(q'), \tag{1}$$

where the LHS of inequality (1) represents the expected payoff of type q' when playing the conjectured equilibrium rebate $r^*_{\frac{1}{3}}(q) = 0$, and the RHS of inequality (1) measures the expected payoff of type q' when playing a strictly positive rebate yielding the same score as type q''. Inequality (1) can be rewritten as follows

$$[1 - c(q')](q')^4 \geq [1 - c(q') - 2\Delta(q')]\big(q' + \Delta(q')\big)^4 \Leftrightarrow$$
$$[1 - c(q')]\left[\big(q' + \Delta(q')\big)^4 - (q')^4\right] \leq 2\Delta(q')\big(q' + \Delta(q')\big)^4. \tag{2}$$

Notice that both the LHS and the RHS of inequality (2) are strictly increasing functions of $\Delta(q')$, and they are both equal to zero when $\Delta(q') = 0$. Call them $LHS_{(2)}(\Delta(q'))$ and $RHS_{(2)}(\Delta(q'))$, respectively. In order inequality (2) to hold it would then suffice to show that

$$\frac{\partial LHS_{(2)}\big(\Delta(q')\big)}{\partial \Delta(q')} \leq \frac{\partial RHS_{(2)}\big(\Delta(q')\big)}{\partial \Delta(q')}, \forall \Delta(q'): r_{\frac{1}{3}}(q') \leq 1 - c(q'),$$

that is

$$4[1-c(q')]\big(q'+\Delta(q')\big)^3 \leq 2\big(q'+\Delta(q')\big)^4 + 8\Delta(q')\big(q'+\Delta(q')\big)^3 \iff$$

$$4[1-c(q')]\big(q'+\Delta(q')\big)^3 \leq 2\big(q'+\Delta(q')\big)^3\big[\big(q'+\Delta(q')\big)+4\Delta(q')\big] \iff$$

$$2[1-c(q')] \leq [q'+5\Delta(q')], \tag{3}$$

which is always fulfilled for every $q' > q^\circ$ and every feasible $\Delta(q')$ given the assumption on the cost function $c(\cdot)$. Hence $r^*_{\frac{1}{3}}(q) = \max\left\{\tilde{r}_{\frac{1}{3}}(q), 0\right\}$ is indeed a BNE.■

APPENDIX B

THE ECONOMETRIC MODEL

For each vector of qualities drawn in any given auction we classify bidders according to their BNE score (that is, the score under the assumption that bidders play the BNE bids of Proposition 1) and obtain an efficiency-based ranking of 5 categories, from RANK1 to RANK5, indicating the bidder with the highest (lowest) BNE score, respectively. Then, we employ a random-effect linear probability model where i = 1, ...,18 identifies the matching group and t = 1, ..., 11 the period (i.e., the auction repetition):

$$y_{it} = \delta_0 + \delta_1\gamma_{2/3} + \delta_2(r - r^*)_{RANK1_{it}} + \delta_3(r - r^*)_{RANK2_{it}} + \delta_4 q_{RANK1_{it}} + \delta_5(q_{RANK1_{it}})^2 + \delta_6 q_{RANK2_{it}} + \\ +\delta_7(q_{RANK2_{it}})^2 + c_i + u_{it} \quad (1)$$

where: i) y_{it} is a binary index, positive if the auction of matching group i at period t is won by the most efficient type; ii) $\gamma_{2/3}$ is a dummy variable, positive if the weight on rebate is equal to 2/3 (remember that the treatment is randomized between groups); iii) $(r - r^*)_{RANK1_{it}}$ ($(r - r^*)_{RANK2_{it}}$) is the difference between observed and equilibrium bidding function of RANK1 (RANK2) player of matching group i in period t; iv) $q_{RANK1_{it}}$ ($q_{RANK2_{it}}$) is the realized quality of RANK1 (RANK2) player of matching group i in period t; v) c_i is a matching group specific random effect, uncorrelated with the independent variables and i.i.d. over the panel and vi) u_{it} is an idiosyncratic error term.

The choice of variables in (iii) in equation (1) is motivated by the evidence that, *for a given distribution of quality levels*, the probability of an efficient outcome increases (decreases) when the efficient (non-efficient) bidder RANK1 (RANK2) overbids. The parameter δ_1 is the marginal effect of our treatment variable γ on the probability to get an efficient outcome –holding all other variables constant at their means– and represents our *direct effect* "A" in Figure 5. Since RANK3 and RANK4 players never win $\gamma_{2/3}$ auctions, our econometric strategy focus on the bidding behavior of RANK1 and RANK2 players.

As motivated in the main text, our treatment variable might have a direct influence in explaining deviations from equilibrium. This is why the latter, $(r - r^*)_{RANK1_{it}}$ and $(r - r^*)_{RANK2_{it}}$, can be treated as *endogenous* to the value of γ as well as to the randomized level of quality and to individual-specific characteristics (such as the results from the Cognitive Reflection Test used in the regressions). In other words,

$$(r - r^*)_{RANK1_{it}} = \beta_0 + \beta_1\gamma_{2/3} + \beta_2 q_{RANK1_{it}} + \beta_3(q_{RANK1_{it}})^2 + \beta_4 CRTgroup_{RANK1_i} + \varepsilon_{it} \quad (2.1)$$

$$(r - r^*)_{RANK2_{it}} = \alpha_0 + \alpha_1\gamma_{2/3} + \alpha_2 q_{RANK2_{it}} + \alpha_3(q_{RANK2_{it}})^2 + \alpha_4 CRTgroup_{RANK2_i} + \mu_{it} \quad (2.2)$$

where: (i) $CRTgroup_{RANK1_{it}}$ ($CRTgroup_{RANK2_{it}}$) is a two-dimensional column vector of dummy variables for the group RANK1 (RANK2) player of matching group i has been assigned to on the basis of the Cognitive

Reflection Test (CRT group, see Section 3.4); (ii) the parameters α_1, β_1 are the marginal effects of our treatment variable γ on bidders' "trembles" around the equilibrium bid.

Substituting (2.1) and (2.2) into equation (1), it is simple to verify that the *indirect effect* described in Section 4.2 is computed by the following sum of products: $(\beta_1 \cdot \delta_2) + (\alpha_1 \cdot \delta_3)$. In particular, $(\beta_1 \cdot \delta_2)$ measures the extent to which efficiency changes exclusively when RANK1 players' bidding functions change by the amount they would have changed had γ moved from 1/3 to 2/3. Similarly, $(\alpha_1 \cdot \delta_3)$ measures the extent to which efficiency changes when RANK2 players' bidding functions change by the amount it would have changed had γ moved from 1/3 to 2/3. The total effect is equal to the sum of the direct and indirect effects: $\delta_1 + (\beta_1 \cdot \delta_2) + (\alpha_1 \cdot \delta_3)$.

The system of equations above has been estimated using a two-stage least-squares random-effect estimator.[13] This method applies an OLS regression on both equations (2.1) and (2.2) and gets predictions for $(r - r^*)_{RANK1_{it}}$ and $(r - r^*)_{RANK2_{it}}$ (Stage 1). After substituting $(r - r^*)_{RANK1_{it}}, (r - r^*)_{RANK2_{it}}$ with their predictions, equation (1) is estimated by OLS and the residuals are used to estimate the covariance matrix of equation errors (Stage 2).

Detailed results are presented in Table B.1. Column 1 presents the results of the final Stage 2; Columns 2 and 3 present the results from Stage 1 where predictions of both the endogenous variables $(r - r^*)_{RANK1_{it}}, (r - r^*)_{RANK2_{it}}$ are computed.[14]

Columns 2 and 3 point towards a significantly role of γ on the observed differences of players' bids from their BNE predictions. On average, the marginal impact of γ is negative and more relevant for RANK1 than for RANK2 player. This means that, *coeteris paribus*, the observed bidding functions are closer to the equilibrium bidding function when γ=2/3. Interestingly, bids from non-efficient reflective players are significantly higher than their BNE predictions compared to bids from other CRT-based groups. Figure 6 shows that, on average, the bidding functions of reflective players are closer to the equilibrium bidding functions compared to those of non-reflective players. This means that they are, on average, more likely to make a correct guess about the level of quality associated with the highest pseudotype. Hence, when endowed with a sub-optimal (second-best) level of quality, reflective players probably realize that they need to overbid if they want to have a chance of winning the auction. Moving to Stage 2 estimation, holding all other variables constant at their means, the probability of getting an efficient outcome when γ=2/3 is 70% smaller than when γ=1/3. Not surprisingly, aggressive bidding strategies of RANK2 players generate a significantly negative effect on the probability of achieving an efficient outcome. Precisely, a 10% increase in the distance of RANK2 players bids from their BNE predictions lowers the likelihood of an efficient outcome by 17%.

[13] The Random Effects (RE) approach provides consistent estimates in our context, since session-specific random effects, c_i, are uncorrelated with the experimental design and the way subjects are randomized across periods, sessions and treatment.

[14] Since the variables $r_{RANK1} - r^*_{RANK1}$ and $r_{RANK1} - r^*_{RANK2}$ range from negative to positive values, we choose to add a constant value to the data (that is the minimum value of $r_{RANK1} - r^*_{RANK1}$ to $r_{RANK1} - r^*_{RANK1}$ and the minimum value of $r_{RANK2} - r^*_{RANK2}$ to $r_{RANK2} - r^*_{RANK2}$) in order to get a clear intuition of the estimated impacts without loss of generality.

VARIABLES	(1) Prob. Efficient Winner	(2) $r_{RANK1} - r^*_{RANK1}$	(3) $r_{RANK2} - r^*_{RANK2}$
$\gamma_{2/3}$	-0.708***	-0.172***	-0.032
	(0.234)	(0.039)	(0.020)
q_{RANK1}	-0.582	0.083	
	(0.779)	(0.149)	
q_{RANK1}^2	0.386	-0.149	
	(0.650)	(0.123)	
q_{RANK2}	-0.947**		-0.331***
	(0.462)		(0.117)
q_{RANK2}^2	1.294**		0.481***
	(0.582)		(0.144)
Period	-0.020*	-0.003	-0.000
	(0.011)	(0.002)	(0.002)
$r_{RANK1} - r^*_{RANK1}$	-0.827		
	(0.618)		
$r_{RANK2} - r^*_{RANK2}$	-1.759***		
	(0.464)		
Impulsive_{RANK1}		-0.015	
		(0.013)	
$\text{Reflective}_{RANK1}$		0.012	
		(0.014)	
Impulsive_{RANK2}			-0.024
			(0.018)
$\text{Reflective}_{RANK2}$			0.033*
			(0.019)
Constant	2.946***	0.792***	0.580***
	(0.688)	(0.065)	(0.033)
Observations	198	198	198
R^2	0.29		
Underidentification test♦			
Kleibergen-Paap rk LM statistic	13.16		
p-value	0.0105		
Overidentification test♦♦			
Hansen J statistic	3.08		
p-value	0.5445		

Standard errors in parentheses, clustered at the group level. *** p<0.01, ** p<0.05, *p<0.1

♦ Kleibergen-Paap rk LM statistic.

♦♦ Hansen J statistic.

Table B.1. Determinants of the probability of the most efficient type to win the auction, Marginal Effect Values (MEs).

Our goodness-of-fit measures indicate that the model fits the data well. In particular, our under-identification tests -based on the Kleibergen-Paap rk LM statistics for estimates with heteroskedasticity-robust and clustered standard errors- reject the null hypothesis that Stage 1 and Stage 2 equations are underidentified (that is, the matrices of reduced form coefficients on the excluded instruments are full column rank). The Hansen statistics do not reject the null hypothesis that the instruments are valid instruments (that is, that they are uncorrelated with the error term), and, hence, the excluded instruments are correctly excluded from the estimated equations.

APPENDIX C

SUPPLEMENTARY STATISTICAL EVIDENCE

Table C.1 presents the summary statistics for selected individual–level variables which we deem to represent a good proxy of the unobserved individual heterogeneity which may have an impact on bidding behavior. With respect to the Cognitive Reflection Test, we find that 38% (28%) [34%] are classified as "impulsive" ("reflective") ["others"], respectively. From Table C.1 we also notice that the sample distribution over the CRT categories has a strong gender component: while 48% of the male sample is categorized as "reflective" (and the remaining 52% is approximately equally distributed across the other categories), the same percentage of females are classified as "impulsive" and only 13% as "reflective". This evidence is in line with previous findings in the literature (take, e.g., Frederick, 2005 and Cueva et al., 2016). With respect to the education field, the majority of our players is enrolled in an economic/business degree at a master level and expect to continue studying further. Parents' education level is relatively high, with over 50% of the sample declaring their father/mother holds a tertiary level of education. Only 20% of our players declare to have worked during the previous week and the reported weekly cash holdings is highly disperse, ranging from 15€ to 450€. Finally, the majority is in favour of merit-based compensation, while 67% of our players are tempered by prudent trust in others.

Variable	**Description**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
CRTgroup	Cognitive Reflection Test: 1 to 3	92	1.95	0.79	1	3
	=1 if Others	31				
	=2 if Impulsive	25				
	=3 if Reflective	26				
age	Age	92	22.73	2.19	19	31
woman	Gender: =1 if woman	92	0.57	0.50	0	1
economics	Field of education: =1 if student of economics	92	0.61	0.49	0	1
law	Field of education: =1 if law	92	0.25	0.44	0	1
political science	Field of education: =1 if political science	92	0.11	0.31	0	1
master_degree	Level of education: =1 if master's degree	92	0.64	0.48	0	1
phd	Level of education: =1 if Ph.D.	92	0.05	0.23	0	1
exp_master_degree	Expected level of education: =1 if master's degree	92	0.46	0.50	0	1
exp_phd	Expected level of education: =1 if Ph.D.	92	0.50	0.50	0	1
employed	Labour market status: =1 if employed	92	0.20	0.40	0	1
cash_holdings	Weekly cash holdings	88	91.08	71.62	15	450
unhappiness	Degree of unhappiness: 1 (happy) to 7 (unhappy)	90	3.79	1.72	1	6
trust	Trust in others: =1 if yes	90	0.37	0.53	0	2
meritocracy	Preference for Meritocracy: =1 if yes	89	0.90	0.30	0	1
inequality	Preference for income inequality: 1 (egalitarianism) to 7 (merit)	90	4.83	1.62	1	7
RSR	Room Size Ratio	90	2.51	1.29	0.8	10

Tab. C.1. Summary statistics.

Because treatment assignment has been randomized, we should observe no statistically significant differences across characteristics between the two treatment groups. Table C.2. shows that differences are indeed very

moderate. In terms of individual characteristics, subjects allocated to auctions with $\gamma = 1/3$ (the control group) are comparable to those allocated to auctions with $\gamma = 2/3$ (the treatment group).

Variable	Description	Obs	Difference in means Treatment-Control	p-value
CRTgroup	Cognitive Reflection Test: 1 to 3	92		
	=1 if Others	31	0.0941	0.3453
	=2 if Impulsive	25	0.0052	0.9596
	=3 if Reflective	26	-0.0993	0.2955
age	Age	92	0.7296	0.1114
woman	Gender: =1 if woman	92	0.1494	0.1517
economics	Field of education: =1 if student of economics	92	0.0170	0.8690
law	Field of education: =1 if law	92	-0.0761	0.4049
political science	Field of education: =1 if political science	92	0.0388	0.5554
master_degree	Level of education: =1 if master's degree	92	-0.0496	0.6242
phd	Level of education: =1 if Ph.D.	92	0.0629	0.1874
exp_master_degree	Expected level of education: =1 if master's degree	92	-0.0634	0.5471
exp_phd	Expected level of education: =1 if Ph.D.	92	0.0652	0.5367
father_sec_educ	Father's level of education: =1 if secondary	92	0.0293	0.7674
father_tert_educ	Father's level of education: =1 if tertiary	92	-0.1957	0.0615
father_phd	Father's level of education: =1 if Ph.D.	92	0.1035	0.1288
mother_sec_educ	Mother's level of education: =1 if secondary	92	0.1811	0.0674
mother_tert_educ	Mother's level of education: =1 if tertiary	92	-0.2189	0.0357
mother_phd	Mother's level of education: =1 if Ph.D.	92	-0.0463	0.3738
employed	Labour market status: =1 if employed	92	0.1220	0.1435
cash_holdings	Weekly cash holdings	88	8.2506	0.5920
unhappiness	Degree of unhappiness: 1 (happy) to 7 (unhappy)	90	-0.0128	0.9720
trust	Trust in others: =1 if yes	90	0.1677	0.1337
meritocracy	Preference for Meritocracy: =1 if yes	89	-0.0562	0.3855
inequality	Preference for income inequality: 1 (egalitarianism) to 7 (merit)	90	0.1707	0.6195
RSR	Room Size Ratio	90	-0.1559	0.5696

Tab. C.2. Summary statistics, by treatment group.

APPENDIX D

EXPERIMENTAL INSTRUCTIONS

WELCOME TO THE EXPERIMENT!

- This is an economic experiment on individual decision-making. We are only interested in your choices, not in who make them. Pay attention to your decisions because your behaviour will affect your final reward.
- In this experiment, you will play for 11 periods in which you must take a decision. Each decision and its result is independent from any other; namely, every decision that you take in a specific period does not have any effect on the results of other experiment's periods.
- At the beginning of the experiment, the computer will match everyone anonymously and randomly in groups of 5 players. This matching will be the same during all the experiment.

How you can gain a reward during the experiment?

- First of all, you will receive € 5 as a "show-up fee", to acknowledge your avaibility. Moreover, at the end of the experiment, one period will be drawn randomly and your winnings in that specific period will be summed up to the show-up fee and it will be privatly paid to you in cash at the end of the experiment.
- In what follows we will explain which decisions you have to take in each situation and how to deal with the user interface of the computer to implement them.
- Please do not disturb other partecipants during the course of the experiment. If you need help, raise your hand up and wait in silence. One of the proctors will come to help you as soon as possible.

Good Luck!

The Experiment

- In each period of the experiment, you will participate, together with the other 4 members of your group, in an auction in which everyone has to make an offer to win a "contract", the object of which is the service that you produce.
- In case you win the auction, your profit will be the difference between your price, paid to buy your service, and the cost incurred to achieve it. In case you lose the auction, you will not receive anything

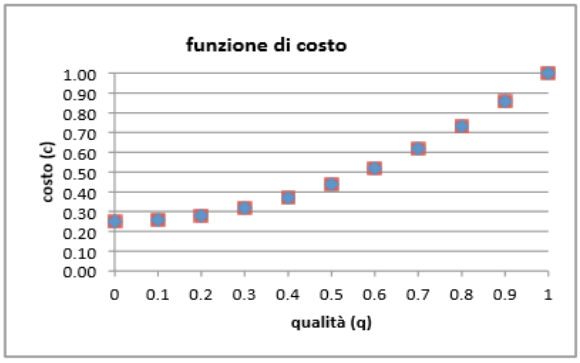
from the buyer and, moreover, you don't have any cost; namely, your profit will be zero. At the beginning of each period, you will receive the most relevant information about your bid, that is:

1. Your service quality, Q, for that period. This parameter (which value goes from 0, 10, 20 …to 100) is not chosen by you (neither by other group members) but is assigned randomly and independently by the computer, with equal probability for each value.

2. The cost, C, associated to the assigned quality, Q, that you have to pay in case you win the auction.

   - The cost parameter C depends on the quality: a higher quality is associated with a higher cost. In the following graph you can see the cost function, which is the same for all the players, in each phase and period. The graph and the table below report the level of quality assigned to each player, Q, and associated the cost, C, that each player has to pay in case s/he wins the auction: e.g., a quality level equal to 40 will always correspond to a cost equal to 37.

q	C
0	25 €
10	25.8 €
20	28.0 €
30	31.8 €
40	37 €
50	43.8 €
60	52 €
70	61.8 €
80	73 €
90	85.8 €
100	100 €



   - In each period, after having observed your quality and the associated cost, you can make your bid as follows:

   - Define your rebate, R, with respect to the baseline price (fixed at 100€ for all the players). The offered price (which correspond to (100-R)€) cannot be lower than your production cost, C, in that period. In this sense, the computer will prevent you from fixing a rebate that is too high so as to yield a monetary loss in case you win the auction (given that the price would cover your production costs for the service).

   - Choose your rebate, R, and put it in the corresponding box of the user interface. The relationship between your rebate and the price you offer is given by

$$\text{Price} = 100 - R$$

- As we said before, the computer does not allow any rebate R greater than € (100 – C).

How is the winner determined?

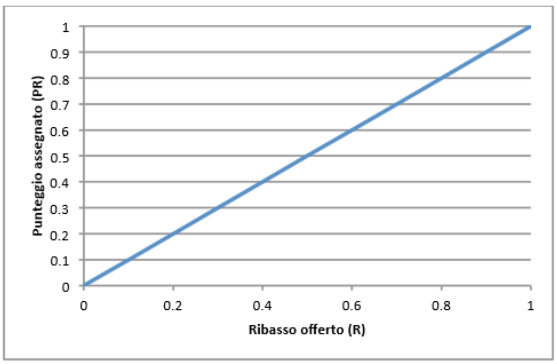
- In each period, after that all the players have chosen their value R, they receive a Total Score (TS), which is a weighted average between the quality, Q, and a score associated to each bid, namely FINANCIAL SCORE (FS).

- The winner will be the player in your group who obtains the highest TS. In case two or more players obtain the highest TS, all of them will be considered as winners and their payoff will be shared across the winners. The Total Score, TS, is calculated with the following formula, which will be the same during the entire experiment:

$$TS = \gamma \ Q + (1-\gamma) \ R$$

where Q is the level of quality determined by computer and R is the rebate, that is, a rebate of 10€ corresponds to 10 points, a rebate of 20€ corresponds to 20 points, and so on.

- The graph below shows how varies the score associated with the rebate as a function of R.



Di seguito, per ulteriore chiarimento, ti presentiamo l'esempio di un possibile esito di un periodo:

Giocatore	1	2	3	4	5
Qualità (Q)	60	40	90	30	50
Costo (C)	52	37	85.8	31.8	43.8
Ribasso (R)	25	39	1	30	37
$PT = \left(Q \ x \ \frac{1}{3}\right) + \left(PR \ x \ \frac{2}{3}\right)$	37	39	31	30	**41.3**
PROFITTO	0	0	0	0	**19.2**

- In this case, the winner is Player 5 because she obtained the best TS. Her profit (assuming a cost C = 43.8) would be:

100 – R – C = 100 – 37 – 43.8 = 19.2 €

To summarize:

- You are competing with other 4 players for the award of a service. The group will be the same during the entire experiment.
- In each period, the computer will assign randomly a value between 0 and 100, which is the quality of your service, Q. A cost C corresponds to each level of quality, Q, where C is an increasing function of Q, as specified in the table above. In each period, you must decide your rebate with respect to the baseline price (fixed for all players in each period to € 100), R.
- The service will be assigned to your group member who obtains the highest TOTAL SCORE, TS, according to this formula:

$$TS = 1/3 \times Q + 2/3 \times FS$$

- The winner's profit is equal to (100 – R – C)€.
- In the case of multiple winners, the profits are equally divided between them.
- The losers in each group obtain a profit equal to zero.
- Before starting the experiment, you will participate in 5 dry periods, to better understand the rules of the game described above. The computer will simulate the behaviour of the other 4 group members of your group, and it will provide you with a feedback at the end of each period.
- After these 5 periods, you will start playing for good. During 11 periods of the experiment, you will not receive any feedback (your score, your opponents' score, your profit, your opponent's profit, etc…). You will just receive all result of this experiment at the end of it.

Ivie

Guardia Civil, 22 - Esc. 2, 1º
46020 Valencia - Spain
Phone: +34 963 190 050
Fax: +34 963 190 055

Website: www.ivie.es
E-mail: publicaciones@ivie.es