That's the ticket: Explicit lottery randomisation and learning in Tullock contests[a]

Subhasish M. Chowdhury[b] Anwesha Mukherjee[b]
Theodore L. Turocy[cd]

July 17, 2017

[a]This project was supported by the Network for Integrated Behavioural Science (Economic and Social Research Council Grant ES/K002201/1). We thank Catherine Eckel, Aidas Masiliunas, and Henrik Orzen, as well as participants at the Contests: Theory and Evidence Conference, Eastern ARC Workshop, Jadavpur University Annual Conference 2016, and the Visions in Methodology Conference for helpful comments. We thank Tim Cason and the Vernon Smith Experimental Economics Laboratory at Purdue University for allowing us to use their facilities. Any errors are the sole responsibility of the authors.
[b]School of Economics, Centre for Behavioural and Experimental Social Science, and Centre for Competition Policy, University of East Anglia, Norwich NR4 7TJ, United Kingdom
[c]Corresponding author: T.Turocy@uea.ac.uk
[d]School of Economics and Centre for Behavioural and Experimental Social Science, University of East Anglia, Norwich NR4 7TJ, United Kingdom

Abstract

Most laboratory experiments studying Tullock contest games find bids significantly exceed equilibrium predictions. We design an experimental protocol which presents and implements the game in terms of frequency counts, using individually-numbered (virtual) lottery tickets. In contrast to the conventional protocol used in many recent experiments, we find that in our implementation (1) initial bid levels are significantly lower and (2) bids adjust more rapidly towards expected-earnings best responses. We demonstrate the robustness of our results by replicating them across two continents at two universities with different student profiles. Our results show that an adequate behavioural game theory of contests must take into account details of how the contest game is implemented.

JEL classifications: C72, C91, D72, D83.

Keywords: lottery contest, learning, frequency, experiment.

1 Introduction

"The hero ... is condemned because he doesn't play the game. [...] But to get a more accurate picture of his character, [...] you must ask yourself in what way (the hero) doesn't play the game".

— Albert Camus, in the afterword of *The Outsider* (Camus, 1982, p. 118)

Economic agents can expend resources strategically in contests in order to increase their chances of winning valuable prizes. Examples of such contests can be found in settings as diverse as rent-seeking, electoral competition, advertising, research and development, and sports. Ultimate success in these competitions can depend both on the expenditures (hereinbelow *bids*) of the contestants, as well as other factors which can be thought of as luck or happenstance.

The model of Tullock (1980) is the workhorse used when the impact of luck is assumed to be sufficiently large. In Tullock's specification of a contest, for any two contestants, their relative chances of victory are given by the ratio of their bids, raised to some exponent. This exponent is frequently set to one in applications, and it is this case we will focus on in the current study. With an exponent of one, the influence of chance is large enough that a contestant's payoff, as a function of their effort investment, is single-peaked, and the corresponding best response changes smoothly as the conjectured investments of other competitors is changed. In a symmetric setting with risk-neutral contestants and no spillovers, there is a unique Nash equilibrium in which the contestants play pure strategies (Szidarovszky and Okuguchi, 2008; Chowdhury and Sheremeta, 2011).

The Tullock contest has been a component in many experimental studies, either standing alone or as part of a broader research question. Most of these studies report that bids in the contests exceed the levels predicted by the risk-neutral Nash equilibrium, and that this overdissipation is persistent even after repeated experience with the game. To facilitate comparability across studies, experimenters often adopt elements in their protocols which are similar to those used in other recent studies. Particularly in a field as active as Tullock contests have been in recent years (cf. the survey of Dechenaux et al., 2015), this can lead to a convergence in protocols to a de-facto standard. However, the Tullock contest's success as a workhorse model is underpinned in part by its versatility, with the same formal strategic-form game used as a metaphor across a variety of settings. Yet, it is well-established that human decision-makers do attend to details of language and process. The abstraction of a particular setting to the noncooperative game model may therefore discard features which are behaviourally relevant, and therefore important for understanding the performance of particular instances of contests.

We undertake a critical review of protocols for implementing the Tullock contest game. This review informs the design of a new protocol which is internally self-consistent, in that it presents and carries out the game entirely in terms of information about frequencies or counts, as opposed to ratios or probabilities. This protocol corresponds to a lottery raffle, an institution which exists and is familiar to most people. We show that that using our protocol, initial bid are lower, and bids adjust more rapidly towards expected-earnings mutual best responses.

Table 1 updates and extends the list of Tullock contest studies compiled in Sheremeta (2013). The rightmost column in this table shows the ratio of observed bids in the experiment to the risk-neutral Nash equilibrium bid. Most studies report bids to be above the Nash prediction. Sheremeta (2013) reports that the median experiment generates bids 1.72 times that of the Nash prediction, on the basis of his meta-analysis of 30 different contest experiments involving 39 experimental treatments. Several papers (Herrmann and Orzen, 2008; Abbink et al., 2010; Cason et al., 2012a; Cohen and Shavit, 2012; Mago et al., 2013) report average bids more than double the Nash level. However, bids relative to the Nash prediction do vary substantially, with some studies finding bids below or close to the Nash benchmark.

We augment the survey of Sheremeta by collecting information on the experimental protocols used for presenting the game.[1] We focus on two qualitative features of these protocols: (1) whether the instructions refer to a ratio of bids mapping into probabilities of winning the prize, and (2) what concrete randomisation mechanism, if any, is mentioned in the instructions.

[Table 1 about here.]

The column "ratio rule" in Table 1 indicates whether the experiment's instructions made an explicit mention that the probability of winning is given by the *ratio* of the contestant's own bid to the total bids of all contestants. A majority of studies do discuss this ratio, with many, including Falucchi et al. (2013), Ke et al. (2013), and Lim et al. (2014) explicitly using displayed mathematical formula like (1).[2]

Many designs supplement mention of a ratio or probability with a concrete description of a mechanism capable of generating the probability. For example, instructions may state that it is "as if" bids translate into lottery tickets, or other objects such as balls or tokens (Potters et al., 1998; Fonseca, 2009; Masiliunas et al., 2014; Godoy et al., 2015), which are then placed together in a container, with one drawn at random to determine the winner. We record this practice as "Lottery" in the "Example" column of Table 1. However, in carrying out the experiment itself, experimenters rarely if ever resolve the outcome of each period using an explicit lottery draw presentation.

An alternative approach used in a minority of studies (Schmidt et al., 2006; Herrmann and Orzen, 2008; Morgan et al., 2012; Ke et al., 2013) is a spinning lottery wheel. On this wheel, bids

[1] We were unable to obtain instructions for a few of the studies listed. Cells with entries marked — indicate studies we were not able to classify.

[2] Another alternative, giving a full payoff table as used by Shogren and Baik (1991), is a rather rare device.

are mapped proportionally onto wedges of a circle. Morgan et al. (2012) and Ke et al. (2013) report bids around 1.5 times that predicted by the equilibrium, while Herrmann and Orzen (2008) report bids just above equilibrium and Schmidt et al. (2006) find bids below the equilibrium prediction.

Based on this survey, the emerging standard of recent years, which we refer to as the *conventional* approach, is for a Tullock contest experiment to

1. introduce the game in terms of a probability or ratio;

2. give a lottery as an example mechanism, but

3. not to play out the lottery explicitly when determining the winner of the prize in each game.

The introduction of the conceit of the lottery mechanism suggests that the total *counts* or frequencies of tickets are helpful for participants to understand and process the strategic situation. Differences in processing probabilistic (ratio) versus frequency (count) information have been studied extensively in psychology. For example, Gigerenzer and Hoffrage (1995) proposed humans are well-adapted to manipulating frequency-based information as this is the format in which information arises in nature.

In our *ticket* treatment we use a design which maintains the integrity of the game's representation throughout both the instructions and play. The instructions introduce the game only as a lottery, in which bids translate into individually-numbered tickets. When playing the game, ticket numbers are assigned, and in realising the outcome of the game the number of the winning ticket is revealed. The play of the game therefore matches the description in the instructions. It likewise corresponds closely to the implementation of real-world lotteries, in which tickets are often collected into a container such as a large drum, with one selected at random to determine the winner.

The self-consistent implementation of the game in the ticket treatment results in significantly lower bids in the first period. As we do not conduct any live sample draws or practice rounds, first-period decisions are made prior to experiencing the feedback of the draw mechanism. This implicates the frequency-based explanation itself as a driver of bids in the contest. In addition, bids move more quickly towards approximate mutual best responses (in expected earnings terms) when the experimental protocol is designed with the explicit lottery structure. The dynamics of behaviour depend on carrying through the mechanism, rather than using it only as a hypothetical "as if" aside in the instructions. We demonstrate the robustness of our findings by replicating our experiment in two countries across two continents, at two universities which have undergraduate student bodies with distinct profiles.

We present the formal description of the Tullock contest game, our two experimental implementations, and the hypotheses in Section 2. The summary of the data and the results are included in Section 3. We conclude in Section 4 with further discussion.

2 Experimental design

The Tullock lottery contest game we study is formally an n-player simultaneous-move game. There is one indivisible prize, which each player values at $v > 0$. Each player i has an endowment $\omega \geq v$, and chooses a bid $b_i \in [0, \omega]$. Given a vector of bids $b = (b_1, \ldots, b_n)$, the probability player i receives the prize is given by

$$p_i(b) = \begin{cases} \frac{b_i}{\sum_{j=1}^{n} b_j} & \text{if } \sum_{j=1}^{n} b_j > 0 \\ \frac{1}{n} & \text{otherwise} \end{cases} \tag{1}$$

If players are risk-neutral, they maximise their expected payoff $u_i(b) = vp_i(b) + (\omega - b_i)$. The unique Nash equilibrium is in pure strategies, with $b_i^{NE} = \min\left\{\frac{n-1}{n^2}v, \omega\right\}$ for all players i.

In our experiment, we choose $n = 4$ and $\omega = v = 160$. We restrict the bids to be drawn from the discrete set of integers, $\{0, 1, \ldots, 159, 160\}$. With these parameters, the unique Nash equilibrium has $b_i^{NE} = 30$.

Participants played 30 contest periods, with the number of periods announced in the instructions. The groups of $n = 4$ participants were fixed throughout the session. Within a group, members were referred to anonymously by ID numbers 1, 2, 3, and 4; these ID numbers were randomised after each period. All interaction was mediated through computer terminals, using zTree (Fischbacher, 2007). A participant's complete history of their own bids and their earnings in each period was provided throughout the experiment. Formally, therefore, the 4 participants in a group play a repeated game of 30 periods, with a common public history.[3] By standard arguments, the unique subgame-perfect equilibrium of this supergame interaction is to play the $b_i^{NE} = 30$ in all periods irrespective of the history of play.

We contrast two treatments, the *conventional* treatment and the *ticket* treatment, in a between-sessions design. The instructions for both treatments present the game as a lottery.[4] In the conventional treatment, the instructions explain the relationship between bids and chances of receiving the prize using the mathematical formula first, with a subsequent sentence mentioning that bids could be thought of as lottery tickets. Our explanation follows the most common pattern found across the studies surveyed in Table 1. In the ticket treatment, each penny bid purchases an individually-numbered lottery ticket. To determine the winner of the prize, the number of one of those lottery tickets is selected and displayed on the screen.

The randomisation in each period was presented to participants in line with the explanations in

[3]That is, all players in the group share the same history of which bids were submitted in each period. Because of the re-assignment of IDs in each round, the private histories differ, because each participant knows which of the four bids in each round was the one they submitted, and whether or not they were the winner in each round.

[4]We provide full text of the instructions in Appendix A.

the instructions. In conventional treatment sessions, after bids were made but before realising the outcome of the lottery, participants saw a summary screen (Figure 1a), detailing the bids of each of the participants in the group. In sessions using the ticket treatment, the explicit ticket metaphor was played out by providing the identifying numbers for each ticket purchased (Figure 1b).

[Figure 1 about here.]

There are two channels through which the implementation of the ticket treatment could lead to behaviour different from that in the conventional treatment.

1. While both mention the lottery ticket metaphor, the conventional treatment instructions discuss the chances of receiving the prize, whereas the ticket treatment uses counts of tickets purchased. An effect due to this change would be identifiable in the first-period bids, which are taken when participants have not had any experience with the mechanism or information about the behaviour of others.

2. Feedback is structured in terms of the individually-identifiable tickets. The number of the winning ticket conveys no additional payoff-relevant information beyond the identity of its owner. An effect due to this feedback structure will be identifiable by looking at the evolution of play within each fixed group over the course of the 30 periods of the session.

We structure our analysis to look for treatment effects via both of these possible channels.

We conducted a total of 14 experimental sessions. Eight of the sessions took place at the Centre for Behavioural and Experimental Social Science at University of East Anglia in the United Kingdom, using the hRoot recruitment system (Bock et al., 2012), and six at the Vernon Smith Experimental Economics Laboratory at Purdue University in the United States, using ORSEE (Greiner, 2015). We refer to the samples as UK and US, respectively. In the UK, there were four sessions of each treatment with 12 participants (3 fixed groups) per session; in the US, there were three sessions of each treatment with 16 participants (4 fixed groups) per session. We therefore have data on a total of 48 participants (12 fixed groups) in each treatment at each site.

The units of currency in the experiment were pence. In the UK sessions, these are UK pence. In the US sessions, we had an exchange rate, announced prior to the session, of 1.5 US cents per pence. We selected this as being close to the average exchange rate between the currencies in the year prior to the experiment, rounded to 1.5 for simplicity.

Participants received payment for 5 of the 30 periods. The five rounds which were paid were selected publicly at random at the end of the experiment, and were the same for all participants in a session.[5] Sessions lasted about an hour, and average payments were approximately £10 in the UK and $15 in the US.

[5]The US participants also received a $5 participation payment on top of their contingent payment, to be consistent with conventions at Purdue.

3 Results

We begin with an overview of all 5,760 bids in our sample. Figure 2 displays dotplots for the bids made in each period, broken out by subject pool and treatment. Table 2 provides summary statistics on the individual bids for each treatment and subject pool. Both the figure and table indicate a treatment difference. Aggressive bids at or near the maximum of 160 are infrequent in the ticket treatment after the first few periods, but persist in the conventional treatment. Figure 3 summarises the distribution of mean bids by group over time. The aggregate patterns of behaviour are similar in the UK and US.

[Table 2 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

Result 1. *In each treatment, there are no significant differences between the distributions of bids in the UK versus in the US.*

Support. We use the group as the unit of independent observation, and compute, for each group, the average bid over the course of the experiment. The Mann-Whitney-Wilcoxon rank-sum test does not reject the null hypothesis of equal distributions of these group means across the subject pools ($p = 0.86$ for the conventional and $p = 0.91$ for the ticket treatment). Similarly, the Mann-Whitney-Wilcoxon test does not reject the null hypothesis if the group means are computed based only on periods 1-10, 11-20, or 21-30.[6] □

In view of the similarities between the data from the two subject pools, we continue by using combined sample for our subsequent analysis. Our next result treats the full 30-period supergame as a single unit for each group, and compares behaviour to the benchmark of the unique subgame-perfect Nash equilibrium in which the stage game equilibrium is played in each period.

Result 2. *Bids are significantly lower over the course of the experiment in the ticket treatment than in the conventional treatment. Both treatments significantly exceed the Nash equilibrium prediction.*

Support. For each group we compute the mean bid over the course of the experiment. The mean over groups is 51.7 in the conventional treatment (standard deviation 14.8) and 40.7 in the ticket

[6] For the conventional treatment, the p-values for the M-W-W test are $p = 0.69$ for periods 1-10, $p = 0.77$ for periods 10-21, and $p = 0.91$ for periods 21-30. For the ticket treatment the corresponding p-values are $p = 0.39$, $p = 0.95$, and $p = 0.29$, respectively.

(standard deviation 9.1). Figure 4 plots the full distribution of these group means; the boxes indicate the locations of the median and upper and lower quartiles of the distributions. Using the Mann-Whitney-Wilcoxon rank-sum test, we reject the hypothesis that the distributions are equal ($p = 0.0036$). □

[Figure 4 about here.]

The difference between the treatments could be attributable to some difference in how experiential learning takes place because of the feedback mechanism in playing out the lottery explicitly, or simply because participants process the explanation of the game differently. We can look for evidence of the latter by considering only the first-period bids.

Result 3. *First-period bids are significantly lower in the ticket treatment than the conventional treatment, and therefore are closer to the Nash equilibrium prediction.*

Support. Figure 5 displays the distribution of first-period bids for all 192 bidders (96 in each treatment). Because at the time of the first-period bids participants have had no interaction, we can treat these as independent observations. The mean first-period bid in the conventional treatment is 71.1, versus 56.8 in the ticket treatment. Put another way, as a point estimate approximately 35% of the observed overbidding relative to the Nash prediction is explained in the first period by the treatment difference. Using the Mann-Whitney-Wilcoxon rank-sum test, we reject the hypothesis that the distributions are equal ($p = 0.020$). □

[Figure 5 about here.]

First-period bids on average are above the equilibrium prediction in both treatments. We therefore turn to the dynamics of bidding over the course of the session. Returning to the group as the unit of independent observation, Figure 6 displays boxplots of the distribution of group average bids period-by-period for each treatment. Bid levels are higher in the conventional treatment in the first period, and both treatments exhibit a trend of average bids decreasing towards the Nash equilibrium prediction.

[Figure 6 about here.]

We are interested in determining whether the ticket-based implementation of the lottery also has an effect on the dynamics of behaviour over the experiment. Under the maintained assumption that participants are interested in the earnings consequences of their actions, we organise our analysis of dynamics in terms of payoffs, rather than bids themselves. Consider a group g in session s of treatment $c \in \{\text{conventional}, \text{ticket}\}$. We construct for this group, for each period t, a measure of disequilibrium based on ε-equilibrium. (Radner, 1980) In each period t, each bidder i in the group submitted a bid b_{it}. Given these bids, bid b_{it} had an expected payoff to i of

$$\pi_{it} = \frac{b_{it}}{\sum_{j \in g} b_{jt}} \times 160 + (160 - b_{it}).$$

For comparison, we can consider bidder i's best response to the other bids of his group. Letting $B_{it} = \sum_{j \in g : j \neq i} b_{jt}$, the best response, if bids were permitted to be continuous, would be given by

$$\tilde{b}^{\star}_{it} = \max\{0, \sqrt{160 B_{it}} - B_{it}\}.$$

Bids are required to be discrete in our experiment; the quasiconcavity of the expected payoff function ensures that the discretised best response $b^{\star}_{it} \in \{\lceil \tilde{b}^{\star}_{it} \rceil, \lfloor \tilde{b}^{\star}_{it} \rfloor\}$. This discretised best response then generates an expected payoff to i of

$$\pi^{\star}_{it} = \frac{b^{\star}_{it}}{b^{\star}_{it} + B_{it}} \times 160 + (160 - b^{\star}_{it}).$$

We then write[7]

$$\varepsilon_{csgt} = \max_{i \in g}\{\pi^{\star}_{it} - \pi_{it}\}.$$

By construction, $\varepsilon_{csgt} \geq 0$, with $\varepsilon_{csgt} = 0$ only at the Nash equilibrium.

Conducting the analysis in the payoff space measures behaviour in terms of potential earnings. The marginal earnings consequences of an incremental change in bid depends on both b_{it} and B_{it}, so a solely bid-based analysis would not adequately capture incentives. In addition, although in general bids are high enough that the best response in most groups in most periods is to bid low, there are many instances in which the best response for a bidder would have been to bid higher than they actually did. A focus on payoffs allows us to track the dynamics without having to account for directional learning in the bid space.

[Figure 7 about here.]

Figure 7 shows the evolution of the disequilibrium measure ε over the experiment. The clustering of this measure at lower values, especially below about 30, is evident in the ticket treatment throughout the experiment, while any convergence in the conventional treatment is slower. While suggestive, these dot plots alone are not enough to establish whether the evolution of play differs between the treatments, because it does not take into account the dynamics of each individual group. Result 3 implies that values of ε in Period 1 are lower in the ticket treatment. Therefore,

[7] Taking the maximum to define the metric ε_{csgt} gives the standard definition of ε-equilibrium. Our results about the treatment effect on dynamics also hold if ε_{csgt} is defined as the average or the median in each group.

the difference seen in Figure 7 could be attributable to the different initial conditions rather than different dynamics, as there is simply less room for ε to decrease among the groups in the ticket treatment given their first-period decisions.

We control for this by investigating the evolution of ε within-group over the experiment. As a first graphical investigation, we plot the average value of $\varepsilon_{csg(t+1)}$ as a function of ε_{csgt} for both treatments in Figure 8.[8] Consider two groups, one in the conventional treatment and one in the ticket treatment, who happen to have the same ε in some period. Figure 8 says that in the subsequent period, on average, the ε measure of the group in the ticket treatment will be lower, that is, they will move further towards an approximate mutual (expected-earnings) best response.[9]

[Figure 8 about here.]

Result 4. *Convergence towards equilibrium, as measured by ε-equilibrium, is significantly faster in the ticket treatment than in the conventional treatment.*

Support. To formalise the intuition provided by Figure 8, we estimate a random-effects panel regression

$$\varepsilon_{csg(t+1)} = \alpha + \beta_0 \varepsilon_{csgt} + \beta_1 \mathbf{1}_{c=\text{ticket}} + \beta_2 \mathbf{1}_{c=\text{ticket}} \times \varepsilon_{csgt},$$

where $\mathbf{1}_{c=\text{ticket}}$ is a dummy variable which takes on the value 1 for groups in sessions using the ticket treatment and 0 otherwise. The resulting parameter estimates, with standard errors clustered at the session level, are reported in Table 3. Both the intercept and slope parameters for the ticket treatment are significantly lower than for the conventional treatment, indicating a faster rate of convergence for groups in the ticket treatment from any given initial value of ε. □

[Table 3 about here.]

Figure 8 and Table 3 show that in both treatments, groups that have very small values of ε in one period tend to increase ε in the subsequent period; that is, they move away from equilibrium. The fixed point for ε using the point estimates is about 37.5 for the conventional treatment, and 22.5 for the ticket treatment. This observation is consistent with previous studies (Chowdhury et al., 2014) which have demonstrated a difference between the Tullock contest with a random outcome, as studied here, and a version in which the prize is shared deterministically among the contestants in proportion to their bids. When behaviour is close to equilibrium, small deviations in bids have small consequences in terms of expected payoffs, leaving open the door for other behavioural

[8] For the purposes of Figure 8 we aggregate observations by rounding ε_{csgt} to the nearest multiple of five, and taking the average over all observations with the same rounded value.

[9] There are very few groups in either treatment with values of ε above about 75, accounting for the instability in the graph for large ε.

factors to come into play. For example, although the outcome of the randomisation contains no new information for participants, they may nevertheless base their bids in subsequent periods in part on the outcomes of previous random draws. The presence of this or similar heuristics would introduce an underlying level of noise in play consistent with the positive intercepts obtained in Table 3.

4 Discussion

We have demonstrated that a self-consistent use of a ticket-based implementation of the Tullock contest, throughout both instructions and the play of the game, has a significant effect on behaviour. Measured relative to the risk-neutral Nash equilibrium commonly used as a benchmark in the literature, roughly one-third of the overbidding in the first period, and one-half of the overbidding over the course of the 30 periods of our experiment, can be attributed to this implementation relative to the conventions used in most recent studies.

In addition to the contrast between frequency-based and rate-based representations as proposed by Gigerenzer and others, there is a literature on mathematical thinking and learning that asserts that feelings of uncertainty, anxiety, or discomfort may arise in individuals when presented with mathematical displays or terminology (Tobias and Weissbrod, 1980; Molina et al., 2016; Aiken, 1970; Goldin, 2002). Experimental instructions often employ a displayed formula patterned after (1). Kapeller and Steinerberger (2013) argue that mere mathematical representation of an argument may hinder understanding among a significant proportion of people. Xue et al. (2017) show that people who self-report that they are not "good at math" make earnings-maximising choices substantially less often in a riskless decision task. On the basis of tasks involving basic numerical additions, they propose their results are consistent with the theory of math anxiety. As quantitative researchers, experimental economists may suppose that the precision of a mathematical display encourages comprehension. The possibility mathematical displays may have the opposite effect would have significant implications for designing and interpreting laboratory and field experiments. Investigating this falls outside the design of this experiment, and is left for potentially fruitful further research.

Our design investigates directly possible differences across participant pools. In addition to the cross-country, cross-continent comparison between the UK and the US, there are differences in the profile of students in the participant pool. Purdue University is noted for its strength in degrees in engineering; University of East Anglia does not offer degrees in engineering but is well-known for its programmes in areas like creative writing. Nevertheless, the data in both treatments is very similar across sites, providing initial evidence that results in contest experiments, including the treatment effects we report here, are portable across at least some participant pools. The use of university students in English-speaking countries allowed us the control of using identical instruction texts, but samples only one part of the population. Because university students as a whole, even those studying less mathematically-intensive subjects, are more likely to have extended formal and recent exposure to concepts of ratios and probability, they may process explanations using those concepts differently than the less-trained general population. In contrast, a ticket-based lottery raffle is a common institution which is broadly familiar to most everyone, at all levels of educational attainment as well as across cultures. This would suggest the hypothesis that the performance of the ticket treatment would be more stable and portable in the population at large.

The lottery implementation lends itself naturally to the revelation of the outcome by an explicit mechanism like the number of the winning ticket. While such a lottery can be modeled as a Tullock contest, there are other settings in which the Tullock model can and has been applied where the resolution of the randomness of the outcome cannot be so succinctly identified. For example, in an application to a sporting contest, it is usually not possible to point to one deciding (random) factor which leads one contestant to prevail over another. We therefore do not claim our ticket-based approach is necessarily the "better" one; the appropriate design depends on the application domain in which the experimenters imagine their results will apply. However, our results do underscore that while there are many settings in which the versatile Tullock model is applied, the strategic equivalence of the assumed underlying normal form games across those settings does not extent to a behavioural equivalence.

A satisfactory behavioural game theory of contests will need to draw out the relationship between implementation features and the performance of the contest mechanism. Meanwhile, our results point out that care must be taken in using statements like "X is a stylised fact in laboratory Tullock contest games." A Tullock contest is a non-cooperative game, which is an abstract, mathematical object. We argue by analogy that this non-cooperative game is a useful abstraction and description of many different real-world settings. However, players interact within the real-world setting, and not within (or with) the abstraction. A critical appraisal of the details of experimental protocol, as we have done here for the case of the Tullock contest, is essential to mapping out the external validity of results in the laboratory to the domains of real-world application of the model.

References

Klaus Abbink, Jordi Brandts, Benedikt Herrmann, and Henrik Orzen. Intergroup conflict and intra-group punishment in an experimental contest game. *The American Economic Review*, 100:420–447, 2010.

T. K. Ahn, R. Mark Isaac, and Timothy C. Salmon. Rent seeking in groups. *International Journal of Industrial Organization*, 29(1):116–125, 2011.

Lewis R. Aiken. Attitudes toward mathematics. *Review of Educational Research*, 40(4):551–596, 1970.

Lisa R. Anderson and Sarah L. Stafford. An experimental analysis of rent seeking under varying competitive conditions. *Public Choice*, 115(1-2):199–216, 2003.

Kyung H. Baik, Subhasish M. Chowdhury, and Abhijit Ramalingam. The effects of conflict budget on the intensity of conflict: An experimental investigation. Working paper, 2015.

Kyung Hwan Baik, Subhasish M. Chowdhury, and Abhijit Ramalingam. Group size and matching protocol in contests. Working paper, 2016.

Olaf Bock, Andreas Nicklisch, and Ingmar Baetge. hRoot: Hamburg registration and organization online tool. WiSo-HH Working Paper Series, Number 1, 2012.

Philip Brookins and Dmitry Ryvkin. An experimental study of bidding in contests of incomplete information. *Experimental Economics*, 17(2):245–261, 2014.

Albert Camus. *The Outsider, translated by Joseph Laredo*. Hamish Hamilton, London, 1982.

Timothy N. Cason, Anya C. Savikhin, and Roman M. Sheremeta. Behavioral spillovers in coordination games. *European Economic Review*, 56(2):233–245, 2012a.

Timothy N. Cason, Roman M. Sheremeta, and Jingjing Zhang. Communication and efficiency in competitive coordination games. *Games and Economic Behavior*, 76(1):26–43, 2012b.

Subhasish M Chowdhury and Roman M Sheremeta. Multiple equilibria in Tullock contests. *Economics Letters*, 112 (2):216–219, 2011.

Subhasish M. Chowdhury, Roman M. Sheremeta, and Theodore L. Turocy. Overbidding and overspreading in rent-seeking experiments: Cost structure and prize allocation rules. *Games and Economic Behavior*, 87:224–238, 2014.

Chen Cohen and Tal Shavit. Experimental tests of Tullock's contest with and without winner refunds. *Research in Economics*, 66(3):263–272, 2012.

Douglas D. Davis and Robert J. Reilly. Do too many cooks always spoil the stew? An experimental analysis of rent-seeking and the role of a strategic buyer. *Public Choice*, 95(1-2):89–115, 1998.

Emmanuel Dechenaux, Dan Kovenock, and Roman M Sheremeta. A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*, 18(4):609–669, 2015.

Cary Deck and Salar Jahedi. Time discounting in strategic contests. *Journal of Economics & Management Strategy*, 24(1):151–164, 2015.

Francesco Fallucchi, Elke Renner, and Martin Sefton. Information feedback and contest structure in rent-seeking games. *European Economic Review*, 64:223–240, 2013.

Marco Faravelli and Luca Stanca. When less is more: Rationing and rent dissipation in stochastic contests. *Games and Economic Behavior*, 74(1):170–183, 2012.

Urs Fischbacher. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10:171–178, 2007.

Miguel A. Fonseca. An experimental investigation of asymmetric contests. *International Journal of Industrial Organization*, 27(5):582–591, 2009.

Gerd Gigerenzer. The psychology of good judgment. Frequency formats and simple algorithms. *Medical Decision Making*, 16:273–280, 1996.

Gerg Gigerenzer and Ulrich Hoffrage. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102:684–704, 1995.

Sara Godoy, Miguel A. Melendez-Jimenez, and Antonio J. Morales. No fight, no loss: Underinvestment in experimental contest games. *Economics of Governance*, 16(1):53–72, 2015.

Gerald A. Goldin. Affect, meta-affect, and mathematical belief structures. In *Beliefs: A hidden variable in mathematics education?*, pages 59–72. Springer, 2002.

Ben Greiner. Subject pool recruitment procedures: Organizating experiments with ORSEE. *Journal of the Economic Science Association*, 1:114–125, 2015.

Benedikt Herrmann and Henrik Orzen. The appearance of homo rivalis: Social preferences and the nature of rent seeking. CeDEx discussion paper series, 2008.

Jakob Kapeller and Stefan Steinerberger. How formalism shapes perception: an experiment on mathematics as a language. *International Journal of Pluralism and Economics Education*, 4(2):138–156, 2013.

Changxia Ke, Kai A Konrad, and Florian Morath. Brothers in arms–an experiment on the alliance puzzle. *Games and Economic Behavior*, 77(1):61–76, 2013.

Erik O. Kimbrough and Roman M. Sheremeta. Side-payments and the costs of conflict. *International Journal of Industrial Organization*, 31(3):278–286, 2013.

Erik O. Kimbrough, Roman M. Sheremeta, and Timothy W. Shields. When parity promotes peace: Resolving conflict between asymmetric agents. *Journal of Economic Behavior & Organization*, 99:96–108, 2014.

Xiaojing Kong. Loss aversion and rent-seeking: An experimental study. CeDEx discussion paper series, 2008.

Wooyoung Lim, Alexander Matros, and Theodore L. Turocy. Bounded rationality and group size in Tullock contests: Experimental evidence. *Journal of Economic Behavior & Organization*, 99:155–167, 2014.

Shakun D. Mago, Roman M. Sheremeta, and Andrew Yates. Best-of-three contest experiments: Strategic versus psychological momentum. *International Journal of Industrial Organization*, 31(3):287–296, 2013.

Shakun D. Mago, Anya C. Samak, and Roman M. Sheremeta. Facing your opponents: Social identification and information feedback in contests. *Journal of Conflict Resolution*, 60:459–481, 2016.

Aidas Masiliunas, Friederike Mengel, and J. Philipp Reiss. Behavioral variation in Tullock contests. Working Paper Series in Economics, Karlsruher Institut für Technologie (KIT), 2014.

Edward L. Millner and Michael D. Pratt. An experimental investigation of efficient rent-seeking. *Public Choice*, 62: 139–151, 1989.

Edward L. Millner and Michael D. Pratt. Risk aversion and rent-seeking: An extension and some experimental evidence. *Public Choice*, 69:81–92, 1991.

Marta Molina, Susana Rodríguez-Domingo, María Consuelo Cañadas, and Encarnación Castro. Secondary school students errors in the translation of algebraic statements. *International Journal of Science and Mathematics Education*, pages 1–20, 2016.

John Morgan, Henrik Orzen, and Martin Sefton. Endogenous entry in contests. *Economic Theory*, 51(2):435–463, 2012.

Jan Potters, Casper G. De Vries, and Frans Van Winden. An experimental examination of rational rent-seeking. *European Journal of Political Economy*, 14(4):783–800, 1998.

Curtis R. Price and Roman M. Sheremeta. Endowment effects in contests. *Economics Letters*, 111(3):217–219, 2011.

Curtis R Price and Roman M Sheremeta. Endowment origin, demographic effects, and individual preferences in contests. *Journal of Economics & Management Strategy*, 24(3):597–619, 2015.

Roy Radner. Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives. *Journal of Economic Theory*, 22:136–154, 1980.

David Schmidt, Robert Shupp, and James M. Walker. Resource allocation contests: Experimental evidence. CAEPR working paper, 2006.

Pamela Schmitt, Robert Shupp, Kurtis Swope, and John Cadigan. Multi-period rent-seeking contests with carryover: Theory and experimental evidence. *Economics of Governance*, 5(3):187–211, 2004.

Roman M. Sheremeta. Experimental comparison of multi-stage and one-stage contests. *Games and Economic Behavior*, 68(2):731–747, 2010.

Roman M. Sheremeta. Contest design: An experimental investigation. *Economic Inquiry*, 49(2):573–590, 2011.

Roman M Sheremeta. Overbidding and heterogeneous behavior in contest experiments. *Journal of Economic Surveys*, 27(3):491–514, 2013.

Roman M. Sheremeta. Impulsive behavior in competition: Testing theories of overbidding in rent-seeking contests. Social Science Research Network working paper 2676419, 2015.

Roman M. Sheremeta and Jingjing Zhang. Can groups solve the problem of over-bidding in contests? *Social Choice and Welfare*, 35(2):175–197, 2010.

Jason F. Shogren and Kyung H. Baik. Reexamining efficient rent-seeking in laboratory markets. *Public Choice*, 69 (1):69–79, 1991.

Robert Shupp, Roman M. Sheremeta, David Schmidt, and James Walker. Resource allocation contests: Experimental evidence. *Journal of Economic Psychology*, 39:257–267, 2013.

Ferenc Szidarovszky and Koji Okuguchi. On the existence and uniqueness of pure nash equilibrium in rent-seeking games. *40 Years of Research on Rent Seeking 1: Theory of Rent Seeking*, 1:271, 2008.

Sheila Tobias and Carol Weissbrod. Anxiety and mathematics: An update. *Harvard Educational Review*, 50(1):63–70, 1980.

Gordon Tullock. Efficient rent seeking. In James M. Buchanan, Robert D. Tollison, and Gordon Tullock, editors, *Toward a theory of the rent-seeking society*, pages 97–112. Texas A&M University Press, College Station, TX, 1980.

Lian Xue, Stefania Sitzia, and Theodore L. Turocy. Mathematics self-confidence and the "prepayment effect" in riskless choices. *Journal of Economic Behavior & Organization*, 135:239–250, 2017.

A Instructions

The session consists of 30 decision-making periods. At the conclusion, any 5 of the 30 periods will be chosen at random, and your earnings from this part of the experiment will be calculated as the sum of your earnings from those 5 selected periods.

You will be randomly and anonymously placed into a group of 4 participants. Within each group, one participant will have ID number 1, one ID number 2, one ID number 3, and one ID number 4. The composition of your group remains the same for all 30 periods but the individual ID numbers within a group are randomly reassigned in every period.

In each period, you may bid for a reward worth 160 pence. In your group, one of the four participants will receive a reward. You begin each period with an endowment of 160 pence. You may bid any whole number of pence from 0 to 160; fractions or decimals may not be used.

If you receive a reward in a period, your earnings will be calculated as:

Your payoff in pence = your endowment – your bid + the reward.

That is, your payoff in pence = 160 – your bid + 160.

If you do not receive a reward in a period, your earnings will be calculated as:

Your payoff in pence = your endowment – your bid.

That is, your payoff in pence = 160 – your bid.

Portion for conventional treatment only

The more you bid, the more likely you are to receive the reward. The more the other participants in your group bid, the less likely you are to receive the reward. Specifically, your chance of receiving the reward is given by your bid divided by the sum of all 4 bids in your group:

$$\text{Chance of receiving the reward} = \frac{\text{Your bid}}{\text{Sum of all 4 bids in your group}}.$$

You can consider the amounts of the bids to be equivalent to numbers of lottery tickets. The computer will draw one ticket from those entered by you and the other participants, and assign the reward to one of the participants through a random draw.

An example. Suppose participant 1 bids 80 pence, participant 2 bids 6 pence, participant 3 bids 124 pence, and participant 4 bids 45 pence. Therefore, the computer assigns 80 lottery tickets to participant 1, 6 lottery tickets to participant 2, 124 lottery tickets to participant 3, and 45 lottery tickets for participant 4. Then the computer randomly draws one lottery ticket out of 255 (80 + 6 + 124 + 45). As you can see, participant 3 has the highest chance of receiving the reward: 0.49 = 124/255. Participant 1 has a 0.31 = 80/255 chance, participant 4 has a 0.18 = 45/255 chance and participant 2 has the lowest, 0.05 = 6/255 chance of receiving the reward.

After all participants have made their decisions, all four bids in your group as well as the total of those bids will be shown on your screen.

Participant ID	Bid
Participant 1	80
Participant 2	6
Participant 3	124
Participant 4	45
Total bids	255

Interpretation of the table: The horizontal rows in the left column of the above table contain the ID numbers of the four participants in every period. The right column lists their corresponding bids. The last row shows the total of the four bids. The summary of the bids, the outcome of the draw and your earnings will be reported at the end of each period.

At the end of 30 periods, the experimenter will approach a random participant and will ask him/her to pick up five balls from a sack containing 30 balls numbered from 1 to 30. The numbers on those five balls will indicate the 5 periods, for which you will be paid in Part 2. Your earnings from all the preceding periods will be throughout present on your screen.

Portion for ticket treatment only

The chance that you receive a reward in a period depends on how much you bid, and also how much the other participants in your group bid. At the start of each period, all four participants of each group will decide how much to bid. Once the bids are determined, a computerised lottery will be conducted to determine which participant in the group will receive the reward. In this lottery draw, there are four types of tickets: Type 1, Type 2, Type 3 and Type 4. Each type of ticket corresponds to the participant who will receive the reward if a ticket of that type is drawn. So, if a Type 1 ticket is drawn, then participant 1 will receive the reward; if a Type 2 ticket is drawn, then participant 2 will receive the reward; and so on.

The number of tickets of each type depends on the bids of the corresponding participant:

- Number of Type 1 tickets = Bid of participant 1
- Number of Type 2 tickets = Bid of participant 2

- Number of Type 3 tickets = Bid of participant 3
- Number of Type 4 tickets = Bid of participant 4

Each ticket is equally likely to be drawn by the computer. If the ticket type that is drawn has your ID number, then you will receive a reward for that period.

An example. Suppose participant 1 bids 80 pence, participant 2 bids 6 pence, participant 3 bids 124 pence, and participant 4 bids 45 pence. Then:

- Number of Type 1 tickets = Bid of participant 1 = 80
- Number of Type 2 tickets = Bid of participant 2 = 6
- Number of Type 3 tickets = Bid of participant 3 = 124
- Number of Type 4 tickets = Bid of participant 4 = 45

There will therefore be a total of 80 + 6 + 124 + 45 = 255 tickets in the lottery. Each ticket is equally likely to be selected.

In each period, the calculations above will be summarised for you on your screen, using a table like the one in this screenshot:

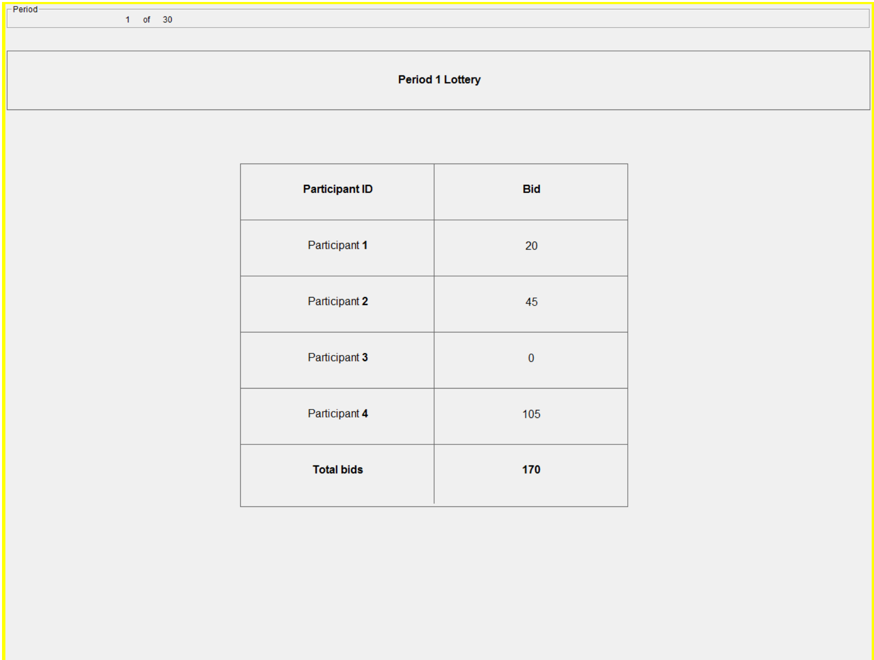
Participant ID	Bid	Ticket Type	Total tickets	Ticket number(s)
Participant 1	80	Type 1	80	1 - 80
Participant 2	6	Type 2	6	81 - 86
Participant 3	124	Type 3	124	87 - 210
Participant 4	45	Type 4	45	211 - 255

Interpretation of the table: The horizontal rows in the above table contain the ID numbers of the four participants in every period. The vertical columns list the participants' bids, the corresponding ticket types, total number of each type of ticket (second column from right) and the range of ticket numbers for each type of ticket (last column). Note that the total number of each ticket type is exactly same as the corresponding participant's bid. For example, the total number of Type 1 tickets is equal to Participant 1's bid.

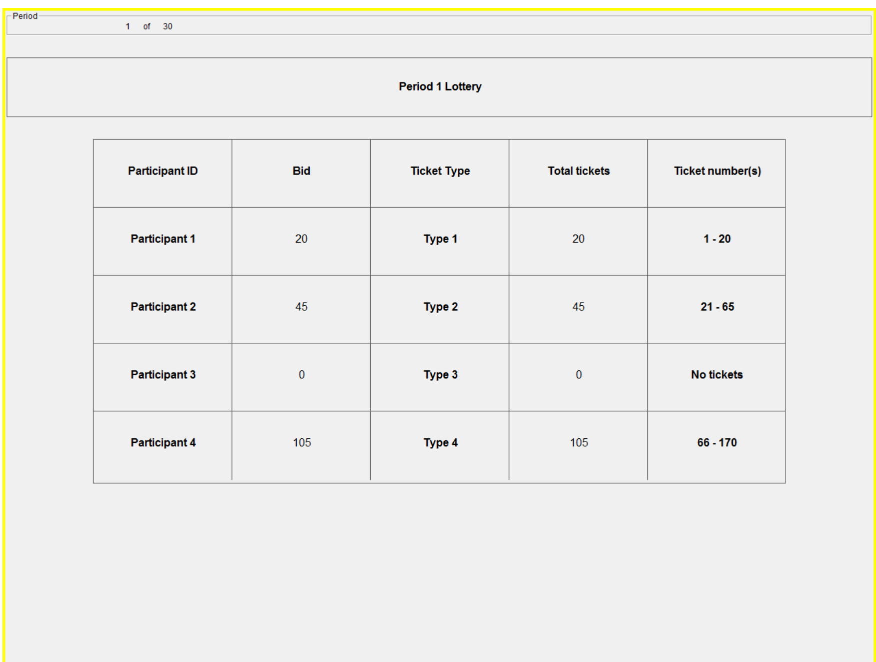
The last column gives the range of ticket numbers for each ticket type. Any ticket number that lies within that range is a ticket of the corresponding type. That is, all the ticket numbers from 81 to 86 are tickets of Type 2, which implies a total of 6 tickets of Type 2, as appears from the 'Total Tickets' column. In case a participant bids zero, there will be no ticket that contains his or her ID number. In such a case, the last column will show 'No tickets' for that particular ticket type.

The computer then selects one ticket at random. The number and the type of the drawn ticket will appear below the table. The ID number on the ticket type indicate the participant receiving the reward.

At the end of 30 periods, the experimenter will approach a random participant and will ask him/her to pick up five balls from a sack containing 30 balls numbered from 1 to 30. The numbers on those five balls will indicate the 5 periods, for which you will be paid in Part 2. Your earnings from all the preceding periods will be throughout present on your screen.

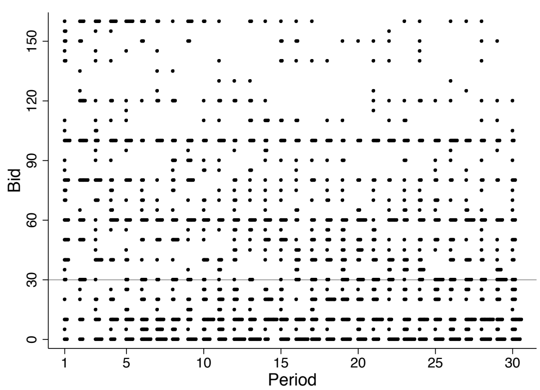


(a) Conventional treatment

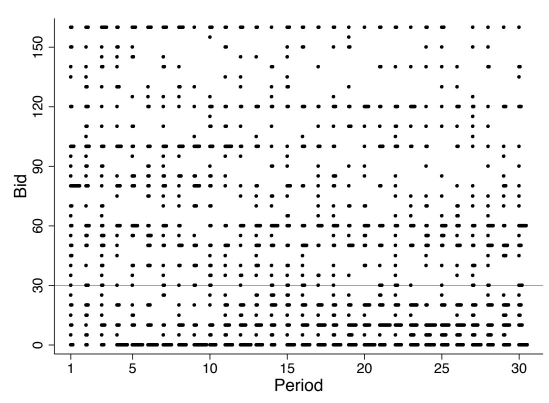


(b) Ticket treatment

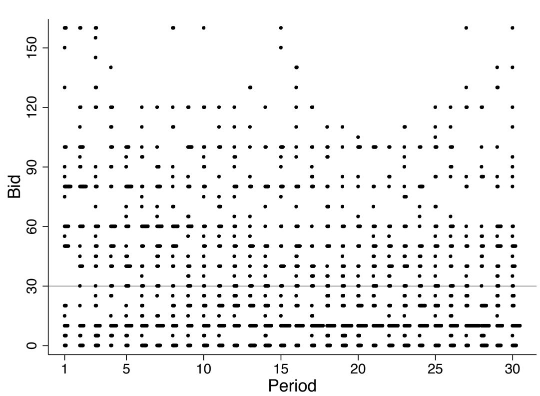
Figure 1: Comparison of bid summary screens



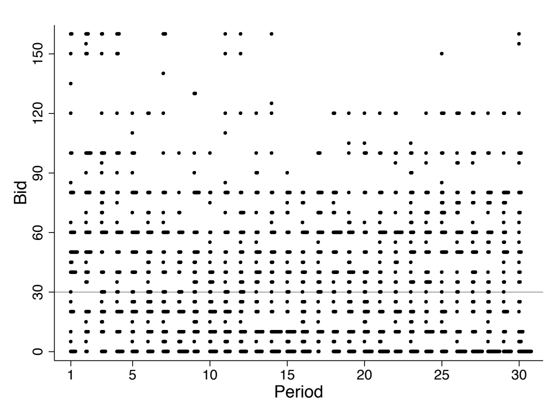
(a) Conventional, UK



(b) Conventional, US

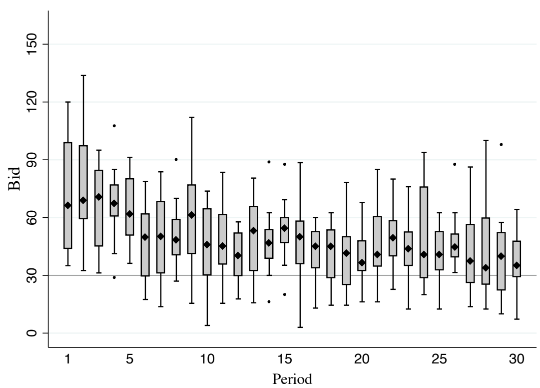


(c) Ticket, UK

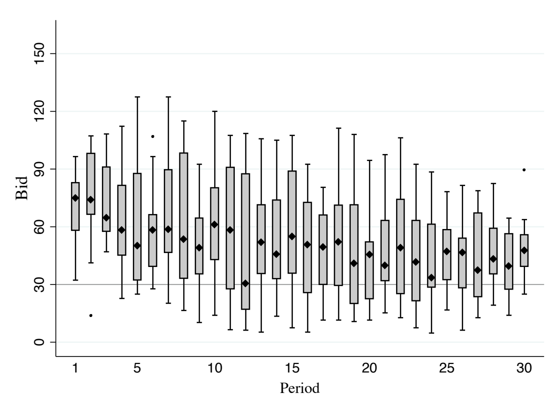


(d) Ticket, US

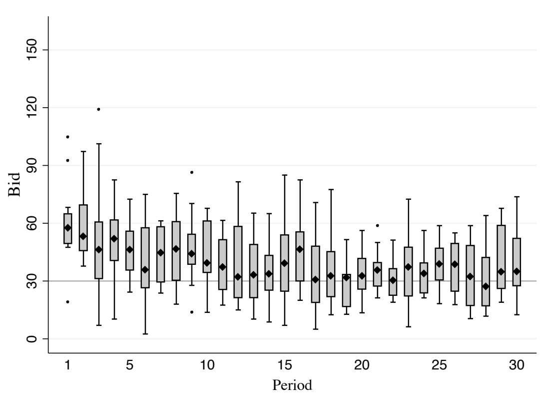
Figure 2: All bids by period, grouped by subject pool and treatment. Each dot represents the bid of one participant in one period. The Nash equilibrium bid of 30 is indicated by a horizontal line.



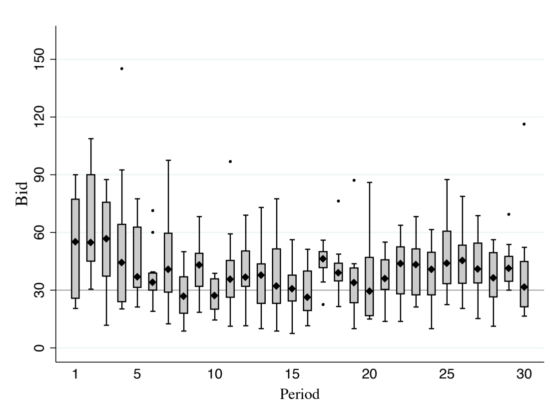
(a) Conventional, UK



(b) Conventional, US



(c) Ticket, UK



(d) Ticket, US

Figure 3: Distribution of group mean bids, by subject pool and treatment. For each period, the vertical boxes plot the interquartile range of average bids across groups. The black diamond indicates the median of the group averages.

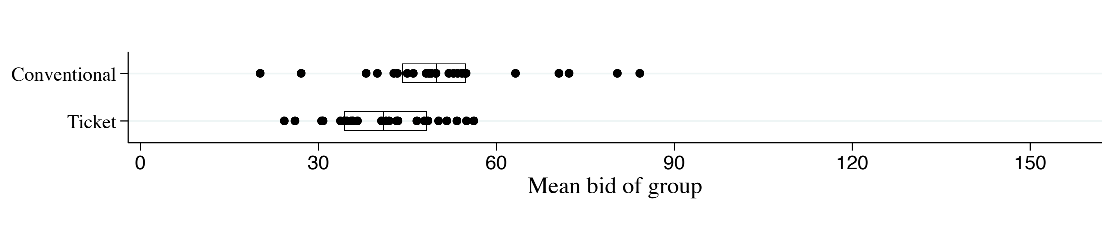


Figure 4: Distribution of mean bids for each group over the experiment. Each dot represents the mean bid of one group.

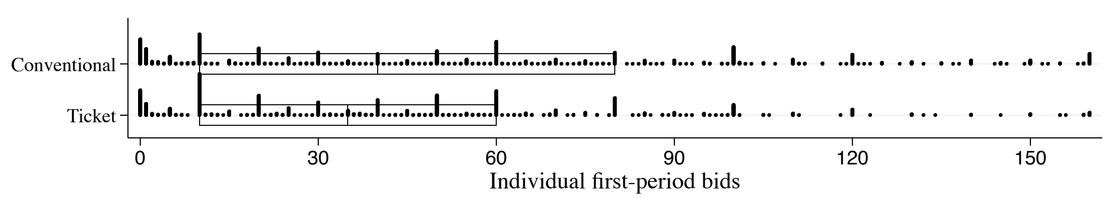
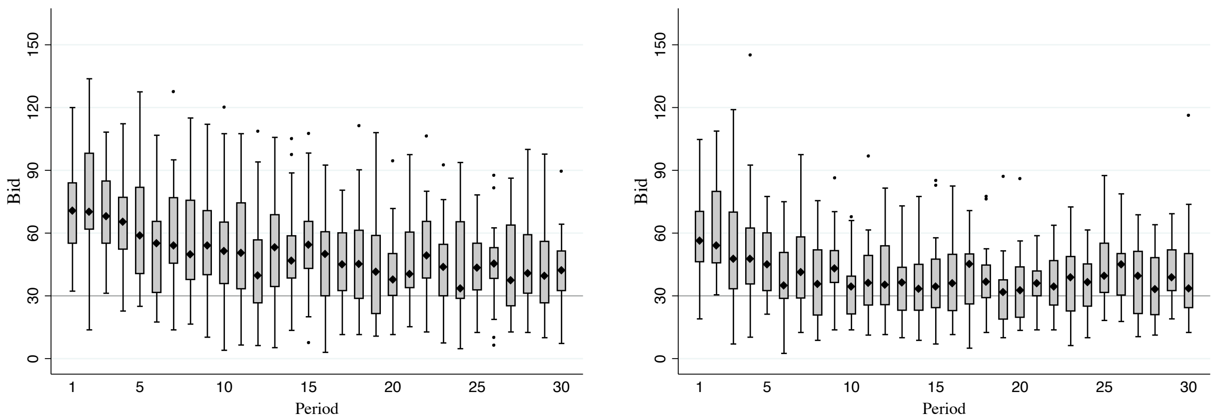


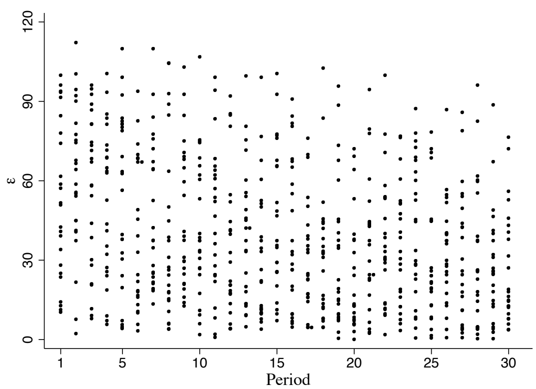
Figure 5: Distribution of first-period bids for all participants. Each dot represents the bid of one participant. For each distribution, the superimposed box indicates the median and the lower and upper quartiles.



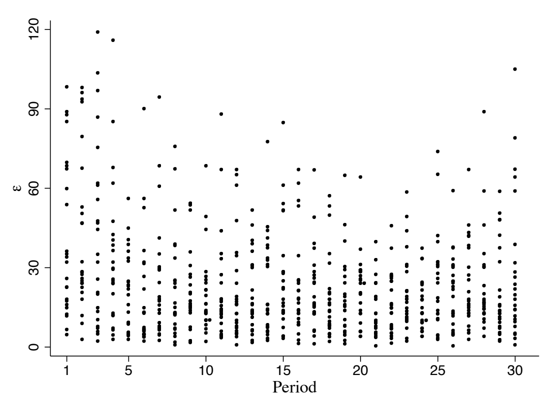
(a) Conventional

(b) Ticket

Figure 6: Evolution of group average bids over time. For each period, the vertical boxes plot the interquartile range of average bids across groups. The black diamond indicates the median of the group averages.



(a) Conventional



(b) Ticket

Figure 7: Ex-post measure of disequilibrium ε within groups, by period. Each dot corresponds to the value of the measure for one group in one period.

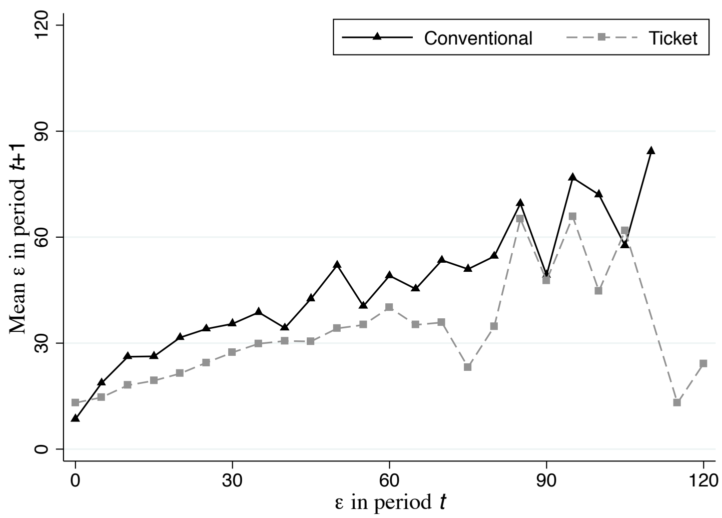


Figure 8: Expected value of disequilibrium measure ε in next period, as a function of a group's current ε.

Study	Treatment	Ratio rule	Example	Actual-to-Nash
Millner and Pratt (1989)	Lottery	Yes	Lottery	1.12
Millner and Pratt (1991)	Less risk averse	Yes	Lottery	1.23
Shogren and Baik (1991)	Lottery	No (payoff table)	—	1.01
Davis and Reilly (1998)	Lottery	—	—	1.46
Potters et al. (1998)	Lottery	Yes	Lottery	1.68
Anderson and Stafford (2003)	One-shot (n=2)	No	Lottery	1.79
	One-shot (n=3)	No	Lottery	1.50
	One-shot (n=4)	No	Lottery	1.87
	One-shot (n=5)	No	Lottery	3.00
	One-shot (n=10)	No	Lottery	3.46
Schmitt et al. (2004)	Static	Yes	Lottery	1.76
Herrmann and Orzen (2008)	Direct repeated	No	Wheel	2.05
Kong (2008)	Loss aversion	Yes	Lottery	1.81
Fonseca (2009)	Simultaneous	No	Lottery	2.00
Abbink et al. (2010)	1:1	No	Lottery	2.05
Sheremeta (2010)	Single	Yes	Lottery	1.52
Sheremeta and Zhang (2010)	Individual	Yes	Lottery	1.95
Ahn et al. (2011)	1V1	—	—	1.37
Price and Sheremeta (2011)	P treatment	Yes	—	1.90
Sheremeta (2011)	GC	Yes	Lottery	1.33
	SC	Yes	Lottery	1.31
Cason et al. (2012b)	Individual	No	—	1.26
Faravelli and Stanca (2012)	Standard Lottery	Yes	Lottery	1.10
Morgan et al. (2012)	Small Prize	No	Wheel	1.45
	Large Prize	No	Wheel	1.19
Cohen and Shavit (2012)	One-shot (w/o refund)	Yes	—	2.52
Mago et al. (2013)	High r	Yes	Lottery	1.90
	High r + IP	Yes	Lottery	1.51
	Low r + IP	Yes	Lottery	2.73
Ke et al. (2013)	Share/fight	Yes	Wheel	1.50
Kimbrough and Sheremeta (2013)	Baseline	Yes	Lottery	1.95
Cason et al. (2012a)	Blue & Green	Yes	Lottery	2.23
Shupp et al. (2013)	Single-prize (real lottery)	Yes	Wheel	0.73
Brookins and Ryvkin (2014)	Complete symmetric	Yes	—	1.42
Chowdhury et al. (2014)	PL	Yes	—	1.75
Lim et al. (2014)	n=2	Yes	Lottery	1.30
	n=4	Yes	Lottery	1.61
	n=9	Yes	Lottery	3.30
Fallucchi et al. (2013)	Lottery-Full	Yes	Lottery	1.25
Kimbrough et al. (2014)	Baseline unbalanced	Yes	Lottery	1.41
	Baseline balanced	Yes	Lottery	1.15
	Random unbalanced	Yes	Lottery	1.17
	Random balanced	Yes	Lottery	1.25
Masiliunas et al. (2014)	N1S1	Yes	Lottery	1.25
Deck and Jahedi (2015)	Experiment 1	Yes	—	1.64
Sheremeta (2015)	One-shot	Yes	Lottery	1.81
Price and Sheremeta (2015)	Gift / Yardstick	Yes	Lottery	1.92
Godoy et al. (2015)	Partner-Random	No	Lottery	0.66
	Partner-No allocation	No	Lottery	1.37
Baik et al. (2015)	Medium	Yes	Lottery	1.73
Baik et al. (2016)	Partner (n=3)	Yes	Lottery	1.53
	Partner (n=2)	Yes	Lottery	1.20
Mago et al. (2016)	NP-NI	Yes	Lottery	1.94
	NP-I	Yes	Lottery	1.89

Table 1: Summary of Tullock contest experiments. "Ratio rule" indicates whether the study gives chances of winning in terms of ratios of bids. "Example" describes whether and how the game is additionally explained in terms of a (pseudo-)physical mechanism. "Actual-to-Nash" is the reported ratio of average bids to the Nash baseline prediction.

	UK	US	All
Conventional	49.85	53.56	51.70
	(43.77)	(48.41)	(46.18)
	$N = 1440$	$N = 1440$	$N = 2880$
Ticket	40.35	41.08	40.72
	(35.96)	(35.36)	(35.65)
	$N = 1440$	$N = 1440$	$N = 2880$

Table 2: Descriptive statistics on individual bids. Each cell contains the mean, standard deviation (in parentheses), and total number of bids. The column All pools the bids from the two sites.

	Coefficient	Standard error	p-value
Intercept	18.77	1.39	$<$0.001***
ε_{csgt}	0.50	0.03	$<$0.001***
$\mathbf{1}_{c=\text{ticket}}$	-4.90	1.85	0.008**
$\mathbf{1}_{c=\text{ticket}} \times \varepsilon_{csgt}$	-0.11	0.05	0.024*

Table 3: Random-effects panel regression of evolution of disequilibrium measure ε over time. Dependent variable is $\varepsilon_{csg(t+1)}$; standard errors clustered at the session level. Overall $R^2 = 0.2955$. *** denotes significantly different from zero at 0.1%; ** at 1%, * at 5%.