Department of Labour Economics

HdBA
Hochschule der
Bundesagentur für Arbeit
University of Applied Labour Studies

HdBA Discussion Papers in Labour Economics

No. 25-01

Peer Evaluation Tournaments

Martin Dufwenberg

Katja Görlitz

Christina Gravert

Peer Evaluation Tournaments

Martin Dufwenberg[*] Katja Görlitz[†] Christina Gravert[‡]

Abstract: Peer evaluation tournaments are common in academia, the arts, and corporate environments. They make use of the expert knowledge that academics or team members have in assessing their peers' performance. However, rampant opportunities for cheating may throw a wrench in the process unless, somehow, players have a preference for honest reporting. Building on Dufwenberg and Dufwenberg's (2018) theory of perceived cheating aversion, we develop a multi-player model in which players balance the utility of winning against the disutility of being identified as a cheater. We derive a set of predictions, and test these in a controlled laboratory experiment.

Keywords: psychological game, cheating, tournaments, laboratory experiment
JEL codes: C91

1 Introduction

How can we reliably judge performance in situations where output is complex, multi-dimensional or involves team work? While a $100m$ track record can be objectively measured with a stopwatch, deciding who to award a Nobel Prize, Oscar or Grammy or promote to CEO requires expert evaluations, i.e., assessment by individuals who are themselves knowledgeable about the subject area. This often means that the ones who are best at evaluating are

[*]Department of Economics, University of Arizona; University of Gothenburg; CESifo; email: martind1@arizona.edu
[†]University of Applied Labour Studies; e-mail: katja.goerlitz@hdba.de
[‡]University of Copenhagen and CEBI; e-mail: cag@econ.ku.dk.

also the ones being evaluated. Peer evaluation tournaments are a common solution to this informational problem. For example, much of the scientific process is built on peer evaluations. Scientists working on related topics evaluate which papers will end up published in top journals, whose research topic will receive funding and which applicants will be hired. Only movie directors nominate other movie directors for the Oscars and vote for the films they directed themselves. In private companies, promotion decisions are often based on peer evaluations since team members can have the most knowledge about the non-tangible performance of their colleagues.[1]

However, winner-takes-all tournaments based on subjective evaluations create incentives for dishonesty. Scientific expert evaluators compete for the same journal space and grants as those whom they evaluate and consultants are rivals for the same promotions or bonuses that are given out based on their feedback. Given the possibility for dishonesty, it is perhaps surprising that these institutions are so prevalent in science, arts and business. For these institutions to work, the desire to cheat needs to be counteracted by, for example, the desire to be seen as an honest person which could affect individuals' professional reputation.

In this paper, we model the tradeoff between the desire to win by cheating in a peer evaluation tournament and the disutility from being perceived as a cheater by peers. We then test the model predictions in a novel lab experiment.

Our modeling approach builds on the recent theory by Dufwenberg and Dufwenberg (2018) (D&D) according to which people suffer perceived cheating aversion in proportion to how much others believe that they cheat. D&D considered a special setting with a single active decision maker, which we extend to a multi-player tournament setting in which players can cheat by over- and underreporting their delivered quality to win the tournament. The other

[1]The 360 degrees feedback app Leapsome claims to have over 1600 companies as clients, including many Fortune 500 companies and explicitly also mentions self-evaluation as a key feature of the tool.

players can observe the true quality of the output of each player (with some noise), but the submitted assessment is confidential and only the winner of the tournament can be observed. Each player suffers from perceived cheating aversion to the extent that, conditional on winning, the other players think they should not have won given their observation of his true quality. The model admits an equilibrium such that everyone cheats with positive probability and players with lower quality win less often. We also introduce a source of ambiguity. Players might cut other players some slack given the subjective nature of the quality ratings. This ambiguity can, however, be exploited by the players to hide some of their cheating.

We test our model predictions in a novel laboratory experiment. A group of five players who compete in a creativity task against each other, coming up with alternative uses for a piece of paper. Then, they evaluate their own and each other's quality. Based on their assessments the winner is determined. The players do not observe the ranking of everyone else, but only who the winner is. This feature makes individual cheating possible and not directly verifiable. The nature of the task, evaluating creative output, leaves room for ambiguity. In a between-subjects treatment, we manipulate the extent of the ambiguity regarding whether or not someone cheated, by providing a payment-irrelevant objective quality ranking to which the winner can be compared.

Our experimental setting is stark. If the players are entirely selfish, then the peer-evaluation institution that we study would be, essentially, useless. The incentives to over-report one's own performance, and to under-report that of others, would be so strong as to rule out any positive correlation between quality and reward. In contrast, if the predictions of our theory are supported, then this would provide some measure of hope that the institution we study is useful. We find that, as predicted, participants do cheat, but nevertheless, higher quality players are more likely to win. Being one rank higher on the objective ranking corresponds to a 9.5 percentage points higher probability of winning.

Our article relates to the literature on dishonesty in tournaments.[2] No previous work, however, has theoretically modeled peer evaluation tournaments. And we are only aware of one experimental paper that has empirically tested peer evaluation in a tournament (Carpenter et al., 2010). In their laboratory experiment, subjects compete in a real-effort task stuffing and addressing envelopes. In one treatment, subjects both counted the envelopes and evaluated the quality of the handwritten address labels of their rivals. Compensation depended in part on these peer evaluations. In line with our model predictions, they find that subjects frequently underreport the output of their competitors, but only by a small amount. In comparison with us, Carpenter et al. (2010) are mainly interested in the productivity effects of tournaments in which peer evaluation is present compared to classic tournaments. They conclude that classic tournaments are generally preferred, as workers anticipate sabotage by their peers and thus reduce their output. While Carpenter et al. (2010) mention that "While it is materially costless to sabotage another worker, it does not appear to be costless from a psychological point of view" (p. 507), they do not aim to explain why there is no complete sabotage or what these psychological costs entail. Understanding what counteracts the desire to cheat in tournaments is crucial in designing better mechanisms to further reduce sabotage.[3]

Given the prevalence of high-stakes peer evaluation tournaments, it is surprising that not more work has been done. Our paper thus addresses an important gap in the literature.

[2]A number of studies have looked at the performance effects of tournaments (Lazear, 2018), sabotage in tournaments (Piest and Schreck, 2021; Lazear, 1989; Cadsby et al., 2010; Harbring and Irlenbusch, 2008, 2011; Conrads et al., 2014; Charness et al., 2014; Balafoutas et al., 2020; Dato et al., 2022; Benistant and Villeval, 2019), or general antisocial behavior at the workplace (Gangadharan et al., 2020).

[3]Our paper also adds to the knowledge on the usefulness of 360 degrees feedback in general. While the previous literature sheds light on psychological or management aspects of 360 degree feedback (Beehr et al., 2001; Atkins and Wood, 2006; Buckingham and Goodall, 2015), less is known from an economic perspective (see Sliwka (2020) for a review of the literature on the economics of incentives in firms that includes subjective performance evaluations).

2 The Model

Die-roll warm-up D&D study the experimental die roll paradigm introduced by Fischbacher and Föllmi-Heusi (2013) (F&FH). Before we extend D&D's ideas to our setting it is useful to recall what F&FH and D&D did:

F&FH ask subjects to privately roll and then report the outcome of a die-roll. Subjects get paid in proportion to how high a number they report. Neither full honesty (each number is equally likely to be reported) nor full selfishness (everyone reports the highest-paying number) is observed, but something in between. Higher numbers are more likely to be reported although all reports occur with positive probability.

In D&D's theory nature draws $x \in \{0, ..., n\}$, $n \geq 1$, with probability $\pi_x \in (0,1)$, $\sum_x \pi_x = 1$.[4] A decision maker (DM) observes x and then reports $y \in \{0, ..., n\}$ after which he is paid y units of money. An audience observes y, but not x. Let $p(x'|y) \in [0,1]$ be the probability the audience assigns to $x = x'$ given y, $\sum_{x'} p(x'|y) = 1$. DM's utility at (x, y) equals

$$y - \theta \cdot \sum_{x'<y} p(x'|y) \cdot (y - x'). \tag{1}$$

The summation reflects how much DM is perceived to cheat by the audience; $\theta \geq 0$ measures DM's sensitivity to that. (1) is independent of x; DM cares about his image, not cheating per se. (1) depends on the audience's beliefs, via $p(x'|y)$, generating a psychological game (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009, 2022). D&D explore equilibria and their fit with data.[5]

Peer-evaluation tournaments We adapt D&D's notion of perceived cheating aversion to a tournament with $N > 1$ players, not a single DM. That

[4]F&FH's setup is the special case where $n = 5$ and $\pi_x = \frac{1}{6}$ for all x.

[5]Abeler et al. (2019) survey the (> 100) experiments that were conducted with the die-roll paradigm. They also discuss various theoretical approaches, and conclude that D&D's theory (along with another approach, represented by Gneezy et al. (2018) and Khalmetski and Sliwka (2019)) is most consistent with the data.

N-some constitute each others' audience. The counterpart to D&D's x is now the players' true qualities $(x_i)_{i \leq N} \in \mathbb{R}^N$, which, however, are given at the start of the analysis and known to all. Choose indices such that $x_j \geq x_i$ if $j > i$. The counterpart to D&D's y is each player i's report $y_i = (y_{ij})_{j \leq N} \in \mathbb{R}^N$, where y_{ij} is i's report of j's quality.[6] A mixed strategy s_i is a probability distribution across a (finite) subset of reports. Write $s_i(y_i)$ for the probability that s_i assigns to y_i; if $s_i(y_i) = 1$ we interpret that as the pure strategy y_i. A single winner j is determined based on who received the highest overall reported quality ($= \max_j \Sigma_{i \leq N} y_{ij}$), ties broken at random. The players do not observe each others' reports, only who won. The (normalized) material payoff equals 1 for the winner and 0 for all others.

To specify players' utilities, first define c_{y_i} and c_{s_i} as how much i cheats given y_i and s_i:

$$c_{y_i} = \max\{y_{ii} - x_i, 0\} + \frac{1}{N-1} \cdot \sum_{j \neq i} \max\{x_j - y_{ij}, 0\} \tag{2}$$

$$c_{s_i} = \sum_{\{y_i | s_i(y_i) > 0\}} s_i(y_i) \cdot c_{y_i} \tag{3}$$

The first term on the r.h.s. of (2) reflects how much i over-reports own quality (if a positive amount). Rather than cheat that way, i could instead under-report each of the others just as much; the second term reflects that. Finally, (3) computes an expected value, c_{s_i}, based on s_i.

Assume that the non-winners' utilities depend only on their material payoffs ($= 0$) while the winner's utility is affected also by how much others perceive that the winner cheats. We restrict attention to equilibria where all players correctly anticipate each other's behavior; for this purpose we only need limited notation about (implications of) beliefs: Let $\widetilde{c}_i$ be the expectation of every $j \neq i$ about c_{s_i}. Let $\theta \geq 0$ measure each player's sensitivity to perceived cheating. If i wins then i's utility equals

[6] We impose (at first) no lower or upper bound on y_{ij}; this allows us to highlight key intuitions in a stark way. We show later how and why our main insight translates to a setting (like our experiment) with more restricted strategy sets.

$$1 - \theta \cdot \widetilde{c}_i. \qquad (4)$$

Let w_i^s be the probability with which player i wins when $s = (s_i)_{i \leq N}$ is played; s is an equilibrium if two conditions hold: (i) each s_i maximizes i's expected utility given $\widetilde{c}_i$, and (ii) $\widetilde{c}_i = c_{s_i}$ whenever $w_i^s > 0$.[7] The following result provides a backdrop for our subsequent main show:

Observation 1: *(i) If $\theta = 0$ no equilibrium exists. (ii) If $\theta > 0$ an equilibrium exists. (iii) If s is an equilibrium in pure strategies and $w_i^s > 0$ for all i, then $x_i = x_j$ for all i, j.*

Proof: (i) Consider any strategy profile $s = (s_i)_{i \leq N}$. Some player j can be identified that does not win with probability 1. By deviating to a pure strategy $y_j = (y_{ji})_{i \leq N}$ such that y_{jj} is high enough while each y_{ji} is low enough j can guarantee a win. Hence s is not an equilibrium.

(ii) Select a pure strategy profile $s = (s_i)_{i \leq N} = (y_i)_{i \leq N}$ with a designated honest winner j, i.e., $c_j = 0$ and $\Sigma_{i \leq N} y_{ij} > \Sigma_{i \leq N} y_{ik}$ for all $k \neq j$. We show that s can be an equilibrium: First, construct $(\widetilde{c}_i)_{i \leq N}$ as required. Since j is not cheating and wins with probability 1 when s is played, we get $\widetilde{c}_j = c_{y_j} = 0$. Since the probability that $k \neq j$ wins equals 0, we can let $(\widetilde{c}_k)_{k \neq j}$ be anything. Choose each $\widetilde{c}_k$ for $k \neq j$ so high that $1 - \theta \cdot \widetilde{c}_k < 0$. Now verify that (playing y_i and s) each player maximizes expected utility, as required in equilibrium: Player j wins for sure, thereby maximizing his material payoff while avoiding any disutility from perceived cheating. Each player $k \neq j$ gets a utility of 0 (by definition, since k is not winning) while k would get less by any deviation such that k wins (since $1 - \theta \cdot \widetilde{c}_k < 0$).

(iii) Since $w_i^s > 0$ for all i, $\Sigma_{i \leq N} y_{ij}$ must be the same for all j. Moreover, j's utility must be non-negative since otherwise j could ensure utility

[7]If $w_i^s = 0$ then $\widetilde{c}_i$ may take any value. Bayes' rule does not restrict out-of-equilibrium beliefs. Formally, we are adapting a version of the notion of sequential equilibrium, as developed for finite psychological games by Battigalli & Dufwenberg (2009).

0 by lowering (only) y_{jj} (guaranteeing a non-win). Also, j's utility must be non-positive since otherwise j could increase his utility by raising (only) y_{jj} (increasing the probability of a win to 1) while losing nothing as regards perceived cheating. Hence, j's utility equals 0, so $1 - \theta \cdot \widetilde{c}_j = 0$ for all j. Hence $\widetilde{c}_j > 0$ takes the same value for all j, which (under equilibrium expectations) is impossible unless $x_i = x_j$ for all i, j. ■

Our main result The equilibrium s defined in the proof of Observation 1(ii) has a designated winner j. Any $k \neq j$ wins with probability 0. This is intuitively questionable for a couple of reasons: First, cheating either does not occur at all, or has no bearing on the outcome. This seems hard to reconcile with the cheating patterns documented by F&FH, where cheating was common even if moderate. Second, the off-path inference $\widetilde{c}_j$ needed to deter a deviation by $k \neq j$ may seem implausible; recall that $\widetilde{c}_j$ was chosen to be large enough, which is a stretch especially for players who would not have to cheat much (or at all) to win (e.g., player N, if the designated winner is some $j < N$). Third, the nature of s may seem at odds with evidence reported by Charness and Dufwenberg (2011). Their results, albeit from a different game-context, suggest that low-talent agents cannot be trusted unless opportunities are provided such that they may participate in a Pareto-improving outcome. Equilibrium s, as described in Observation 1(ii), does not exhibit such a property; there are no gains to be reckoned with for any $k \neq j$ relative to what k can get as a non-winner ($= 0$).

This begs the question whether there are equilibria where all players have a positive probability of winning,[8] and which are useful for selecting talents in the sense that players with higher quality are more likely to win. Observation 1(iii) reveals that, barring the knife-edge case where $x_i = x_j$ for all i, j, which seems unlikely to hold in most situations, the answer is *no* as regards equilibria in pure strategies. However, if we instead look for equilibria in

[8] Inspired by the title of the article by Charness & Dufwenberg (2011), one might say that such equilibria would involve participation (in expectation, as *all* players gain materially with some probability).

mixed strategies we have the following positive result. As economists should, we justify dubbing it our Theorem on grounds of economic significance rather than mathematical complexity:

Theorem: *If $\theta > 0$ then an equilibrium $s = (s_i)_{i \leq N}$ in mixed strategies exists such that $w_i^s > 0$ for all i and $x_j > x_i$ implies $w_j^s > w_i^s$.*

Proof: Such an s can be constructed in many ways; we highlight one possibility: Assume that each i randomizes across precisely two pure strategies: i either reports honestly, i.e., chooses $y_i^H := (y_{ij})_{j \leq N} = (x_j)_{j \leq N}$, or i cheats-to-T, i.e., chooses $y_i^T := (y_{ij})_{j \leq N}$ such that $y_{ij} = x_j$ for all $j \neq i$ and $y_{ii} > x_i$ such that $(N-1) \cdot x_i + y_{ii} = T$. Note that, by construction, $w_i^s > 0$ for all i.

For s to be an equilibrium, i must be indifferent between y_i^H and y_i^T, since $s_i(y_i^H) > 0$ and $s_i(y_i^T) > 0$. Player i must also be indifferent between winning and losing since these are the only two possible outcomes and he could have achieved either for sure by raising or lowering his cheating. It follows that i's utility of winning equals 0, since that's what i would get from losing (by assumption). Using (4), and the equilibrium condition that $\widetilde{c}_i = c_{s_i}$ for all i (which must hold since all i win with positive probability), the following equality is implied:

$$0 = 1 - \theta \cdot s_i(y_i^T) \cdot (T - N \cdot x_i). \tag{5}$$

The l.h.s. [r.h.s.] is the utility of losing [winning]. Note that $s_i(y_i^T) \cdot (T - N \cdot x_i) = c_{s_i}$ since $T - N \cdot x_i$ is the amount that i cheats when choosing y_i^T. We need to show that (5) can be satisfied for all i. Consider player N. Make (5) hold by selecting $s_N(y_N^T) \in (0, 1)$ and $T > N \cdot x_i$ appropriately. (This can be done in infinitely many ways, depending on the size of T.) Then consider $i < N$. Make (5) hold by selecting $s_i(y_i^T) \in (0, s_i(y_N^T)]$, appropriately. (There is a unique way to do this, as T is now given.) Finally, note that, by construction, $x_j > x_i$ implies $w_j^s > w_i^s$. ■

In the equilibrium s constructed in the proof, all players cheat, and win, with positive probablity. Regardless of θ, no one can be fully trusted. However, there is also good news: The higher is a player's quality, the more likely that player is to win.[9] Hence, if the players are motivated by perceived cheating aversion, a peer evaluation tournament may to some degree reveal competence even if the precision of that instrument is not **100**%.

ε-doubts It is arguably unrealistic to assume that qualities (i.e., $(x_i)_{i \leq N}$) are observed with full confidence, because there are many sources of ambiguity that we so far neglected.[10] As a means to take such doubts into account, we can introduce $\varepsilon \in [0, 1)$: Assume that player i uses strategy s_i. How much does i cheat? So far, we assumed that the answer was given by c_{s_i}, as defined by (3). However, to capture the effect of doubts one may replace $\widetilde{c}_i$ in (4) by $\frac{(1-\varepsilon)\cdot\widetilde{c}_i}{(1-\varepsilon)\cdot\widetilde{c}_i+\varepsilon}$ to get a new ε-affected utility of winning for player i:

$$1 - \theta \cdot \frac{(1-\varepsilon)\cdot\widetilde{c}_i}{(1-\varepsilon)\cdot\widetilde{c}_i+\varepsilon}. \tag{4'}$$

If $\varepsilon = 0$ we reproduce (4) but if $\varepsilon > 0$ this biases the calculation. The idea: j cuts i some slack, and this also affects i's image concern. The interpretation is that j assigns probability ε to the possibility that, for each y_i and c_{y_i} that figure in the summation to get (3), i actually isn't knowingly cheating but rather making his realized choice for any other reason. Our final result shows that the statement in the Theorem is still valid, although the overall amount of cheating increases with ε:

[9]In the constructed equilibrium, cheating involved only over-reporting of own quality. However, other equilibria could be constructed that also (or instead) involve under-reporting of others' qualities. To see this, recall (2) and note that the marginal effect to every player of i adding an amount Δ to his reported own score is the same as that of i deducting Δ from the reported score of every other player. *Mutatis mutandis*, all equilibria thus constructed would share the highlighted properties.

[10]Maybe there is noise in how players evaluate quality, others' & own. Maybe i mistakenly believes his quality is higher than it is, so that he isn't knowingly cheating even if he gives himself a high score. Or maybe i has high quality and it is j who is mistaken.

Observation 2: *Redefine the definition of equilibrium using (4') rather than (4), ceteris paribus. (i) If* $\theta > 0$ *then an equilibrium* $s = (s_i)_{i \leq N}$ *in mixed strategies exists such that* $w_i^s > 0$ *for all* i *and* $x_j > x_i$ *implies* $w_j^s > w_i^s$*. (ii) Moreover,* c_{s_i} *is strictly increasing in* ε*, for all* i*.*

Proof: (i) Note that, apart from drawing on (4') rather than (4), the arguments in the proof of our Theorem reproduce exactly.

(ii) Let s be any equilibrium in mixed strategies such that $w_i^s > 0$ for all i and $x_j > x_i$ implies $w_j^s > w_i^s$. Reasoning analogously as in the proof of the Theorem, one sees that each i must be indifferent between winning and not winning. Hence (4') takes the value of 0, for all i. In equilibrium, $\widetilde{c}_i = c_{s_i'}$, so we get $\frac{(1-\varepsilon)\cdot\widetilde{c}_i}{(1-\varepsilon)\cdot\widetilde{c}_i+\varepsilon} = \frac{1}{\theta}$, implying that $c_{s_i'}$ is strictly increasing in ε. ■

While Observation 2(ii) shows that more ambiguity induces more cheating, this does not imply that ambiguity necessarily adversely affects the effectiveness of a team evaluation tournaments for revealing competence. To see this, fix an equilibrium $s = (s_i)_{i \leq N}$ as highlighted in the proof of our Theorem, where each i randomizes over two choices y_i^H and y_i^T. Using the new equilibrium definition, based on (4') rather than (4), it is clear that if $\varepsilon > 0$ then s is not sustainable as an equilibrium, as the amount of cheating is too low (given s). However, we can create a new, similar in spirit, equilibrium s' $= (s_i')_{i \leq N}$ where each i randomizes over choices y_i^H and $y_i^{T'}$. Note that this can be done in many ways, depending on the sizes of T' and $s_i'(y_i^{T'})$. If, for example, $T' > T$ while $s_i'(y_i^{T'}) = s_i(y_i^T)$ for all i, then w_i is not changed for any i; the peer evaluation tournament with ambiguity is then as efficient at selecting high-quality winners as the tournament without ambiguity is. On the other hand, if $T' = T$ while $s_i'(y_i^{T'}) > s_i(y_i^T)$, then the differences between the players' win-probabilities is reduced. All in all, under our theory, it is a open (and empirical) question whether ambiguity adversely affects the efficiency of a peer evaluation tournament institution.

Restricted strategy sets The theory we have so far developed imposes no bounds on players' strategy sets. Our experimental design, however, uses restricted strategy sets. We now explain how & why the spirit and relevance of the above analysis is nevertheless preserved: Suppose that s is an equilibrium, as described in the Theorem (or Observation 2(i)). Let $\overline{y}$ be the highest value of y_{ij}, chosen with positive probability given s, across all i and j. Now consider a modified game such that reported qualities $y_{ij} \leq M$ for all i, j. Obviously, as long as $M \geq \overline{y}$, s remains an equilibrium in the new game. However, it may seem that a problem occurs if instead $M < \overline{y}$. This is, however, a mirage in the sense that one can redefine units, making $\overline{y}' < M$ the new $\overline{y}$, as follows: Redefine qualities $(x_j)_{j \leq N}$ and sensitivity θ as $(x'_j)_{j \leq N}$ and θ' such that $x'_j = \frac{\overline{y}'}{\overline{y}} \cdot x_j$ for all j and $\theta' = \frac{\overline{y}}{\overline{y}'} \cdot \theta$. Mutatis mutandis, this allows for re-creating the old equilibrium s, except that all reports are scaled down by the factor $\frac{\overline{y}'}{\overline{y}}$. θ is raised to θ', but this is just matching the quality adjustments, like the value of money is invariant to currency conversions.[11]

We also presumed that there are no integer constraints on reports. Further adjustments render that issue moot. For example, and in anticipation of our experiment let $N = 5$ and suppose that $y_{ij} \in \{0, 1, ..., 10\}$ for all i, j. We can now create equilibria such that properties implied under the Theorem (or Observation 2(i)) occur. For example, following the previous paragraph, describe qualities as fractions of 1: $x'_i \in [0, \delta]$ where $\delta \leq 1$. Now, in the spirit of the construction made in the proof of the Theorem, assume that each player i randomizes between two reports: $y_i^0 = (y_{ij})_{j \leq N} = (0, 0, ..., 0)$ or y_i^T which is just like y_i^0 except that $y_{ii} = T = \{1, 2, .., 10\}$. As long as the difference $T - \delta$ is large enough, this pattern is compatible with an equilibrium s in mixed strategies such that $w_i^s > 0$ for all i. Of course, that particular example is crude and extreme, as only two choices are made; however, that feature can

[11]The arguments made here assume that $\theta > 0$. In instead $\theta = 0$, unlike in Observation 1(i), we get equilibrium existence: Each i has a dominant strategy to cheat maximally, and in equilibrium all players do so. The outcome is obviously useless as regards promoting winners with higher quality, as all players tie on their total scores and thus have the same (=$1/N$) chance of winning.

be relaxed. The essential point is that players either back off or compete, and in equilibrium they are indifferent between these modes of behavior.

Wrap up To sum up, in the equilibrium described in the Theorem (and Observation 2(i)) players either report honestly, and likely don't win, or they exaggerate, and have a decent shot of winning. They are indifferent between these two modes of behavior, because, in order to win, the degree of cheating needed is just high enough that (conditional on winning) the sweetness of the material prize is exactly counterbalanced by pangs delivered via others' inference that one cheated. If there is ambiguity as regards to quality-evaluations, as we propose to capture via the parameter ε, then the higher is ε the more sheltered players are from others' opprobrium and the more likely they are to cheat. This, however, does not imply that the institution of a peer evaluation tournament is less efficient at selecting winners with some precision.

These results provide reasons to be hopeful that peer evaluation tournaments can function well, despite that they would be destined to fail if players were selfish rather than affected by perceived cheating aversion. In the next sections, we test whether the rosy parts of the theoretical story we have told are empirically relevant.

3 Experimental Design

We now turn to the experiment that tests the theoretical predictions of our model. Participants compete in groups of five for a prize. Each player performs a task that all players (including the player him or herself) rate on a scale from 0 to 10. The person with the highest total score in the group wins. To make peer evaluation meaningful and to introduce some ambiguity ε, we chose a creativity task that is subjective, leaving scope for cheating when rating one's competitors. We used the "unusual uses task" (Guilford, 1967) where players should come up with as many unconventional uses for a piece of paper (e.g. make a hat, dry wet shoes, insulate a house) that they could think of. The experiment was run in sessions consisting of ten players. Within each session, participants were randomly assigned to their competitors.

The experimental set-up All participants were informed that they take part in a tournament consisting of five players where the winner gets a 500 SEK (≈50 USD) prize and the losers received a 50 SEK (≈5 USD) show up fee. To guarantee anonymity, each subject got an ID number assigning them randomly to a group of five players who they would compete against for the prize. Next, the creativity task was introduced and explained in detail. Subjects were informed that it was in their best interest to perform well in the task to increase their chances of winning. Importantly, they were not informed about the scoring mechanism when performing the task to rule out that the scoring could affect the creative performance of the players.

After a three-minute practice round on unusual uses for an old tire, the experimenter distributed the sheets for the incentivized task. Subjects had three minutes to come up with unusual uses for a piece of paper. When the time was up, the experimenter collected all answer sheets and handed out new instructions with the scoring rules. Subjects then received a copy of the answer sheet of each member of their group including their own and were asked to score each answer sheet on a score from 0 to 10. The instructions stated: The winner of the 500 SEK will be determined by the following procedure: You will now evaluate your own answer and the answers of the other four players with whom you compete for the 500 SEK. Please evaluate the answers with respect to their originality. Originality is scored for each person on a scale from 0 to 10 where 0 indicates overall not at all original answers and 10 very original answers. For scoring, take into consideration i) the number of answers, ii) their degree of being unusual and iii) the number of different categories they come from. The other players in your group will also do the same scoring. The points given and the points received are kept anonymous by the research team as well as the information who are

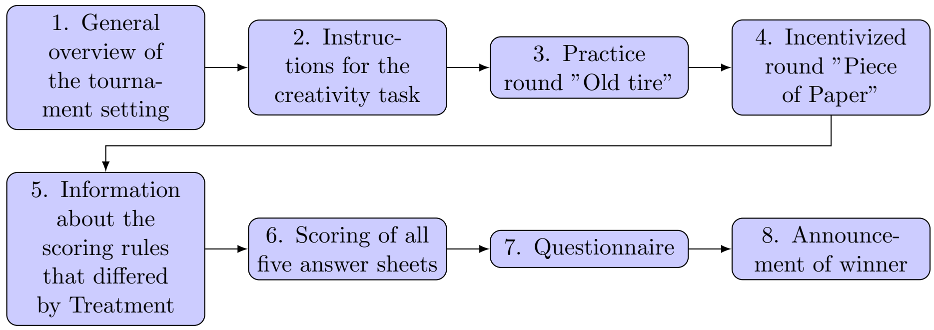

Figure 1: Experimental Design

the players in your group. For each player, the research team will add up the points given to a TOTAL SCORE of a minimum of 0 and a maximum of 50. The person with the highest TOTAL SCORE out of your group will receive 500 SEK (including the show-up fee). The ID of the winner was announced at the end of the experiment. Individual scores and the identity of the winner were kept anonymous by the experimenter. If there was a tie for the highest TOTAL SCORE, the winner was determined randomly which only happenend once. After scoring, subjects filled out a questionnaire on general characteristics and demographics. We summarize the process in Figure 1.

Treatments There is a baseline treatment and an objective-ranking treatment. Both treatments followed the process as outlined above and the scoring rules were the same. The only difference was that players in the baseline treatment only got the scoring rules, while players in the objective-ranking treatment obtained further information after the scoring rules were announced, but before they made their scoring. In particular, they were informed that we used the answers of more than 100 test persons who did the same task in a prior experiment to generate (what we called) an objective score for each participant. We also informed them that each player would be able to see these OBJECTIVE SCORES for each player in their group at the end of the experiment. The instructions emphasized again that winning is solely determined based on the TOTAL SCORE (composed of the scores given by one's players in the group and one's own assessment of the task). The objective-ranking treatment reduced the uncertainty about the objective creativity. With this treatment variation, we intended to reduce ε from our model, which is defined as any ambiguity in regard to assessing a players cheating.[12]

Calculating the objective rank We created the objective ranking following Bradler et al. (2019). The objective creativity increases with i) the number of valid answers, ii) the number of distinct categories the ideas come from (e.g. ring, necklace, bracelet belong only to one category which is jewelry) and iii) originality. Original ideas are those being mentioned less than 8% and very original answers were named less than 1%. Because Bradler et al. (2019) provided us with their participants' answers of the unusual uses task for a piece of paper, we could calculate for each answer how often it was mentioned. [13] The objective quality rank measures each players' quality by group based on the objective score on a scale from 1 (lowest objective score) to 5 (highest objective score).

Further information The experiment was conducted at the experimental laboratory of the University of Gothenburg. It was a pen and paper experiment and all earnings were paid out in cash or via a direct payment app right after the experiment. We conducted 14 sessions with ten participants in each session, 70 participants per treatment. Treatments were assigned at the session level. The sessions lasted up to 60 minutes. Average earnings were 140

[12]We conducted one additional treatment which is not suitable to test the theory, as individuals did not see the objective ranking in the end.

[13]To test the validity of the objective ranking, we asked 278 students at the University of Applied Labour Studies to assess the objective quality of the answers of five players of a random group in a survey in 2023. Their answers are highly correlated with the objective rank.

SEK. 57 percent of the participants were female.[14] See the Appendix for the full instructions and consent form.

4 Results

Do players cheat? The strategy that maximizes each player's probability of winning is to give themselves the best possible score (10) and all the other players the worst possible score (0). However, this strategy of reporting 10-0-0-0-0 is only followed by 5.7 percent of the individuals which corresponds to 8 individuals only (6 in the baseline treatment and 2 in the objective scores treatment). This low amount of individuals playing the payoff-maximizing strategy suggests that the majority of players were not solely motivated by winning the prize.

Nevertheless, we find that participants are significantly more likely to give themselves higher scores on average than what they give to the rest of their group. If everyone was correctly reporting x_i, then on average self-reportings and peer evaluations should have the same mean. In Figure 2, for a sample pooling both treatment groups, the average number of points given to oneself is 7.14 and the average number of points to others is 4.32 ($p < 0.001$). This difference suggests that individuals are cheating to their advantage, even if they are not choosing to cheat maximally.

Do all players win with positive probability? Our theorem implies that all players have a positive probability of winning, but that higher-quality individuals win more often. Figure 3 documents that this is generally the case. The probability of winning increases by objective rank (Corr = 0.34). Figure 3 shows that individuals with the best objective quality (rank 5) wins with a probability of 38 percent and the second best player with 30 percent. The individuals with third and fourth highest quality win with

[14] We have a short questionnaire at the end of the experiment, but as the demographics to not relate to our model, the analysis is not included in this paper.

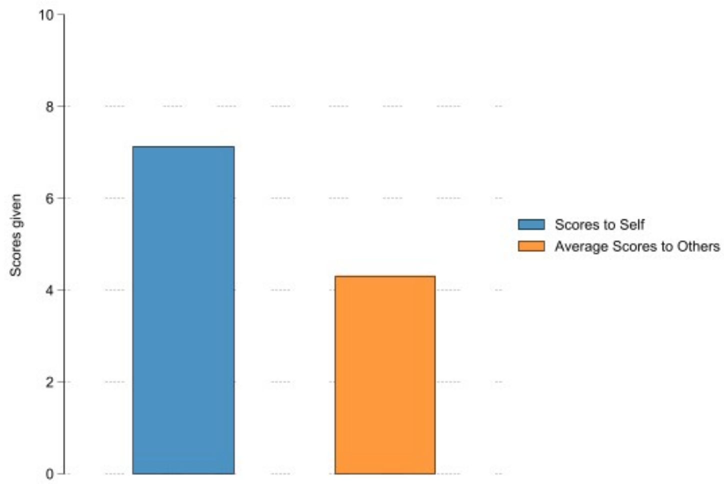

Figure 2: Average scores given to oneself and to others in the group

19 and 11 percent, respectively. However, the worst player (rank 1) never wins in our data. In sum, the data show that also lower ranked players have a positive probability of winning, even though we cannot fully confirm our hypothesis. To estimate how the effect of having a higher quality rank affects the probability of winning, we estimate a simple linear probability model. Rank is a catagorical variable (1-5) and winning is a binary variable that is one if the player won and zero otherwise. We cluster standard errors by session. Our linear probability model regression in column 1 of Table 1 shows that each point higher on the objective ranking, increases players' probability of winning by 9.5 percentage points (p=0.001).

Does reducing ambiguity affect cheating? Finally, we test whether changing the ambiguity of the evaluations affects cheating. Our Observation 2 (ii) states that more ambiguity induces more cheating, but we show theoretically, that this does not imply that ambiguity necessarily adversely affects the effectiveness of a team evaluation tournaments for revealing competence. We vary the level of ambiguity in our experiment by informing participants

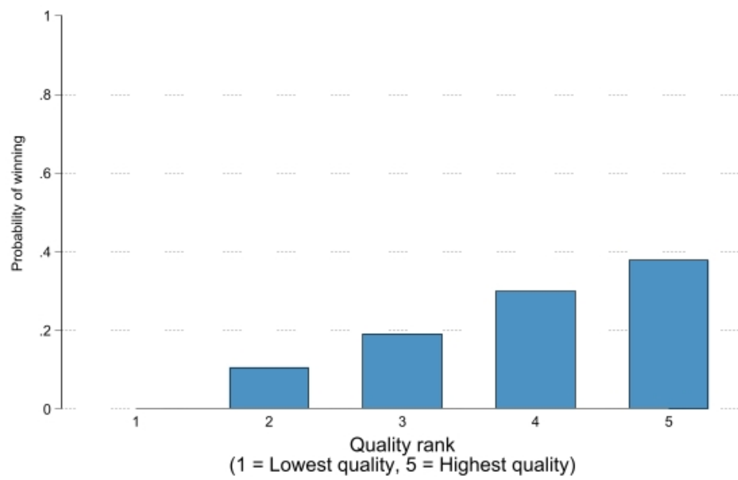

Figure 3: Winning probability by the objective rank according to the objective score

Table 1: Winning Probability

	(1) Winner	(2) Fair Winner
Rank (1-5)	0.095***	
	(0.023)	
Objective Ranking		0.014
		(0.043)
Constant	-0.089	0.071**
	(0.071)	(0.028)
Observations	140	140
R^2	0.114	0.001

OLS regression. Standard errors clustered by session. Winner is a dummy equal to one if chosen as winner. Fair Winner is a dummy equal to one if rank 5 was chosen as winner * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

in our objective-ranking treatment group that everyone will receive an objective ranking of the participants in their group to compare their rankings and the winner to. In the control group, this objective ranking was unannounced. We would therefore expect higher levels of cheating in the control group than in the treatment group. We make no predictions about whether this affects the effectiveness of the tournament in declaring the best player as the winner.

Figure 4 shows the winning probabilities by treatment. The figure suggests that individuals with the highest rank (5) are more likely to win when ambiguity is reduced. To test whether our treatment increases the probability that the rightful winner is chosen, we regress our treatment dummy on a binary variable that is equal to one if the group winner is also the person with the highest objective rank. The second column of Table 1 shows a positive coefficient for the objective treatment increasing the probability that the player with the highest objective quality wins, but we are underpowered to detect so small effects. Hence, we cannot conclude that in our experiment reducing ambiguity had a significant effect on the effectiveness of the tournament in selecting a high-quality winner.

5 Conclusion

Many high-stakes environments such as science, arts and business rely on expert evaluators, who can adequately judge the contribution of an individual in their field. Given the high level of expertise needed, the evaluators are often also competitors of the ones they are evaluating. This setting creates incentives to cheat. If individuals were completely selfish, peer tournaments would be more or less useless, as winning would be unrelated to the actual performance of the persons who are evaluated. Given the importance and prevalence of these types of peer evaluation tournaments, it is surprising that there has been so far very little work on their usefulness in selecting high-quality winners.

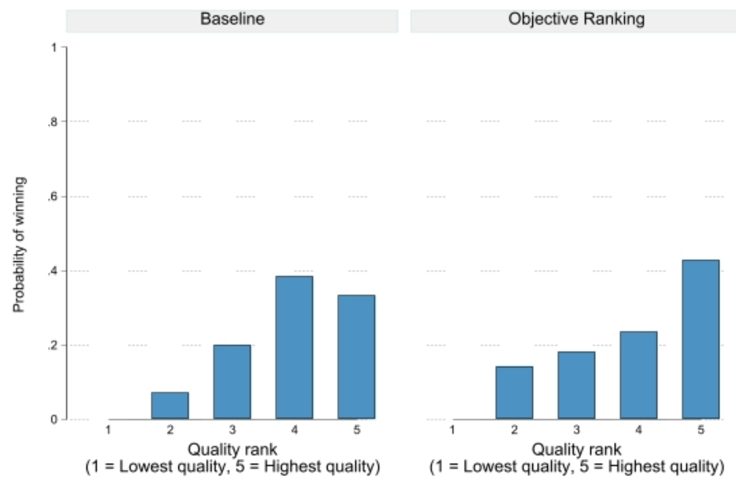

Figure 4: Winning probability by rank according to the objective score - by treatment

In this paper, we provide the first model on decision making in a peer evaluation tournament. Players trade off the desire to win with the disutility from being perceived as a cheater. Applied to the scientific process this would mean that if scientists care about their reputation among their peers, they will balance their desire to achieve professional success with the potential image loss of being perceived as dishonest.

In our laboratory experiment, we find evidence that supports our model predictions. While cheating is prevalent, higher quality winners are significantly more likely to win. On average, the players with the highest quality win 38% of the time and those with the lowest quality never win. Thus, our model and our empirical findings provide (fairly) good news for expert evaluation systems. We do not find conclusive evidence of whether reducing ambiguity affects the effectiveness of peer evaluation tournaments. In practice, peer evaluation tournaments are already often complemented by for example, an editor reviewing reports from referees or a manager supplementing peer evaluations from colleagues with additional performance data. Future work could test further strategies to increase honesty and test our model in a field setting, such as a 360 degrees app or other peer evaluation tournament settings.

References

Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for truth-telling. *Econometrica 87*(4), 1115–1153.

Atkins, P. W. B. and R. Wood (2006). Self- versus others' ratings as predictors of assessment center ratings: Validation evidence from 360-degree feedback programs. *Personnel Psychology 55*(4), 871–904.

Balafoutas, L., S. Czermak, M. Eulerich, and H. Fornwagner (2020). Incentives for dishonesty: An experimental study with internal auditors. *Economic Inquiry 58*(2), 764–779.

Battigalli, P. and M. Dufwenberg (2009). Dynamic psychological games. *Journal of Economic Theory 144*(1), 1–35.

Battigalli, P. and M. Dufwenberg (2022). Belief-dependent motivations and psychological game theory. *Journal of Economic Literature 60*(3), 833–882.

Beehr, T. A., L. Ivanitskaya, C. P. Hansen, D. Erofeev, and D. M. Gudanowski (2001). Evaluation of 360 degree feedback ratings: relationships with each other and with performance and selection predictors. *Journal of Organizational Behavior 22*(7), 775–788.

Benistant, J. and M. C. Villeval (2019). Unethical behavior and group identity in contests. *Journal of Economic Psychology 72*, 128–155.

Bradler, C., S. Neckermann, and A. J. Warnke (2019). Incentivizing creativity: A large-scale experiment with performance bonuses and gifts. *Journal of Labor Economics 37*(3), 793–851.

Buckingham, M. and A. Goodall (2015). Reinventing performance management. how one company is rethinking peer feedback and the annual review, and trying to design a system to fuel improvement. *Harvard Business Review*, 1–10.

Cadsby, C. B., F. Song, and F. Tapon (2010). Are you paying your employees to cheat? an experimental investigation. *The BE Journal of Economic Analysis & Policy 10*(1).

Carpenter, J., P. H. Matthews, and J. Schirm (2010). Tournaments and office politics: Evidence from a real effort experiment. *American Economic Review 100*(1), 504–17.

Charness, G. and M. Dufwenberg (2011). Participation. *American Economic Review 101*(4), 1211–1237.

Charness, G., D. Masclet, and M. C. Villeval (2014). The dark side of competition for status. *Management Science 60*(1), 38–55.

Conrads, J., B. Irlenbusch, R. M. Rilke, A. Schielke, and G. Walkowitz (2014). Honesty in tournaments. *Economics Letters 123*(1), 90–93.

Dato, S., E. Feess, and P. Nieken (2022). Lying in competitive environments: A clean identification of behavioral impacts.

Dufwenberg, M. and M. A. Dufwenberg (2018). Lies in disguise–a theoretical analysis of cheating. *Journal of Economic Theory 175*, 248–264.

Fischbacher, U. and F. Föllmi-Heusi (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association 11*(3), 525–547.

Gangadharan, L., P. J. Grossman, and J. Vecci (2020). Antisocial behavior in the workplace. *Handbook of Labor, Human Resources and Population Economics*.

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989). Psychological games and sequential rationality. *Games and Economic Behavior 1*(1), 60–79.

Gneezy, U., A. Kajackaite, and J. Sobel (2018). Lying aversion and the size of the lie. *American Economic Review 108*(2), 419–53.

Guilford, J. P. (1967). Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior 1*(1), 3–14.

Harbring, C. and B. Irlenbusch (2008). How many winners are good to have?: On tournaments with sabotage. *Journal of Economic Behavior & Organization 65*(3-4), 682–702.

Harbring, C. and B. Irlenbusch (2011). Sabotage in tournaments: Evidence from a laboratory experiment. *Management Science 57*(4), 611–627.

Khalmetski, K. and D. Sliwka (2019). Disguising lies—image concerns and partial lying in cheating games. *American Economic Journal: Microeconomics 11*(4), 79–110.

Lazear, E. P. (1989). Pay equality and industrial politics. *Journal of Political Economy 97*(3), 561–580.

Lazear, E. P. (2018). Compensation and incentives in the workplace. *Journal of Economic Perspectives 32*(3), 195–214.

Piest, S. and P. Schreck (2021). Contests and unethical behavior in organizations: A review and synthesis of the empirical literature. *Management Review Quarterly 71*(4), 679–721.

Sliwka, D. (2020). Bonus plans, subjective performance evaluations, and career concerns. *Handbook of Labor, Human Resources and Population Economics*.

6 Acknowledgements

We thank Arne Jonas Warnke for providing us with their data from the unusual uses task. We further thank participants at various conferences and seminars for their feedback and comments. The activities of CEBI are financed by the Danish National Research Foundation, Grant DNRF134.

7 Online Appendix - Experimental Instructions

First set of instructions is for Baseline and second set is for objective-ranking treatment.

Consent Form – Personal copy (do not return)

Information Christina Gravert is conducting this research study. You have been asked to participate in this study because you expressed interest in volunteering your time and you are an adult 18 years or older.

Data protection The data from this study will only be used for the purpose of scientific investigations. All the information will be analyzed anonymously and reported only in aggregated form. Personal information will only be used for payment purposes. By signing this document and participating in this study you agree to these terms and conditions.

Risks The study is designed so that participation in it does not involve some added risks or discomforts. Because this is a research study, there may be some unknown risks that are currently unforeseeable.

Benefits Except for the earnings you obtain you will not benefit personally from this study. However, we hope that our results will add to the knowledge about the processes underlying human decision making, and society may benefit from this knowledge.

Earnings The specific earnings for the experiment will depend on your and others' decisions as well as some luck. The exact earnings you may receive for each part of the experiment will be specified clearly before the beginning of each part.

You will be paid either by Swish or in cash.

Participating in an economic experiment is an activity of independent character and does not constitute an employment relationship. Participants are therefore responsible with regard to fiscal consequences. An Income Statement will be sent to Skatteverket if the compensation exceeds 100 SEK. Copy of the Income Statement (KU) will be sent to the payee.

Participation Your participation in this study is voluntary. The researchers may remove you from the study without your consent if they feel it is in your best interest or the best interest of the study. You may also be withdrawn from the study if you do not follow the instructions given you.

Contact information Christina Gravert, Department of Economics, christina.gravert@economics.gu.se. Please keep this email address in case you later have any questions regarding the payment process.

You have received a copy of this consent document. You agree to participate.

Name

Subject's signature Date

Consent Form – Please return a signed version of this copy

Information Christina Gravert is conducting this research study. You have been asked to participate in this study because you expressed interest in volunteering your time and you are an adult 18 years or older.

Data protection The data from this study will only be used for the purpose of scientific investigations. All the information will be analyzed anonymously and reported only in aggregated form. Personal information will only be used for payment purposes. By signing this document and participating in this study you agree to these terms and conditions.

Risks The study is designed so that participation in it does not involve some added risks or discomforts. Because this is a research study, there may be some unknown risks that are currently unforeseeable.

Benefits Except for the earnings you obtain you will not benefit personally from this study. However, we hope that our results will add to the knowledge about the processes underlying human decision making, and society may benefit from this knowledge.

Earnings The specific earnings for the experiment will depend on your and others' decisions as well as some luck. The exact earnings you may receive for each part of the experiment will be specified clearly before the beginning of each part.

You will be paid either by Swish or in cash.

Participating in an economic experiment is an activity of independent character and does not constitute an employment relationship. Participants are therefore responsible with regard to fiscal consequences. An Income Statement will be sent to Skatteverket if the compensation exceeds 100 SEK. Copy of the Income Statement (KU) will be sent to the payee.

Participation Your participation in this study is voluntary. The researchers may remove you from the study without your consent if they feel it is in your best interest or the best interest of the study. You may also be withdrawn from the study if you do not follow the instructions given you.

Contact information Christina Gravert, Department of Economics, christina.gravert@economics.gu.se. Please keep this email address in case you later have any questions regarding the payment process.

You have received a copy of this consent document. You agree to participate.

Name

Subject's signature Date

Group: Player: 1

Thank you for participating in our experiment. We will begin shortly. Today's experiment will last up to 60 minutes.

In this experiment you will have the opportunity to earn some money. Therefore, it is in your interests to pay attention to the instructions and make careful choices. Your earnings will be added to your show up fee of 50 SEK and will be paid out to you after the session.

Anonymity:

Should you agree to participate your name will not be connected to any decision that you make here today or any answer that you provide. All of your actions and information you provide are kept completely anonymous.

Some Rules:

Please switch your cell phone on completely silent (no vibration) and put away anything else that you have brought with you. Please do not talk to other participants during the experiment or attempt to look at the questionnaires of other participants. Do not skip ahead in the papers. Wait for the instructions by the experimenter.

If you have any questions at this point, or at any later point during the experiment, please simply raise your hand. A member of the research team will come to you and answer them in private.

Anyone violating these rules may be excluded from the experiment. In this case he/she will forfeit any earnings.

Structure of the experiment:

The experiment consists of solving a task and of filling out a questionnaire. Before solving the task, there will be a detailed description of the task and a practice round. Please find more instructions on the task and the practice round on the next page.

Please write your group and your player number on top of each page!

Please take your time to read all instructions carefully before making any decisions!

Group: Player: 2

Task description

In the experiment, you will be randomly assigned to a group of five players. You will all do the same task. The group winner of this task will receive 500 SEK (including show-up fee). The other four players receive their show-up fee of 50 SEK. You will never be informed by the research group whom you compete with.

Before starting the real task, you will first do a practice round on the example below. You will not be able to earn money in the practice round, but you can prepare yourself for the real task that will start afterwards.

Explanation of the task for the practice round:
Please list as many different and unusual uses for a **rubber tire** (gummi däck) as you can think of. Do not restrict yourself to a specific size of a tire. You can also list uses that require several tires. Do not restrict yourself to uses you are familiar with, but think of as many new uses as possible!

Please write unusual uses. Using a tire as a "car tire" is not an unusual use, using it as a swing is. Answers are also invalid when they would be impossible to create. Using a tire as a flower pot is possible, using it as a space rocket is not. Try to think of different categories for usage. A "garden seat" and a "swing seat" both fall under the category "tires as seats" while using a tire as a "flower pot" is a different usage category ("container"). You should not name any categories.

In order to win it is in your best interest to write as many, original uses that come from as many different categories as you can think of.

Please quietly raise your hand, if there are any questions.

Do not turn to the next page before instructed, otherwise you will be excluded from the experiment and you will not receive any earnings.

Group: Player:

3

Practice Round

The Task

Please list up as many different and unusual uses for **a rubber tire** as you can think of. Do not restrict yourself to a specific size of a tire. You can also list uses that require several tires. Do not restrict yourself to uses you are familiar with, but think of as many new uses as possible!

You have three minutes for the task.

Please write your ideas below in **CAPITAL LETTERS**!

Group: Player:

4

Official Round

The Task

Please list up as many different and unusual uses for **a sheet of paper** as you can think of. Do not restrict yourself to a specific size of the paper. You can also list uses that require several sheets of paper. Do not restrict yourself to uses you are familiar with, but think of as many new uses as possible!

You have three minutes for the task.

Please write your ideas below in **CAPITAL LETTERS**!

Group: Player: 5

The Scoring Rules

The winner of the 500 SEK will be determined by the following procedure:

You will now evaluate your own answer and the answers of the other four players with whom you compete for the 500 SEK. Please evaluate the answers with respect to their originality. Originality is scored for each person on a scale from 0 to 10 where 0 indicates overall “not at all original answers” and 10 “very original” answers. For scoring, take into consideration i) the number of answers, ii) their degree of being unusual and iii) the number of different categories they come from. The other players in your group will also do the same scoring.
The points given and the points received are kept anonymous by the research team as well as the information who are the players in your group.

For each player, the research team will add up the points given to a TOTAL SCORE of a minimum of 0 and a maximum of 50. The person with the highest TOTAL SCORE out of your group will receive the 500 SEK (including show-up fee). Who has won will be announced by the research team at the end of the experiment. The others receive their show-up fee of 50 SEK. If there is a tie, the winner will be chosen randomly.

You will now receive the five sheets of paper with the answers of each player of your group. Please write your score from 0 to 10 in the BOX above the text field. Please write your player number in the CIRCLE.

Group: Player:

6

Questionnaire

1. **How do you see yourself: Are you generally willing to take risks (risk-prone), or do you try to avoid risks (risk-averse)? Please answer on a scale from 0 to 10 where 0 means risk-averse and 10 means risk-prone.**

Risk-averse 1	2	3	4	5	6	7	8	9	Risk-prone 10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

2. **Do you believe that most people…**

   - would use you if they had the chance…….. ☐

   - or that they would try to be fair to you?... ☐

3. **Would you say that people usually…**

   - try to be helpful... ☐

   - or try to pursue their own interests?......... ☐

4. **Please state how much the following statements describe you…**

	Strongly disagree 1	2	3	4	5	6	Strongly agree 7
- I enjoy working in situations involving competition with others…	☐	☐	☐	☐	☐	☐	☐
- It is important to me to perform better than others on a task…..	☐	☐	☐	☐	☐	☐	☐
- I feel that winning is important in both work and games…………….	☐	☐	☐	☐	☐	☐	☐
- I try harder when I am in competition with other people…………..	☐	☐	☐	☐	☐	☐	☐

Group: Player:

7

5. What kind of personality are you? People can have many different qualities—some are listed below. You will probably find that some of these descriptions fit you completely, some not at all and others may fit to a certain extent. Please answer on a scale from 1 to 7, where 1 means "does not describe me at all", and 7 meaning "describes me perfectly".

I am someone who…	Does not describe me at all **1**	**2**	**3**	**4**	**5**	**6**	Describes perfectly **7**
- does a thorough job ……................................	☐	☐	☐	☐	☐	☐	☐
- is talkative…………………………………………………………	☐	☐	☐	☐	☐	☐	☐
- believes that laws and policies should change to reflect the needs of a changing world………….	☐	☐	☐	☐	☐	☐	☐
- I often tries new things just for trying………………	☐	☐	☐	☐	☐	☐	☐
- is sometimes rude to others……………………………	☐	☐	☐	☐	☐	☐	☐
- is original, comes up with new ideas………………	☐	☐	☐	☐	☐	☐	☐
- worries a lot…………………..…......................	☐	☐	☐	☐	☐	☐	☐
- prefers to spend time in familiar surroundings….	☐	☐	☐	☐	☐	☐	☐
- has a forgiving nature…………………………….........	☐	☐	☐	☐	☐	☐	☐
- tends to be lazy…………………………………………………	☐	☐	☐	☐	☐	☐	☐
- is curious about many different things………………	☐	☐	☐	☐	☐	☐	☐
- outgoing, sociable..	☐	☐	☐	☐	☐	☐	☐
- values artistic, aesthetic experiences…………………	☐	☐	☐	☐	☐	☐	☐
- gets nervous easily………………………..................	☐	☐	☐	☐	☐	☐	☐
- believes that religious authorities should be involved when deciding moral issues…………………	☐	☐	☐	☐	☐	☐	☐
- does things efficiently…………………………………….	☐	☐	☐	☐	☐	☐	☐
- is reserved, quiet…………………………………………………	☐	☐	☐	☐	☐	☐	☐
- is considerate and kind to almost everyone…..	☐	☐	☐	☐	☐	☐	☐
- has an active imagination................................	☐	☐	☐	☐	☐	☐	☐
- is relaxed, handles stress well……………………………	☐	☐	☐	☐	☐	☐	☐
- believes that ideals and principles are more important than open-mindedness…………………….	☐	☐	☐	☐	☐	☐	☐

Group: Player:

8

6. **For each of the following actions or activities, please indicate whether you think that it can always be justified, never be justified, or something in between. You may use any response from 1 to 10 to reflect the strength of your feeling. "1" indicates that it is never justifiable and "10" that it is always justifiable. Make a cross in each of the three rows.**

	Never justifiable 1	2	3	4	5	6	7	8	9	Always justifiable 10
Claiming government benefits to which you are not entitled……………………………	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Avoiding a fare on public transport……….	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cheating on taxes if you have a chance..	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

7. **Do you belong to a religion or religious denomination?**

Yes…………………….. ☐

No …………………….. ☐

Prefer not to answer ……………….. ☐

8. **With which one of the following statements do you agree most?**

The basic meaning of religion is:

To follow religious norms and ceremonies………………… ☐

To do good for other people……………………………… ☐

9. **Assume that an essential commodity of daily use gets broken. How easily would you be able to afford 2,000 SEK to replace the commodity within two weeks without having to borrow the money?**

Very easily	Rather easily	Rather difficult	Not at all able
☐	☐	☐	☐

Group: Player:

10. What is your gender?

Male……………..

Female…………

Prefer not to answer…..……..

Thanks for participating in the questionnaire! Please wait for the next instructions.

Group: Player: 10

Objective Scores

Prior to this experiment, we had more than 100 test persons do the same task as you. With these results, we created a catalogue of possible answers.

We used this catalogue to calculate an OBJECTIVE SCORE (with scores between 0 and 10) for each participant in your group taking into consideration the number of answers, their degree of being unusual and the number of different categories they come from.

Important: These OBJECTIVE SCORES are not used to determine the winner. The winner is solely determined by the TOTAL SCORE (composed of the scores given by one's players in the group and one's own assessment of the task). The TOTAL SCORES will not be revealed except for announcing the winner of each group.

You now have the chance to see the OBJECTIVE SCORES we have calculated. It is for information only.

You do not need to look at the OBJECTIVE SCORES, if you don't want to. But you will have to wait until all people are finished. Your choice has, thus, no influence on the time you stay here for the experiment.

Please circle your answer below. "Yes", if you want to see the OBJECTIVE SCORES and "No", if you do not want to see them.

In the case of "No" you will receive an empty piece of paper.

YES **NO**

Group: Player:

11

Choices

We would like to understand how you made your decisions in the experiment. Please tell us why you chose to allocate the points in the way you did.

Group: Player:

Thank you for participating in our experiment. We will begin shortly. Today's experiment will last up to 60 minutes.

In this experiment you will have the opportunity to earn some money. Therefore, it is in your interests to pay attention to the instructions and make careful choices. Your earnings will be added to your show up fee of 50 SEK and will be paid out to you after the session.

Anonymity:

Should you agree to participate your name will not be connected to any decision that you make here today or any answer that you provide. All of your actions and information you provide are kept completely anonymous.

Some Rules:

Please switch your cell phone on completely silent (no vibration) and put away anything else that you have brought with you. Please do not talk to other participants during the experiment or attempt to look at the questionnaires of other participants. Do not skip ahead in the papers. Wait for the instructions by the experimenter.

If you have any questions at this point, or at any later point during the experiment, please simply raise your hand. A member of the research team will come to you and answer them in private.

Anyone violating these rules may be excluded from the experiment. In this case he/she will forfeit any earnings.

Structure of the experiment:

The experiment consists of solving a task and of filling out a questionnaire. Before solving the task, there will be a detailed description of the task and a practice round. Please find more instructions on the task and the practice round on the next page.

Please write your group and your player number on top of each page!

Please take your time to read all instructions carefully before making any decisions!

Group: Player:

2

Task description

In the experiment, you will be randomly assigned to a group of five players. You will all do the same task. The group winner of this task will receive 500 SEK (including show-up fee). The other four players receive their show-up fee of 50 SEK. You will never be informed by the research group whom you compete with.

Before starting the real task, you will first do a practice round on the example below. You will not be able to earn money in the practice round, but you can prepare yourself for the real task that will start afterwards.

Explanation of the task for the practice round:
Please list as many different and unusual uses for a **rubber tire** (gummi däck) as you can think of. Do not restrict yourself to a specific size of a tire. You can also list uses that require several tires. Do not restrict yourself to uses you are familiar with, but think of as many new uses as possible!

Please write unusual uses. Using a tire as a “car tire” is not an unusual use, using it as a swing is. Answers are also invalid when they would be impossible to create. Using a tire as a flower pot is possible, using it as a space rocket is not. Try to think of different categories for usage. A “garden seat” and a “swing seat” both fall under the category “tires as seats” while using a tire as a “flower pot” is a different usage category (“container”). You should not name any categories.

In order to win it is in your best interest to write as many, original uses that come from as many different categories as you can think of.

Please quietly raise your hand, if there are any questions.

Do not turn to the next page before instructed, otherwise you will be excluded from the experiment and you will not receive any earnings.

Group: Player:

3

Practice Round

The Task

Please list up as many different and unusual uses for **a rubber tire** as you can think of. Do not restrict yourself to a specific size of a tire. You can also list uses that require several tires. Do not restrict yourself to uses you are familiar with, but think of as many new uses as possible!

You have three minutes for the task.

Please write your ideas below in **CAPITAL LETTERS**!

Group: Player:

4

Official Round

The Task

Please list up as many different and unusual uses for **a sheet of paper** as you can think of. Do not restrict yourself to a specific size of the paper. You can also list uses that require several sheets of paper. Do not restrict yourself to uses you are familiar with, but think of as many new uses as possible!

You have three minutes for the task.

Please write your ideas below in **CAPITAL LETTERS**!

Group: Player:

5

The Scoring Rules

The winner of the 500 SEK will be determined by the following procedure:

You will now evaluate your own answer and the answers of the other four players with whom you compete for the 500 SEK. Please evaluate the answers with respect to their originality. Originality is scored for each person on a scale from 0 to 10 where 0 indicates overall "not at all original answers" and 10 "very original" answers. For scoring, take into consideration i) the number of answers, ii) their degree of being unusual and iii) the number of different categories they come from. The other players in your group will also do the same scoring.

The points given and the points received are kept anonymous by the research team as well as the information who are the players in your group.

For each player, the research team will add up the points given to a TOTAL SCORE of a minimum of 0 and a maximum of 50. The person with the highest TOTAL SCORE out of your group will receive the 500 SEK (including show-up fee). Who has won will be announced by the research team at the end of the experiment. The others receive their show-up fee of 50 SEK. If there is a tie, the winner will be chosen randomly.

Prior to this experiment, we had more than 100 test persons do the same task as you. With these results, we created a catalogue of possible answers.

We will use this catalogue to calculate an OBJECTIVE SCORE (with scores between 0 and 10) for each participant in your group taking into consideration the number of answers, their degree of being unusual and the number of different categories they come from.

You will be able to see these OBJECTIVE SCORES for each player in your group at the end of the experiment.

Important: These OBJECTIVE SCORES are not used to determine the winner. The winner is solely determined by the TOTAL SCORE (composed of the scores given by one's players in the group and one's own assessment of the task).

You will now receive the five sheets of paper with the answers of each player of your group. Please write your score from 0 to 10 in the BOX above the text field. Please write your player number in the CIRCLE.

Group: Player:

6

Questionnaire

1. **How do you see yourself: Are you generally willing to take risks (risk-prone), or do you try to avoid risks (risk-averse)? Please answer on a scale from 0 to 10 where 0 means risk-averse and 10 means risk-prone.**

Risk-averse									Risk-prone
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

2. **Do you believe that most people…**

- would use you if they had the chance…….. ☐
- or that they would try to be fair to you?... ☐

3. **Would you say that people usually…**

- try to be helpful... ☐
- or try to pursue their own interests?……… ☐

4. **Please state how much the following statements describe you…**

	Strongly disagree **1**	**2**	**3**	**4**	**5**	**6**	Strongly agree **7**
- I enjoy working in situations involving competition with others…	☐	☐	☐	☐	☐	☐	☐
- It is important to me to perform better than others on a task…..	☐	☐	☐	☐	☐	☐	☐
- I feel that winning is important in both work and games…………….	☐	☐	☐	☐	☐	☐	☐
- I try harder when I am in competition with other people…………..	☐	☐	☐	☐	☐	☐	☐

Group: Player:

7

5. **What kind of personality are you? People can have many different qualities—some are listed below. You will probably find that some of these descriptions fit you completely, some not at all and others may fit to a certain extent. Please answer on a scale from 1 to 7, where 1 means "does not describe me at all", and 7 meaning "describes me perfectly".**

I am someone who…	Does not describe me at all **1**	**2**	**3**	**4**	**5**	**6**	Describes perfectly **7**
- does a thorough job	☐	☐	☐	☐	☐	☐	☐
- is talkative..	☐	☐	☐	☐	☐	☐	☐
- believes that laws and policies should change to reflect the needs of a changing world..............	☐	☐	☐	☐	☐	☐	☐
- I often tries new things just for trying..................	☐	☐	☐	☐	☐	☐	☐
- is sometimes rude to others..................................	☐	☐	☐	☐	☐	☐	☐
- is original, comes up with new ideas....................	☐	☐	☐	☐	☐	☐	☐
- worries a lot..	☐	☐	☐	☐	☐	☐	☐
- prefers to spend time in familiar surroundings....	☐	☐	☐	☐	☐	☐	☐
- has a forgiving nature...	☐	☐	☐	☐	☐	☐	☐
- tends to be lazy...	☐	☐	☐	☐	☐	☐	☐
- is curious about many different things.................	☐	☐	☐	☐	☐	☐	☐
- outgoing, sociable..	☐	☐	☐	☐	☐	☐	☐
- values artistic, aesthetic experiences....................	☐	☐	☐	☐	☐	☐	☐
- gets nervous easily...	☐	☐	☐	☐	☐	☐	☐
- believes that religious authorities should be involved when deciding moral issues.....................	☐	☐	☐	☐	☐	☐	☐
- does things efficiently...	☐	☐	☐	☐	☐	☐	☐
- is reserved, quiet..	☐	☐	☐	☐	☐	☐	☐
- is considerate and kind to almost everyone.....	☐	☐	☐	☐	☐	☐	☐
- has an active imagination................................	☐	☐	☐	☐	☐	☐	☐
- is relaxed, handles stress well................................	☐	☐	☐	☐	☐	☐	☐
- believes that ideals and principles are more important than open-mindedness.........................	☐	☐	☐	☐	☐	☐	☐

Group: Player: 8

6. **For each of the following actions or activities, please indicate whether you think that it can always be justified, never be justified, or something in between. You may use any response from 1 to 10 to reflect the strength of your feeling. "1" indicates that it is never justifiable and "10" that it is always justifiable. Make a cross in each of the three rows.**

	Never justifiable 1	2	3	4	5	6	7	8	9	Always justifiable 10
Claiming government benefits to which you are not entitled……………………………	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Avoiding a fare on public transport……….	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cheating on taxes if you have a chance..	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

7. **Do you belong to a religion or religious denomination?**

Yes…………………….. ☐

No …………………….. ☐

Prefer not to answer ……………….. ☐

8. **With which one of the following statements do you agree most?**

The basic meaning of religion is:

To follow religious norms and ceremonies………………… ☐

To do good for other people…………………………….. ☐

9. **Assume that an essential commodity of daily use gets broken. How easily would you be able to afford 2,000 SEK to replace the commodity within two weeks without having to borrow the money?**

Very easily	Rather easily	Rather difficult	Not at all able
☐	☐	☐	☐

Group: Player:

9

10. What is your gender?

Male…………….. ☐

Female………… ☐

Prefer not to answer…......... ☐

Thanks for participating in the questionnaire! Please wait for the next instructions.

Group: Player: 10

Objective Scores

Prior to this experiment, we had more than 100 test persons do the same task as you. With these results, we created a catalogue of possible answers.

We used this catalogue to calculate an OBJECTIVE SCORE (with scores between 0 and 10) for each participant in your group taking into consideration the number of answers, their degree of being unusual and the number of different categories they come from.

Important: These OBJECTIVE SCORES are not used to determine the winner. The winner is solely determined by the TOTAL SCORE (composed of the scores given by one's players in the group and one's own assessment of the task). The TOTAL SCORES will not be revealed except for announcing the winner of each group.

You now have the chance to see the OBJECTIVE SCORES we have calculated. It is for information only.

You do not need to look at the OBJECTIVE SCORES, if you don't want to. But you will have to wait until all people are finished. Your choice has, thus, no influence on the time you stay here for the experiment.

Please circle your answer below. "Yes", if you want to see the OBJECTIVE SCORES and "No", if you do not want to see them.

In the case of "No" you will receive an empty piece of paper.

YES **NO**

Group: Player:

11

Choices

We would like to understand how you made your decisions in the experiment. Please tell us why you chose to allocate the points in the way you did.