First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision *

David Gill †
Zdenka Kissová ‡
Jaesun Lee §
Victoria Prowse ¶

This version: January 30, 2018
First version: July 22, 2015

Abstract

Rank-order relative-performance evaluation, in which pay, promotion, symbolic awards and educational achievement depend on the rank of individuals in the distribution of performance, is ubiquitous. Whenever organizations use rank-order relative-performance evaluation, people receive feedback about their rank. Using a real-effort experiment, we aim to discover whether people respond to the specific rank that they achieve. In particular, we leverage random variation in the allocation of rank among subjects who exerted the same effort to obtain a causal estimate of the *rank response function* that describes how effort provision responds to the content of rank-order feedback. We find that the rank response function is U-shaped. Subjects exhibit 'first-place loving' and 'last-place loathing', that is subjects work hardest after being ranked first or last. We discuss implications of our findings for the optimal design of performance feedback policies, workplace organizational structures and incentives schemes.

Keywords: Relative performance evaluation; Relative performance feedback; Rank order feedback; Dynamic effort provision; Real effort experiment; Flat wage; Fixed wage; Taste for rank; Status seeking; Social esteem; Self esteem; Public feedback; Private feedback.

JEL Classification: C23; C91; J22; M12.

*This manuscript was accepted for publication in *Management Science*. The version of record can be found at https://doi.org/10.1287/mnsc.2017.2907. We thank the Nuffield Centre for Experimental Social Sciences at the University of Oxford for hosting our experiment and Jozef Dobos for excellent programming assistance. We also thank the Editor, the Associate Editor, the anonymous referees and numerous seminar participants for their helpful comments. Finally, we thank the Economic and Social Research Council, the British Academy and the University of Southampton for funding the experiment.

†Department of Economics, Purdue University; gill53@purdue.edu
‡OC&C Strategy Consultants; zdenka.dobosova@occstrategy.com
§Department of Economics and Finance, Tongji University; jaesunlee@tongji.edu.cn
¶Department of Economics, Purdue University; vprowse@purdue.edu

1 Introduction

Relative-performance evaluation (RPE) is ubiquitous. Bonuses, promotions, performance appraisals, symbolic awards, public sector 'league tables' and educational achievement often depend on how well individuals and institutions perform relative to others (Gibbons and Murphy, 1990; Prendergast, 1999; Propper and Wilson, 2003; Kosfeld and Neckermann, 2011), while some companies base executive compensation on across-firm performance comparisons (Gibbons and Murphy, 1990).

Ranking of performance is particularly popular in the business world, in the public sector, and in education. Rank-order RPE, in which pay, promotion, employee appraisals and non-pecuniary awards depend on the rank of individuals in the distribution of performance, is common. Since supervisors need information only on rank and not on absolute performance, rank-order RPE is simple to implement (Prendergast, 1999). Rank-order RPE also prevents ratings compression, that is the tendency of supervisors to rate good and bad workers as too similar to each other (Moers, 2005). Furthermore, competition for promotions naturally takes the form of rank-order RPE when the number of more senior positions that people are competing for is fixed.[1] Prendergast (1999)'s survey concludes that firms primarily provide incentives using competitions for promotion rather than within-grade variation in pay according to performance, while Baker et al. (1988) note that promotion tournaments not only motivate workers but also help to sort workers into jobs according to ability. Hazels and Sasse (2008) provide evidence of the growing popularity of 'forced ranking' RPE, such as General Electric's 'vitality curve' under which each supervisor has to identify the top 20% and bottom 10% of performers. Symbolic awards are also often allocated according to rank: award schemes include McDonald's "Employee of the Month", IBM's "Bravo Award" and Inuit's "Spotlight" employee recognition scheme (Kosfeld and Neckermann, 2011). Finally, rank-order RPE is common beyond the confines of the traditional workplace: the outcomes of sports contests, examinations, innovation races and elections are often determined by rank in performance, while Frey (2007) provides evidence of the broader popularity of rank-based award schemes such as orders, medals, decorations and prizes.

In this paper, we aim to identify how individuals respond to the specific rank that they achieve: that is, does the *content* of rank-order feedback influence effort provision?[2] This question matters because whenever organizations use rank-order RPE, people receive feedback about their specific rank in the distribution of performance. As we discuss in detail below, it is therefore important that organizations understand how people respond to the content of rank-order feedback so that they can design effective performance feedback policies, workplace organizational structures and incentives schemes that take into account the implicit incentives generated by individuals' preferences over the specific rank that they achieve. Understanding how people

[1] Rank-order RPE also has the advantage that it filters out from compensation the effects of shocks that are common across workers (Lazear and Rosen, 1981) and, in the presence of incentives to underreport performance with subjective performance evaluation, it allows employers to commit to the total amount of compensation (Malcomson, 1984). These last two features are shared by other forms of RPE. However, in the presence of workers that care about how their pay compares to how much they feel they deserve, Gill and Stone (2010) show that rank-order RPE can dominate other forms of RPE that are continuous in relative-performance differences.

[2] As is common in the experimental literature on real-effort provision, we use the term 'effort' to correspond to measurable performance in a work task rather than the cost associated with work effort (see, e.g., Abeler et al., 2011, Gill and Prowse, 2012, and Charness et al., 2014).

respond to the content of rank-order feedback will also help policy-makers to formulate rules and regulations pertaining to incentive schemes, such as bonus pay, that involve rank-order relative-performance feedback. The effective design of transparency policies, such as public disclosure of official hospital, school and university league tables in the United Kingdom (Propper and Wilson, 2003), or public disclosure of income tax records in Scandinavian countries (Bø et al., 2015), also hinges on how people respond to the content of rank-order feedback.

We want to identify the effects of a pure taste or preference for rank in the distribution of performance uncontaminated by any preference over rank in the distribution of earnings or a desire for the longer-term reputational benefits associated with higher rank. To do so, we designed a laboratory experiment in which our subjects repeatedly exerted real effort and were paid using a single flat wage that did not depend on performance. The flat wage ensures that our subjects were not motivated to work hard to earn money or improve their rank in the distribution of earnings.[3] Our laboratory setting with random selection of subjects to be invited to participate in each session from a large laboratory-maintained subject pool further ensures that our subjects were not motivated by longer-term reputational concerns. In the Baseline sessions, no rank-order feedback was provided to the subjects. In the Treatment sessions, at the end of every round each subject was informed about her rank in that round among the 17 participants in the session. Furthermore, subjects were always informed of their own absolute performance, but were never told the absolute or mean performance of the other participants.

Our interest lies in recovering the *rank response function* that describes how effort provision in the current round responds to the content of the rank-order feedback received at the end of the previous round. This is challenging because serial dependence in the unobserved drivers of effort will give rise to non-causal correlation between rank in round r and effort in round $r+1$.[4] Such serial dependence in unobservables can take innumerable forms. Possible causes of serial dependence include permanent between-subject differences in unobserved ability, across-subject heterogeneity in rates of learning over rounds, subjects who work in spurts and regression to the mean.

To side-step these potentially serious confounds, we propose and apply an econometric approach that provides a causal estimate of the rank response function. In particular, we use randomness in the allocation of rank in order to identify cleanly the causal effect of the content of rank-order feedback on subsequent effort provision. We do this by using random variation in rank among subjects that exerted the same effort in a given round. By using large sessions of 17 subjects, our design creates many ties in effort. As was made clear to the subjects at the beginning of the experiment, ties were broken randomly. By breaking ties at random when allocating rank, we create random variation that we use in our econometric analysis. It is important that we have enough ties to be able to estimate with precision the effect of the content of rank-order

[3] Of course, in many organizations remuneration and promotion are tied to performance. We use a flat wage as a device to cleanly separate different motives for behavior. Nonetheless, we note that the use of flat wages by employers remains surprisingly common (see Charness et al., 2014, pp. 39-40, for a review of the evidence). As noted by Holmstrom and Milgrom (1991): "It remains a puzzle for this theory that employment contracts so often specify fixed wages and more generally that incentives within firms appear to be so muted, especially compared to those of the market." Furthermore, organizations often provide relative-performance feedback even when pay does not depend on relative performance (Charness et al., 2014, p. 40, also review this evidence).

[4] Throughout, we think of rank as ordered from its lowest value of 17 to its highest value of 1. Thus, we say that rank increases when rank changes from a higher number to a lower number.

feedback on effort provision, while at the same time ensuring that ties are not so common that rank becomes uninformative: in our dataset, 18% of observations involve ties within a given session and round, which we feel strikes an appropriate balance. The laboratory offers us the degree of control necessary to create and observe the random variation that is essential to our identification strategy: the extensive empirical literature on status-seeking suggests that people care about rank without being able to identify the causal effect of rank feedback on later performance (see e.g., the survey by Heffetz and Frank, 2011, as well as the literature that we cite below).

We find that subjects respond strongly to the specific rank that they achieve. In particular, we find that the rank response function is U-shaped. Subjects increase their effort the most in response to the content of rank-order feedback when they are ranked first or last: we call this motivating effect of high and low rank 'first-place loving' and 'last-place loathing'. Being ranked first increases effort by 21% relative to the average level of effort in the Treatment group that receives rank-order feedback, while being ranked last increases effort by 13%. By contrast, being ranked in the middle of the pack, that is being ranked 9th or 10th, reduces effort by more than 10% relative to the average level of effort in the Treatment group (although the effort of the subjects ranked 9th or 10th is still higher than the average level of effort in the Baseline group that does not receive any rank-order feedback). This U-shaped rank response function can be explained by a combination of pride or 'joy of winning' from achieving high rank together with an aversion to low rank. We also find that the U-shaped rank response function does not vary by gender, country of birth, age or subject of study, suggesting that the phenomena of first-place loving and last-place loathing are not restricted to specific demographic groups, but instead are more universal in their manifestation.

Our finding of a U-shaped response to the content of rank-order feedback has a number of implications for how organizations might choose to design their performance feedback policies. In particular, it might be profitable for organizations to emphasize feedback of very high or very low relative performance, e.g., by awarding symbolic prizes to the best performers or scheduling regular appraisal meetings with senior managers for the worst performers. On the other hand, employers might want to exercise caution when providing relative-performance feedback to avoid demoralizing workers of intermediate ability. This concern will be of particular importance in settings in which middle ranking workers are the most loyal, perhaps because they are the least likely to be fired or poached, or in settings where teamwork and cooperation between workers of different abilities are important to production. Our finding that the effects of rank-order feedback do not depend on demographics such as gender suggests that organizations need not worry about designing different feedback policies for groups of people that vary in their demographic characteristics. The U-shaped pattern of response also has implications for optimal organizational design. For instance, firms might want to divide workers into small comparison groups, e.g., by adopting a decentralized organizational structure or designing highly specialized jobs, in order to reduce the number of middle ranks that solicit relatively low subsequent effort provision. Organizations might also find it productive to match individuals into groups with similar abilities, so that everybody has a realistic prospect of obtaining top ranks. Finally, the U-shaped pattern of response also has implications for the optimal design of incentive schemes. Organizations should be aware that incentive schemes that involve rank-order relative-performance evaluation are likely to generate implicit incentives via responses to rank-order feedback in addition to the more obvious pecuniary incentives that standard economic theory emphasizes. For example, the fact that people strive to maintain high rank generates implicit incentives for higher performers, which suggests that marginal pecuniary incentives should be focused more towards middle performers than standard economic theory would suggest.

The novelty of our analysis lies in estimating the causal effect of the content of rank-order feedback on subsequent effort provision. As described above, by leveraging random variation in the allocation of rank, we purge completely the confounding effects of serially dependent unobservables. Furthermore, we show that standard random and fixed effects panel data estimators applied to our data give estimates of the rank response function that differ markedly from our causal estimate, illustrating that our approach is critical to obtaining reliable results on how effort responds to the content of rank-order feedback. Charness et al. (2014) focus on the interaction between relative-performance feedback and sabotage and cheating; in an ancillary analysis they use a standard random effects panel data estimator to calculate correlations between a subject's rank (first, second, or third) in a given round and the change in the subject's effort between that round and the next.[5] Unlike us, Barankay (2011), Barankay (2012) and Kuhnen and Tymula (2012) do not study whether the specific rank that a subject achieves influences effort; however, they do regress performance on a dummy variable that captures whether rank feedback was worse than expected.

There is also an empirical literature on the impact of interim rank information during the course of a competition for prizes, which provides information about the within-competition pecuniary return to effort (e.g., Ehrenberg and Bognanno, 1990, Fershtman and Gneezy, 2011, Genakos and Pagliero, 2012, and Delfgaauw et al., 2013).[6] Relatedly, a small literature looks at the impact of the outcome of competition for monetary prizes on later effort (Gill and Prowse, 2014; Legge and Schmid, 2015): these papers do not identify the pure effect of rank since: (i) they confound rank and monetary prizes; and (ii) competitive outcomes generally provide information about relative ability and hence about the pecuniary return to effort in later competitions.[7]

We study the effects of rank in performance determined by the real work effort of experimental subjects on subsequent effort provision; in contrast Clark et al. (2010) and Kuziemko et al. (2014) study experimentally the effect of rank in the distribution of money on monetary gift-exchange reciprocity (Clark et al., 2010), willingness to take gambles (Kuziemko et al., 2014) and willingness to give to others (Kuziemko et al., 2014). Importantly, in these papers rank has no relationship to work effort or individual merit. Instead, experimental subjects are endowed with an initial rank in the distribution of money that is uncorrelated with their behavior or

[5] In settings different to ours, Hannan et al. (2008), Freeman and Gelber (2010) and Bradler et al. (forthcoming) use methods similar to Charness et al. (2014) to calculate correlations between a subset of ranks (Freeman and Gelber, 2010; Bradler et al., forthcoming) or performance deciles (Hannan et al., 2008) and subsequent performance. In Hannan et al. (2008) and Freeman and Gelber (2010), the feedback was about performance on a task with piece-rate pay and thus was informative about relative earnings. In Freeman and Gelber (2010) and Bradler et al. (forthcoming), feedback was unannounced and provided a single time. Finally, using observational data on school children, Murphy and Weinhardt (2014) and Elsner and Isphording (2015) find a correlation between rank in school and later educational achievement.

[6] Performance feedback also underlies the hypothesis that psychological momentum can help explain 'hot hand' streaks in sports (Iso-Ahola and Dotson, 2014).

[7] In a setting in which subjects compete for monetary prizes by investing money rather than exerting effort, Dutcher et al. (2015) consider the impact of receiving the highest 'winner' prize or the lowest 'loser' prize on later investment.

characteristics.[8]

As well as discovering how people respond to the content of rank-order feedback, we also validate our design by replicating the finding that with flat wages performance increases substantially *on average* when subjects are given rank-order feedback: we find that effort is about 20% higher in the Treatment group that receives feedback compared to effort in the Baseline group without feedback. Falk and Ichino (2006), Kuhnen and Tymula (2012), Hannan et al. (2013), Cadsby et al. (2014) and Charness et al. (2014) also find that in a real-effort task with flat wages performance increases substantially on average when subjects know that full rank-order feedback will be available; as far as we are aware only Eriksson et al. (2015) fail to find an effect.[9] Our work also adds to the substantial body of evidence that supports the importance of status-seeking behavior in a variety of contexts (e.g., Festinger, 1954, Frank, 1985, Huberman et al., 2004, Ellingsen and Johannesson, 2007, and Heffetz and Frank, 2011).[10]

Finally, we extend the literature on how the mode of rank-order feedback influences performance. Across sessions we varied whether feedback was provided privately via the subjects' computer terminals or publicly in front of all the subjects in the session, and we find no statistically significant differences in how people respond to the content of rank-order feedback according to whether the feedback was provided publicly or privately. The existing literature only considers whether public and private rank-order feedback have different effects on average performance: as described in Supplementary Web Appendix C, we find no differences in the average level of effort by mode of feedback, and perhaps surprisingly this null result is not out of line with the existing evidence.[11]

The paper proceeds as follows: Section 2 describes the experimental design; Section 3 con-

[8]In Clark et al. (2010), an 'employer' offers money (interpreted as a flat wage) to an 'employee', who can reciprocate by spending money to the benefit of the employer. The employer offers the wage blindly, knowing nothing about the characteristics or past behavior of the employee. The higher the rank in the distribution of offers by a set of employers to their respective employees, the more reciprocal the employee. Kuziemko et al. (2014) find that people are more willing to gamble and less willing to give to less-fortunate others when they are ranked at, or near the bottom, in the distribution of money in a setting in which endowments of money are allocated randomly by the experimenter.

[9]In other settings (e.g., with performance pay, reputational effects, or where comparisons only to average performance were provided), the evidence on the impact of relative-performance feedback on average performance is mixed. A number of papers find that people work harder or perform better with relative-performance feedback (Gneezy and Rustichini, 2004; Hannan et al., 2008; Mas and Moretti, 2009; Azmat and Iriberri, 2010; Freeman and Gelber, 2010; Murthy, 2010; Blanes i Vidal and Nossol, 2011; Kosfeld and Neckermann, 2011; Murthy and Schafer, 2011; Tran and Zeckhauser, 2012; Tafkov, 2013; Gerhards and Siemer, 2014; Lount Jr. and Wilk, 2014; Jalava et al., 2015; Azmat and Iriberri, 2016; Bradler et al., forthcoming). However, some papers find no effect (Azmat and Iriberri, 2016, when subjects were paid a fixed wage and comparisons only to average performance were provided), report lower performance (Bellemare et al., 2010; Barankay, 2011; Barankay, 2012; Bandiera et al., 2013; Ashraf et al., 2014) or find no clear pattern (Gino and Staats, 2011; Bhattacharya and Dugar, 2012; Rosaz et al., 2012; Georganas et al., 2015), while others find a negative impact on other dimensions (Eriksson et al., 2009; Ebeling et al., 2012; Hannan et al., 2013).

[10]Status seeking may be underpinned by a competitive desire for dominance (Rustichini, 2008) or a 'joy of winning' (Coffey and Maloney, 2010; Sheremeta, 2010), and evidence from neuroeconomics shows that outperforming others activates brain areas related to reward processing (Dohmen et al., 2011). Furthermore, recent happiness research links well-being to the ordinal rank of an individual's wage or income within a comparison group (Brown et al., 2008; Clark et al., 2009; Boyce et al., 2010), while Tincani (2015) finds that responses to changes in the peer ability distribution in the classroom caused by an earthquake are consistent with a model in which students prefer higher rank.

[11]Tran and Zeckhauser (2012), Ashraf et al. (2014), Cadsby et al. (2014) and Gerhards and Siemer (2014) find little or no difference between performance under public and private feedback, while Tafkov (2013) and Hannan et al. (2013) find higher performance with public feedback, although Hannan et al. (2013) also find that public feedback increased inefficient time allocation. Eriksson et al. (2015) do not compare public and private feedback, but they do find that subjects ranked lowest are willing to pay to avoid public exposure.

siders how the content of rank-order feedback influences effort provision; Section 4 concludes; and the Supplementary Web Appendix provides the experimental instructions and additional analysis.

2 Experimental design

2.1 Procedures

We ran 18 experimental sessions at the Nuffield Centre for Experimental Social Sciences (CESS) at the University of Oxford. Each session included 17 student subjects (who did not report Psychology as their main subject of study) and lasted approximately 90 minutes. The 306 participants were drawn from the CESS subject pool, which is managed using the Online Recruitment System for Economic Experiments (Greiner, 2015). For each session, invited students were randomly drawn from the CESS subject pool. Seating positions were randomly assigned. The experimental instructions (Supplementary Web Appendix A) were provided to each subject in written form and were read aloud to the subjects. Questions were answered privately. Subjects were paid in cash at the end of the session.

Each session consisted of a practice round followed by 6 paying rounds, with a demographic questionnaire at the end. We do not use the data from the practice round in the analysis in Section 3. In each round, subjects worked on two real-effort tasks. First, they worked on a computerized verbal task for 3 minutes; second, they worked on a computerized numerical task for a further period of 3 minutes; finally, the subjects were given a 4-minute break. In the paying rounds, the subjects were given treatment-specific feedback during this break. Details of the tasks and feedback follow below. The subjects had no access to the Internet and we did not provide them with any leisure activities. The subjects were paid a show-up fee of £5 and were paid £2.50 in each paying round independently of their performance in the tasks, giving a total payment of £20 per subject. All payments were in pounds sterling. The number of rounds, the real-effort tasks, the nature of the treatment-specific feedback in the paying rounds, and the payment scheme were described in detail before the start of the practice round. We framed the payment scheme as a fixed payment of £2.50 per round to replicate as closely as possible real-world work environments with a fixed wage per work period. After the practice round and before the start of the first paying round, the subjects were reminded that they would be paid £2.50 in each round and we carefully explained that the payment would not depend on their performance in the tasks. They were also reminded about the nature of the treatment-specific feedback that they would receive at the end of each round.

2.2 Real-effort tasks

The verbal task was a 'word-spotting' task. Subjects were presented with a 15 $\times$ 15 grid of capital letters and scored one point for each valid English word that they correctly spotted.[12] In the numerical task, subjects added up pairs of 2-digit numbers and scored one point for each pair that they correctly added up.[13] In both cases, subjects were not penalized for incorrect answers. During each task, a banner at the top of the screen displayed the round number, the time remaining and the subject's score so far in the task. The subjects were told that, in any given round, all the subjects in the session would be presented with the same grid of letters and sequence of pairs of numbers. The round-specific grid of letters and sequence of pairs of numbers were also kept constant across all 18 sessions to ensure that difficulty did not vary by treatment. The grids of letters and sequences of pairs of numbers were chosen to avoid systematic variation in difficulty across rounds.

We chose to use two real-effort tasks, one based on mathematical ability and one based on verbal ability, to ensure that rank reflected performance across multiple skills and so limit the extent to which rank could be driven by ability with respect to just one skill. Performance on the numeral task explains 16.3% of the variation in performance on the verbal task (the correlation between points scored on the verbal task and points scored on the numerical task was 0.403).

We ran a non-incentivized calibration pilot to ensure that the two tasks were of approximately equal difficulty. On average in the incentivized experiment, across the practice round and 6 paying rounds, the subjects scored 41.4 in the verbal task and 39.1 in the numerical task. A subject's 'total points score' in a round is the sum of her scores in the verbal and numerical tasks. We use the term 'effort' to mean 'total points score': as is common in real-effort experiments, our use of the term 'effort' therefore corresponds to measurable performance in a work task rather than the cost associated with work effort (see, e.g., Abeler et al., 2011, Gill and Prowse, 2012, and Charness et al., 2014).

2.3 Treatment-specific feedback

As noted above, at the end of each paying round the subjects were given a 4-minute break during which treatment-specific feedback was provided, and the subjects were reminded about the nature of this feedback just before the start of the first paying round. In every paying round, the subject's total points score was first displayed on her screen for 1 minute.[14] In the Baseline,

[12] Grid-based word-spotting and word-search tasks have been used by, e.g., Burrows and Loomes (1994). The specific implementation here was custom-designed for our experiment using Java. To propose a word, subjects used their mouse to move the cursor to the first letter of the word and then hold down the mouse button to drag the cursor to the last letter of the word. If the proposed word was valid, it was then highlighted in yellow. Valid words could appear horizontally, vertically, diagonally, forwards or backwards, and valid words could share letters. British and American spellings were valid, as were singular and plural forms. Proper nouns and abbreviations were not valid. Valid words were taken to be the words appearing in the "WordNet" (Miller, 1995; wordnet.princeton.edu) or "iSpell" (lasr.cs.ucla.edu/geoff/ispell.html) databases of English words.

[13] Tasks based on adding numbers have been used by, e.g., Niederle and Vesterlund (2007). The specific implementation here was custom-designed for our experiment using Java. Subjects proposed answers by using their mouse to click on a 9-digit virtual number-pad that appeared on the screen and then click on a 'submit' button. The subjects were not allowed to use paper or pens. If the proposed answer was correct, the subject moved on to a new pair of numbers to add up. If the answer was incorrect, the subject had to try again until she summed the pair of numbers correctly.

[14] The subject's point score was also displayed in the practice round.

the subjects then waited for 3 minutes. In the Treatment, during the next 3 minutes of the break each subject was informed about her rank in that round among the 17 participants in the session. As was made clear to the subjects at the beginning of the experiment, ties were broken randomly.[15] Depending on the sub-treatment, this rank-order feedback was provided in one of three different ways. In Sub-treatment 1, the subject's rank was displayed on her screen. In Sub-treatment 2, the experimenter personally and privately informed each subject about her rank. The experimenter did this by handing each subject a card indicating the subject's rank and pointing to where the subject's rank was written on the card. In Sub-treatment 3, all participants were asked to stand up and the experimenter publicly informed each subject about her rank.[16,17]

The experiment used a between-subject design. There were 51 subjects in the Baseline group (3 sessions), 102 in Sub-treatment 1 (6 sessions), 68 in Sub-treatment 2 (4 sessions) and 85 in Sub-treatment 3 (5 sessions). We collected more data in the Treatment than in the Baseline in order to have sufficient power to study how effort provision responds to the content of previous rank-order feedback.[18]

2.4 Questionnaire

At the end of the session, the subjects completed a questionnaire on demographics and competitiveness. The subjects were asked their gender, age, country of birth and main field of study. To preserve anonymity, we collapsed the answers into two categories: male vs. female; aged 22 or above vs. below 22; United Kingdom vs. other country; and Sciences, Technology, Engineering, Maths and Medicine (STEMM) vs. Humanities, Social Sciences, Business, Law and Education. The subjects were also asked to self-report their degree of competitiveness.

3 How effort responds to the content of rank-order feedback

As shown in Figure 1, we find that providing rank-order feedback increases effort substantially on average. Supplementary Web Appendix C provides statistical tests and more detail. We show there that effort does not vary significantly by sub-treatment. The fact that we find that our subjects respond strongly to private rank-order feedback suggests that the subjects care about their own (first-order) beliefs about their rank in the distribution of performance; that is subjects have a desire for 'self esteem'. The fact that we find no difference in the degree to

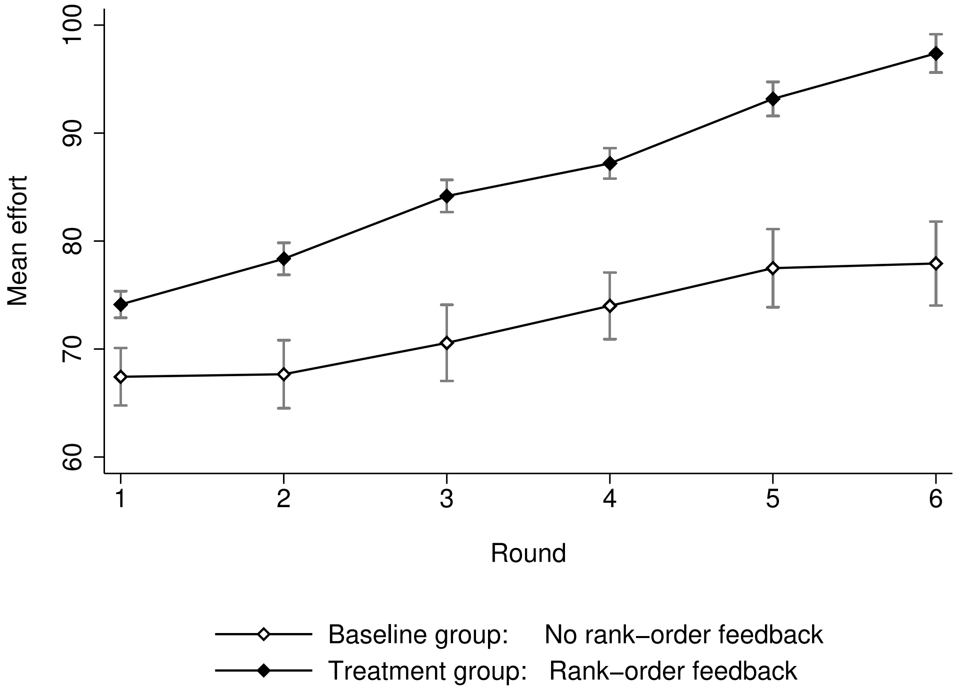
[15] The subjects were told that: "The participant with the highest total points score in that round will be ranked first, the participant with the second highest points score will be ranked second, and so forth. Any ties will be broken at random."

[16] We included these three sub-treatments to test whether mean performance with rank-order feedback depends on the degree to which rank is made public. As we move from Sub-treatment 1 to Sub-treatment 3, the feedback becomes more and more public.

[17] Conceivably, the absence of an outside option while the subjects were working could amplify treatment effects due to experimenter demand effects. However, if such demand effects were important we would expect performance to respond more strongly to rank feedback provided directly by the experimenter in Sub-treatments 2 and 3, but as outlined later we do not find such an effect. Furthermore, in a real-effort laboratory experiment with and without an outside option (browsing the Internet), perhaps surprisingly Goerg et al. (2017) find that subjects respond less strongly to the size of incentives when the outside option is not available (unsurprisingly, the baseline level of performance is higher without the outside option).

[18] We attempted to collect 21 sessions of data, but due to technical difficulties we were able to collect only 18 sessions.

which public and private rank-order feedback motivate our subjects suggests that the subjects do not care about their (second-order) beliefs about the beliefs of others about their rank in the distribution of performance; that is subjects have little desire for 'social esteem'. Our laboratory setting with random selection of subjects ensures that any taste for social esteem is independent of longer-term reputational concerns.[19]



Note: Vertical bars show ± one standard error of the estimate of the round-by-round mean of effort.

Figure 1: Round-by-round mean effort.

The increase in average effort with rank-order feedback is consistent with several different patterns of response to the content of rank-order feedback. Even though subjects increase effort on average when rank-order feedback is provided, they might completely ignore the content of the feedback. Alternatively, effort might respond linearly to feedback about rank order in the distribution of performance: as rank increases, the subjects might become more and more motivated; instead, effort could decline linearly in rank if higher rank demotivates the subjects. More complex scenarios are also possible: for instance, high or low rank might increase motivation relative to intermediate rank-order feedback.

[19]As noted in Section 2.4, at the end of the experiment we asked subjects to self-report their degree of competitiveness. In Supplementary Web Appendix D we describe the question on competitiveness and consider the interaction between competitiveness and rank-order feedback. In summary, with or without rank-order feedback, strongly competitive subjects are motivated to work much harder on average, while the extra motivation induced by rank-order feedback is the same whether subjects are strongly competitive or not. To understand what might underlie this result, note that subjects might form beliefs about their rank in the distribution of performance even in the absence of rank-order feedback, and the subjects might care about these beliefs. Furthermore, the questionnaire equates being 'strongly competitive' with always being interested in how one's performance compares to that of others. Thus, one possible explanation of our findings is that strongly competitive subjects are more highly motivated to work hard to improve their beliefs about their rank in the distribution of performance even when they do not receive any feedback, while the extra motivation induced by providing the rank-order feedback does not vary according to subject competitiveness.

In this section, we seek to shed light on exactly how people respond to the specific rank that they achieve. To this end, we propose and apply an econometric approach that provides a causal estimate of the effect of the content of rank-order feedback on subsequent effort provision. Section 3.1 introduces the estimation problem. Section 3.2 outlines our identification strategy, which exploits randomness in the allocation of rank, and Section 3.3 describes the estimation sample. Section 3.4 presents our results.

3.1 Explaining the rank response function

To fix ideas, suppose that in each of S sessions, N subjects complete a real-effort task for R rounds, with $N \times S$ subjects in total. At the end of every round, each subject is informed of the rank of her effort among the efforts of the N session members. In our experiment, the relevant subjects are those in the Treatment group who receive rank-order feedback, and so $S = 15$, N=17 and $R = 6$. As noted in footnote 4, we think of rank as ordered from its lowest value of N to its highest value of 1. Any within-round ties in the efforts of the session members are broken at random. Thus, for each session and round combination, one and only one subject receives each possible rank. The effort provision of subject n in round $r \geq 2$ of session s is given by:

$$\text{Effort}_{n,s,r} = H(\text{Rank}_{n,s,r-1}) + \gamma_r + \beta X_{n,s} + \varepsilon_{n,s,r} \quad \text{for} \quad n = 1, ..., N; \; s = 1, ..., S; \; r = 2, ..., R. \quad (1)$$

In the above: $\text{Rank}_{n,s,r-1} \in \{1, ..., N\}$ denotes the subject's rank in the previous round; γ_r for $r = 2, ..., R$ are round fixed effects, which capture all round-specific effort shifters that are common across subjects, including common learning effects; $X_{n,s}$ denotes observed subject-specific characteristics that impact effort provision; and $\varepsilon_{n,s,r}$ denotes all unobserved or unmodeled effort shifters, such as ability, motivation or ranks in rounds prior to the previous round.

Our interest lies in recovering the *rank response function*, $H(\text{Rank}_{n,s,r-1})$, which describes how effort provision in the current round responds to the content of the rank-order feedback received at the end of the previous round.[20] In order to separate the effect of the content of rank-order feedback from the effects of other drivers of effort provision, the average value of the rank response function over the N possible ranks is normalized to zero, i.e., we impose that $\sum_{k=1}^{N} H(k) = 0$.[21] Given this normalization, a rank response function that is zero at all previous ranks represents the case where effort provision does not respond to the content of rank-order relative-performance feedback. Meanwhile, a rank response function that exhibits both positive and negative values corresponds to a setting where effort provision increases in response to some previous ranks (that is, those ranks that are associated with positive values of the rank response function) and decreases in response to some other previous ranks (that is, those ranks that are associated with negative values of the rank response function).[22] Next, we explain our strategy for identifying the shape of the rank response function.

[20] The rank response function can be interpreted as the average response to each specific rank across subjects and rounds.

[21] Note that the other drivers of effort include any responses to the provision of rank-order feedback that apply irrespective of the content of the feedback: these can vary by subject and are captured by the round effects, subject-specific observed characteristics and unobservables.

[22] An alternative normalization would be to set $H(k) = 0$ for a particular value of k. Such alternative normalizations would shift the rank response function up or down but would not change how the rank response function varies with previous rank.

3.2 Identification strategy: The tied-groups estimator

Estimation of the rank response function is challenging because the unobserved drivers of effort are likely to be serially dependent over rounds. To understand the complications posed by serial dependence, note that a subject's rank in the previous round was partly determined by her effort and hence her unobservables in the previous round. Thus, in the presence of serially-dependent unobservables, rank in the previous round will not be independent of current-round unobservables. The non-independence of previous rank and current-round unobservables confounds the estimate of the rank response function obtained by running an Ordinary Least Squares (OLS) regression on (1).[23]

Serial dependence in the unobserved drivers of effort can take innumerable forms. Possible causes of serial dependence include permanent between-subject differences in unobserved ability or motivation, across-subject heterogeneity in rates of learning over rounds, subjects who work in spurts and regression to the mean. For instance, if subjects work in spurts, meaning that they alternate over rounds between working hard and resting, then the unobservables will be negatively autocorrelated, giving rise to non-causal negative correlation between rank in round r and effort in round $r+1$ (recall that we think of rank as ordered from its lowest value of N to its highest value of 1). To give another example, across-subject heterogeneity in rates of learning over rounds will cause positive autocorrelation in the unobservables that cannot be controlled using a common learning trend; as a result, faster learners will tend to both improve their rank and increase effort over rounds, giving rise to non-causal positive correlation between rank in round r and effort in round $r+1$. Between-subject differences in permanent unobserved ability will also generate a positive non-causal correlation between rank in round r and effort in round $r+1$: subjects with high ability will tend to have high effort and, therefore, high rank in round r, and will also tend to have high effort in round $r+1$, while the opposite holds for low ability subjects. To side-step the potential confounds arising from serially-dependent unobservables, we propose an econometric approach that uses random variation in the allocation of rank among subjects that exerted the same effort in a given round. Our approach allows us to parse out the causal effect of the content of rank-order feedback on subsequent effort provision while allowing any form of serially-dependent unobservables.

In more detail, our econometric approach exploits two sources of random variation in the allocation of rank among subjects that exerted the same effort in a given round. First, by breaking ties at random, we create random variation in rank for subjects from the same session that exerted the same effort in a given round ('within-session ties'). Within groups of subjects that tied in a given session and round, the random allocation of rank ensures that rank is independent of subject characteristics, whether observed or unobserved. Thus, we can use within-session ties from round 1, before the subjects experienced any feedback, and after round 1, when they might have experienced and responded to different feedback. Second, we can also use random variation in rank for subjects from different sessions that exerted the same effort in a given round ('across-session ties'); this source of random variation in rank is due to random variation in the subject composition of sessions. However, we need to ensure that the subjects

[23]Note, because effort depends on previous rank via the potentially non-linear rank response function, $H(\text{Rank}_{n,s,r-1})$, independence rather than zero correlation is required for OLS to provide a causal estimate of how effort responds to the content of rank-order feedback.

who exerted the same effort across different sessions do not differ systematically from each other. As a result, we can only use across-session ties in the first round, since effort after the first round might have been influenced by previous feedback that will differ with the random variation in session composition. Furthermore, we can only use across-session ties for subjects in the same sub-treatment, since effort correlates with sub-treatment (although not statistically significantly: see Supplementary Web Appendix C).[24]

Formally, we define a 'tied group' as follows. In round 1, a tied group is a set of subjects (of cardinality greater than one) from the same sub-treatment that all exerted the same effort. When more than 2 subjects from the same sub-treatment exerted a given level of effort, they are all included in the same tied group. In this case, the tied group can sometimes include both within-session and across-session ties simultaneously: for example, a tied group could include 3 subjects that all exerted effort of 100 in round 1, with 2 subjects coming from the same session and the third subject coming from a different session in the same sub-treatment. In rounds 2-5, a tied group is a set of subjects from the same session that all exerted the same effort in a given round.[25] Thus, tied groups include all within-session ties, and further include all the across-session ties that we are permitted to use (as explained above, those in round 1 that only include subjects from the same sub-treatment). As outlined in Table 1 in Section 3.3, we have 151 tied groups: 99 include only within-session ties, 37 include only across-session ties, and 15 include both.

We can estimate the rank response function in (1) by focusing on the sample $\mathcal{G}$ of subject-session-round observations for which the subject was part of a tied group in the previous round. In particular, we estimate a fully flexible specification of the unknown rank response function, which includes a dummy variable for each of the $N = 17$ possible values of rank in the previous round:

$$H(\text{Rank}_{n,s,r-1}) = \sum_{k=1}^{N} \varphi_k \mathbf{1}_{\{k\}}(\text{Rank}_{n,s,r-1}), \tag{2}$$

where the indicator function $\mathbf{1}_{\{k\}}(\text{Rank}_{n,s,r-1})$ takes the value 1 if $\text{Rank}_{n,s,r-1} = k$ and zero otherwise. Letting $g = 1, ..., G$ index all the tied groups, the equation to be estimated is therefore:

$$\text{Effort}_{n,s,r} = \sum_{k=1}^{N} \varphi_k \mathbf{1}_{\{k\}}(\text{Rank}_{n,s,r-1}) + \eta_g + \beta X_{n,s} + \epsilon_{n,s,r} \quad \text{for} \quad (n, s, r) \in \mathcal{G}, \tag{3}$$

where: η_g for $g = 1, ..., G$ are fixed effects for the subject's tied group in the previous round, which absorb the round fixed effects in (1); $X_{n,s}$ continues to denote observed subject-specific characteristics; and $\epsilon_{n,s,r}$ denotes unobserved effort shifters that remain after controlling for the tied-group fixed effects. $X_{n,s}$ consists of dummy variables for each combination of demographic

[24]Using across-session ties requires that subjects do not condition effort on any characteristics of the session itself, such as characteristics of the other subjects in the session or the time or day of the session. To test whether, within sub-treatment, subjects condition effort on session characteristics, we test the joint significance of the effects of the session dummies on effort provision after controlling for sub-treatment level effects by including sub-treatment dummies. We comfortably fail to reject the hypothesis that subjects do not condition effort on session characteristics ($p = 0.346$). Supplementary Web Appendix B.1 provides further details.

[25]Ties in round 6 are not relevant, since there is no subsequent round in which we can measure effort.

characteristics (see Section 2.4 for a description of the demographic characteristics); because we impose that $\sum_{k=1}^{N} H(k) = 0$ (see Section 3.1), the estimated rank response function is invariant to the chosen reference category.

We call the estimator of the rank response function obtained by applying a fixed-effects regression with tied-group fixed effects to (3) the 'tied-groups estimator'. The tied-group fixed effects η_g absorb all between-tied-group variation in effort, and so the rank response function is identified purely from within-tied-group variation in rank. Moreover, as explained above, rank within a tied group is allocated randomly, either by the random breaking of within-session ties or due to randomness in session composition; thus, the identifying variation in rank is independent of all other drivers of effort provision. Consequently, the tied-groups estimator is an unbiased estimator of the rank response function.[26]

Finally, we want to make clear that our identification strategy is robust to differences in ability. As in any real-effort experiment, our subjects are likely to exhibit heterogeneous ability. Since we use "effort" to denote performance, we think of differences in ability as being captured by differences in the cost of effort function: for a more able subject, each unit of effort/performance is less costly. Equivalently, we could think of each of unit of effort as having the same cost, but different performance effects, across subjects. Either way, two or more subjects with the same performance in a tied group might be of different ability and hence have experienced different effort costs. However, the allocation of rank within groups of tied subjects (who share the same performance) is random and independent of ability, and hence the estimated effect of rank on subsequent performance is not driven by selection effects on ability but is instead driven by changes in effort.

3.3 Estimation sample

Table 1 reports descriptive statistics for the sample $\mathcal{G}$ of subject-session-round observations for which the subject was part of a tied group in the previous round.[27] This sample contains 350 subject-session-round observations, divided between 151 tied groups, and includes 203 distinct

[26] The random allocation of rank within tied groups also ensures that rank is independent of observed subject characteristics. Therefore, the inclusion of observed subject characteristics in (3) is not critical for the identification strategy, but may increase precision by absorbing variation in effort not explained by the content of rank-order feedback. Similarly, due to the random allocation of rank within tied groups, it is not necessary to include subject fixed effects when applying the tied-groups estimator. We do not include subject fixed effects: since subject fixed effects fully absorb all information on subjects who appear only once in the sample of tied groups, including subject fixed effects would substantially reduce the number of observations that are used to estimate the rank response function.

[27] In a small number of cases, there was no variation in rank among the subjects in a tied group. Note that this was only possible in tied groups including only across-session ties. To avoid misleading the reader about the amount of useful variation in the data, we exclude such tied groups from the descriptive statistics and the analysis; however, including them does not alter the results. Subject-session-round observations in $\mathcal{G}$ have a slight tendency to come from the middle, rather than the top or the bottom, of the distribution of ranks: the middle seven ranks make up 41% of ranks and 56% of observations, while the top five ranks make up 29% of ranks and 23% of observations, and the bottom five ranks make up 29% of ranks and 21% of observations. In more detail, for each rank we provide the number of subjects who were in a tied group in which at least one subject achieved that rank, together with the number of subjects who achieved that rank: $(1, 10, 5)$; $(2, 35, 17)$; $(3, 36, 17)$; $(4, 34, 16)$; $(5, 59, 25)$; $(6, 62, 26)$; $(7, 68, 28)$; $(8, 78, 31)$; $(9, 70, 27)$; $(10, 71, 29)$; $(11, 72, 29)$; $(12, 63, 25)$; $(13, 52, 21)$; $(14, 44, 20)$; $(15, 40, 17)$; $(16, 29, 13)$; $(17, 8, 4)$. The non-uniformity of the distribution of observations across ranks will not bias our estimate of the rank response function but may reduce precision. We also note that, since the distribution of ranks is uniform in the full sample, the overall increase in effort in response to the provision of rank-order relative-performance feedback that we reported in Figure 1 is not disproportionately driven by middle-ranking observations.

subjects. Groups of tied subjects contain an average of 2.3 subjects, and the largest tied group contains 5 subjects. The random allocation of rank within tied groups generated considerable within-group variation in subjects' ranks: the within-tied-group standard deviation of rank is 1.4 and the average within-tied-group range of rank is 1.9.

Subject-session-round observations in $\mathcal{G}$	350
Tied groups	151
Only within-session ties	*99*
Only across-session ties	*37*
Both within-session and across-session ties	*15*
Subjects	203
Mean number of subjects per tied group	2.318
Minimum number of subjects in a tied group	2
Maximum number of subjects in a tied group	5
Within-tied-group standard deviation of rank	1.389
Mean within-tied group range of rank	1.914
Minimum within-tied-group range of rank	1
Maximum within-tied-group range of rank	6

Notes: Descriptive statistics refer to the sample of subject-session-round observations for which the subject was part of a tied group in the previous round. The within-tied-group standard deviation of rank and the mean within-tied-group range of rank are weighted by the number of subjects in each group.

Table 1: Descriptive statistics for the sample of tied groups.

3.4 Results

3.4.1 Shape of the fully flexible rank response function

As explained in Section 3.2, we use the tied-groups estimator to establish the causal effect of the content of rank-order feedback on subsequent effort provision. In particular, we apply a fixed-effects regression with tied-group fixed effects to (3) in order to establish the shape of the fully flexible rank response function given by (2), which includes a dummy variable for each of the $N = 17$ possible values of rank in the previous round. The estimation sample $\mathcal{G}$ is described in Table 1. Figure 2 shows the fully flexible rank response function.

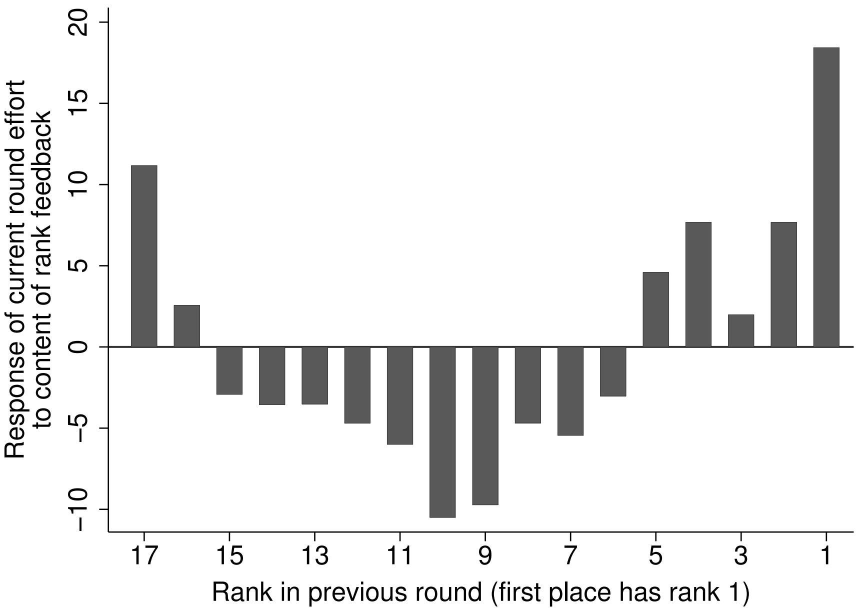


Figure 2: Fully flexible rank response function.

Figure 2 shows that subjects respond strongly to the specific rank that they achieve (we report tests of statistical significance in Section 3.4.2). The rank response function is broadly U-shaped. In particular, we can see that subjects increase their effort the most in response to the content of rank-order feedback when they are ranked first or last: we call this motivating effect of high and low rank 'first-place loving' and 'last-place loathing'. Recall from Section 3.1 that, in order to isolate the response to the content of rank-order feedback from any responses to the provision of rank-order feedback that apply irrespective of the content of the feedback, we have normalized the average value of the rank response function over the 17 possible ranks to zero. Compared to this base level of zero, being ranked first increases effort by 18.4 units, while being ranked last increases effort by 11.2 units. These increases are substantial: the average level of effort across all rounds in the Treatment group that receives rank-order feedback is 85.7; relative to this average, being ranked first increases effort by 21%, while being ranked last increases effort by 13%. Figure 2 also shows that being ranked in the middle of the pack, that is being ranked 9th or 10th, reduces effort by about 10 units, a decrease of more than 10% relative to the average level of effort in the Treatment group (although the effort of the subjects ranked 9th or 10th is still higher than the average level of effort in the Baseline group that does not receive any rank-order feedback). In Section 3.4.2, we show that a quadratic specification of the rank response function captures the pattern accurately.

It is important to emphasize that we have identified the *causal* and *pure* effect of the content of feedback about rank order in the distribution of performance on effort provision. As explained in Section 3.2, the tied-groups estimator leverages randomness in the allocation of rank to give a causal estimate that purges completely the confounding effects of serially dependent unobservables. Furthermore, since we used a flat-wage payment scheme, we have identified the pure effect of the content of feedback about rank in the distribution of performance, uncontaminated by responses to feedback about rank in the distribution of earnings that might be driven by preferences over rank in the distribution of money or a desire for monetary benefits associated with higher rank. Our laboratory setting with random selection of subjects further ensures that responses to the content of rank-order feedback are not driven by longer-term reputational concerns.

Finally, we find that the tied-groups estimation procedure is critical to obtaining reliable results on how effort responds to the content of rank-order feedback. Supplementary Web Appendix B.2 shows that the rank response functions obtained from standard random and fixed effects panel data estimators differ markedly from the rank response function obtained using the tied-groups estimator and illustrated in Figure 2, suggesting that the standard panel data estimators suffer from confounds discussed in Section 3.2. In particular, the standard panel data estimators do not detect first-place loving or last-place loathing.

3.4.2 Quadratic specification, significance tests and robustness checks

In order to conduct tests of statistical significance and check for robustness, we impose some structure on the rank response function. In particular, we replace the fully flexible rank response function (2) with a quadratic specification:

$$H(\text{Rank}_{n,s,r-1}) = \delta_1 \text{Rank}_{n,s,r-1} + \delta_2 (\text{Rank}_{n,s,r-1})^2. \tag{4}$$

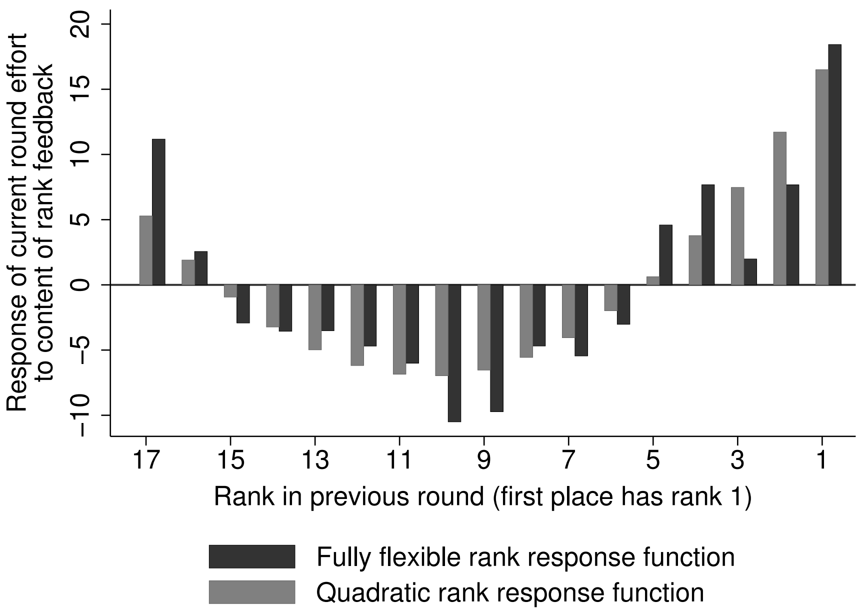


Figure 3: Fully flexible rank response function and quadratic rank response function.

Figure 3 compares the fully flexible rank response function, obtained in Section 3.4.1 using the tied-groups estimation procedure, to the U-shaped quadratic rank response function obtained using the same tied-groups estimation procedure. Conveniently, we find that the quadratic rank response function closely approximates the fully flexible rank response function. We therefore proceed to conduct significance tests and robustness checks using the quadratic rank response function.[28]

	(1)	(2)	(3)
Rank in the previous round (δ_1)	-5.604^{***}	-4.478^{**}	-5.186^{***}
	[0.000]	[0.014]	[0.004]
	(1.216)	(1.821)	(1.780)
Squared rank in the previous round (δ_2)	0.272^{***}	0.213^{**}	0.254^{***}
	[0.000]	[0.021]	[0.007]
	(0.063)	(0.092)	(0.095)
Demographic controls	Yes	No	Yes
First-round effort controls	No	No	Yes
Subject-session-round observations	350	350	350
Tied groups	151	151	151
Subjects	203	203	203
Test for no response to previous rank ($\delta_1 = \delta_2 = 0$), p value	0.007	0.051	0.016

Notes: Parameter estimates of the quadratic specification (4) were obtained using the tied-groups estimation procedure described in Section 3.2. Two-sided p values are shown in square brackets and heteroskedasticity-consistent standard errors, with clustering at the tied-group level, are shown in round brackets (see footnote 28 for discussion of the appropriate level for clustering the standard errors). Demographic controls are dummy variables for each combination of demographic characteristics (see Section 2.4 for a description of the demographic characteristics). First-round effort controls are first-round effort raised to the power of j for $j = 1, 2, 3, 4$. *, ** and *** denote significance at the 10%, 5% and 1% levels (2-sided tests).

Table 2: Significance tests and robustness checks for the quadratic rank response function.

The first column of Table 2 reports the coefficients of the quadratic rank response function (4), that is the coefficients on rank in the previous round (δ_1) and squared rank in the previous round (δ_2), obtained using the tied-groups estimation procedure. As described in Section 3.2, we control for demographics by including a dummy variable for each combination of demographic characteristics. The coefficient on squared rank is statistically significantly different from zero at the 1% level. We also reject the hypothesis that effort does not respond to the content of rank-order feedback, i.e., we reject the joint hypothesis that $\delta_1 = 0$ and $\delta_2 = 0$ ($p = 0.007$). When we augment the specification to include interactions between the previous rank variables and sub-treatment indicators, we find no statistically significant differences according to whether the feedback was provided publicly or privately.[29] Also, when we estimate separate rank response functions for effort on the numerical task and effort on the verbal task, we find no statistically

[28]The significance tests use heteroskedasticity-consistent standard errors, with clustering at the tied-group level. By clustering in this way, we account for any across-subject correlation in unobservables within tied groups. Angrist and Pischke (2008, p. 311) show that clustering at a particular level is not required if the explanatory variable is uncorrelated within clusters, even if unobservables are correlated within clusters. This result implies that we do not need to additionally cluster at the subject, session or round levels because the random assignment of rank within tied groups ensures that the within-tied-group variation in rank is uncorrelated across observations of: (i) the same subject; (ii) different subjects in different rounds of the same session; and (iii) subjects in different tied groups in the same round and session. However, the sum of ranks is fixed across subjects from the same session in a tied group, which creates negative correlation in the explanatory variable within tied groups.

[29]A test of the joint significance of the interactions between the previous rank variables and sub-treatment indicators gives $p = 0.784$.

significant differences in the pattern of response across the two real-effort tasks.[30] Together, these results provide statistical support for a U-shaped rank response function.

Columns 2 and 3 of Table 2 show that our findings are robust to varying the control variables included in the estimation (as explained in footnote 26, the-tied groups estimator identifies the rank response function irrespective of whether or how we control for observables). In Column 2, we drop demographic controls from the estimation: this leads to a slight reduction in precision, as expected, but the coefficients do not change much. Similarly, Column 3 shows that the coefficients on the previous rank variables are robust to adding controls for first-round effort.

The estimates of the parameters of the quadratic rank response function are also robust to restricting the sample to within-session ties: estimating specification (1) in Table 2 using this subsample gives $\hat{\delta}_1 = -4.106$ and $\hat{\delta}_2 = 0.231$.[31] When we estimate separate rank response functions for observations from within-session tied groups and across-session tied groups, we find no statistically significant difference in the pattern of response across the two types of tied group ($p = 0.816$).

3.4.3 Demographics

Finally, we consider whether responses to the content of rank-order feedback vary according to demographic characteristics. As described in Section 2.4, we collected data on four demographic characteristics: gender (male vs. female), age (aged 22 or above vs. below 22), country of birth (United Kingdom vs. other) and field of study (STEMM vs. non-STEMM). First, we augment the quadratic specification (4) that was estimated in Column 1 of Table 2 to include interactions between the previous rank variables and an indicator for being male. We find no statistically significant differences according to gender: a test of the joint significance of the interactions between the previous rank variables and an indicator for being male gives $p = 0.885$. We then repeat the exercise for each of our other demographic characteristics. Again we find no statistically significant differences: the tests for country of birth, age and subject of study give p values of 0.675, 0.900 and 0.526, respectively. Our results thus suggest that the phenomena of first-place loving and last-place loathing are not restricted to specific demographic groups, but instead are more universal in their manifestation.

4 Conclusion

Ranking performance is a popular approach for the business world and in the public sector. Organizations frequently use rank-order relative-performance evaluation to motivate individuals: bonuses, promotions, performance appraisals and symbolic awards often depend on rank in the distribution of performance. Public sector institutions are often ranked through public league tables. Students exerting effort in educational environments also receive rank feedback from the results of examinations and tests.

It is important to know what effect this rank feedback has on those receiving the news and how this feeds into how best to communicate levels of pay, promotion decisions, appraisals,

[30] A test of the joint significance of the cross-task differences in the coefficients on the previous rank variables gives $p = 0.531$.

[31] The p-values are 0.265 and 0.174, respectively.

and educational performance. Despite its popularity, there is little consensus about exactly how people respond to rank-order relative-performance evaluation: there is an active debate about the effectiveness of rank-order feedback provided by organizations (e.g., Prendergast, 1999, Grote, 2005, Hazels and Sasse, 2008), while companies such as General Electric, Yahoo, and Whirlpool continue to experiment with different forms of relative-performance feedback (Kuhnen and Tymula, 2012).

The existing experimental literature was able to show that rank feedback motivates people on average. We extend this literature by providing causal evidence of how individuals respond to the content of rank-order feedback. Our findings show that the specific rung that someone occupies in the ranking ladder determines how much effort they put in afterwards: people have a pure taste for rank in the distribution of performance that operates independently of long-term reputational considerations or any desire for higher relative or absolute compensation. We also show that the effects of ranking are quantitatively important: people who are told that they are among the best or worst performers respond by increasing effort provision substantially, relative to those who are informed that they rank in the middle of the pack.

The U-shaped rank response pattern that we find in this study has important implications for the design of effective performance feedback policies, workplace organizational structures and incentive schemes. In particular, the design of feedback policies and organizational structures should take into account the implicit incentives generated by people's preferences over rank. Our results suggest that ranking is particularly effective in incentivizing individuals who put in very good or very poor performances. The results highlight the value of awarding symbolic prizes to the high performers or scheduling regular appraisals with the worst performers. However, our findings are also cautionary: ranking might demoralize those in the middle if done too often, and middle-ranking employees are often the most loyal (possibly because they are less likely to be poached or fired). The U-shaped pattern of response also has implications for organizational design: decentralized organizational structures might help to reduce the number of middle ranks that induce relatively low subsequent performance, and organizations should consider matching people into groups with similar abilities, so everyone can realistically compete for high rank.

Our study looks at the response of individuals to their own rank when they exert real effort in the laboratory. The power of laboratory experiments allows us to identify cleanly the causal effects of a pure taste for rank in the distribution of performance uncontaminated by a preference for rank in the distribution of earnings or a desire for reputational benefits associated with higher rank. Just like any experiment, we cannot be sure how the effects that we identify might vary with the specifics of our experimental set-up: it would be useful for future work to check that our findings are robust to introducing, for example, different group sizes and other real-effort tasks. We also hope that future work will help us to understand how the pure task for rank that we have identified in the laboratory interacts with the richness of work environments in the field.

Future research could also investigate whether rank feedback about organizational performance induces similar behavioral responses from the individuals in the ranked organization. If so, our findings will also help policymakers to design feedback policies for organizational performance such as hospital and school league tables. Finally, we hope that our findings will spur theoretical and empirical work that develops further the implications of responses to the content of rank-order feedback for how organizations should interact with and attempt to motivate their staff, for whether policymakers should report rank-order organizational performance, and for how educational institutions should provide relative-performance feedback to their students.

References

Abeler, J., **Falk, A.**, **Goette, L.**, and **Huffman, D.** (2011). Reference points and effort provision. *American Economic Review*, 101(2): 470–492

Angrist, J.D. and **Pischke, J.S.** (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press

Ashraf, N., **Bandiera, O.**, and **Lee, S.S.** (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior & Organization*, 100: 44–63

Azmat, G. and **Iriberri, N.** (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, 94(7): 435–452

Azmat, G. and **Iriberri, N.** (2016). The provision of relative performance feedback: An analysis of performance and satisfaction. *Journal of Economics and Management Strategy*, 25(1): 77–110

Baker, G.P., **Jensen, M.C.**, and **Murphy, K.J.** (1988). Compensation and incentives: Practice vs. theory. *Journal of Finance*, 43(3): 593–616

Bandiera, O., **Barankay, I.**, and **Rasul, I.** (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association*, 11(5): 1079–1114

Barankay, I. (2012). Rank incentives: Evidence from a randomized workplace experiment. *Mimeo, University of Pennsylvania*

Barankay, I. (2011). Rankings and social tournaments: Evidence from a crowd-sourcing experiment. *Mimeo, University of Pennsylvania*

Bellemare, C., **Lepage, P.**, and **Shearer, B.** (2010). Peer pressure, incentives, and gender: An experimental analysis of motivation in the workplace. *Labour Economics*, 17(1): 276–283

Bhattacharya, H. and **Dugar, S.** (2012). Status incentives and performance. *Managerial and Decision Economics*, 33(7-8): 549–563

Blanes i Vidal, J. and **Nossol, M.** (2011). Tournaments without prizes: Evidence from personnel records. *Management Science*, 57(10): 1721–1736

Bø, E.E., **Slemrod, J.**, and **Thoresen, T.O.** (2015). Taxes on the Internet: Deterrence effects of public disclosure. *American Economic Journal: Economic Policy*, 7(1): 36–62

Boyce, C.J., **Brown, G.D.A.**, and **Moore, S.C.** (2010). Money and happiness: Rank of income, not income, affects life satisfaction. *Psychological Science*, 21(4): 471–475

Bradler, C., **Dur, R.**, **Neckermann, S.**, and **Non, A.** (forthcoming). Employee recognition and performance: A field experiment. *Management Science*

Brown, G.D.A., **Gardner, J.**, **Oswald, A.J.**, and **Qian, J.** (2008). Does wage rank affect employees' well-being? *Industrial Relations*, 47(3): 355–389

Burrows, P. and **Loomes, G.** (1994). The impact of fairness on bargaining behaviour. *Empirical Economics*, 19: 201–221

Cadsby, C.B., **Engle-Warnick, J.**, **Fang, T.**, and **Song, F.** (2014). Psychological incentives, financial incentives, and risk attitudes in tournaments: An artefactual field experiment. *University of Guelph Department of Economics and Finance Discussion Paper 2014-03*

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

Clark, A.E., **Masclet, D.**, and **Villeval, M.C.** (2010). Effort and comparison income: Experimental and survey evidence. *Industrial & Labor Relations Review*, 63(3): 407–426

Clark, A.E., **Westergård-Nielsen, N.**, and **Kristensen, N.** (2009). Economic satisfaction and income rank in small neighbourhoods. *Journal of the European Economic Association*, 7(2-3): 519–527

Coffey, B. and **Maloney, M.T.** (2010). The thrill of victory: Measuring the incentive to win. *Journal of Labor Economics*, 28(1): 87–112

Delfgaauw, J., **Dur, R.**, **Sol, J.**, and **Verbeke, W.** (2013). Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics*, 31(2): 305–326

Dohmen, T., **Falk, A.**, **Fliessbach, K.**, **Sunde, U.**, and **Weber, B.** (2011). Relative versus absolute income, joy of winning, and gender: Brain imaging evidence. *Journal of Public Economics*, 95(3): 279–285

Dutcher, E.G., **Balafoutas, L.**, **Lindner, F.**, **Ryvkin, D.**, and **Sutter, M.** (2015). Strive to be first or avoid being last: An experiment on relative performance incentives. *Games and Economic Behavior*, 94: 39–56

Ebeling, F., **Fellner, G.**, and **Wahlig, J.** (2012). Peer pressure in multi-dimensional work tasks. *University of Cologne Working Paper in Economics 57*

Ehrenberg, R.G. and **Bognanno, M.L.** (1990). Do tournaments have incentive effects? *Journal of Political Economy*, 98(6): 1307–1324

Ellingsen, T. and **Johannesson, M.** (2007). Paying respect. *Journal of Economic Perspectives*, 21(4): 135–150

Elsner, B. and **Isphording, I.E.** (2015). A big fish in a small pond: Ability rank and human capital investment. *IZA Discussion Paper 9121*

Eriksson, T., **Mao, L.**, and **Villeval, M.C.** (2015). Saving face and group identity. *GATE Working Paper 1515*

Eriksson, T., **Poulsen, A.**, and **Villeval, M.C.** (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, 16(6): 679–688

Falk, A. and **Ichino, A.** (2006). Clean evidence on peer effects. *Journal of Labor Economics*, 24(1): 39–57

Fershtman, C. and **Gneezy, U.** (2011). The tradeoff between performance and quitting in high power tournaments. *Journal of the European Economic Association*, 9(2): 318–336

Festinger, L. (1954). A theory of social comparison processes. *Human relations*, 7(2): 117–140

Frank, R.H. (1985). *Choosing the Right Pond: Human Behavior and the Quest for Status.* New York: Oxford University Press

Freeman, R.B. and **Gelber, A.M.** (2010). Prize structure and information in tournaments: Experimental evidence. *American Economic Journal: Applied Economics*, 2(1): 149–164

Frey, B.S. (2007). Awards as compensation. *European Management Review*, 4(1): 6–14

Genakos, C. and **Pagliero, M.** (2012). Interim rank, risk taking, and performance in dynamic tournaments. *Journal of Political Economy*, 120(4): 782–813

Georganas, S., **Tonin, M.**, and **Vlassopoulos, M.** (2015). Peer pressure and productivity: The role of observing and being observed. *Journal of Economic Behavior and Organization*, 117: 223–232

Gerhards, L. and **Siemer, N.** (2014). Private versus public feedback - The incentive effects of symbolic awards. *Aarhus University Economics Working Paper 2014-01*

Gibbons, R. and **Murphy, K.J.** (1990). Relative performance evaluation for chief executive officers. *Industrial & Labor Relations Review*, 43(3): 30S–51S

Gill, D. and **Prowse, V.** (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1): 469–503

Gill, D. and **Prowse, V.** (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics*, 5(2): 351–376

Gill, D. and **Stone, R.** (2010). Fairness and desert in tournaments. *Games and Economic Behavior*, 69(2): 346–364

Gino, F. and **Staats, B.R.** (2011). Driven by social comparisons: How feedback about coworkers' effort influences individual productivity. *Harvard Business School Working Paper 11-078*

Gneezy, U. and **Rustichini, A.** (2004). Gender and competition at a young age. *American Economic Review: Papers & Proceedings*, 94(2): 377–381

Goerg, S.J., **Kube, S.**, and **Radbruch, J.** (2017). The effectiveness of incentive schemes in the presence of implicit effort costs. *IZA Discussion Paper 10546*

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1): 114–125

Grote, R.C. (2005). *Forced Ranking: Making Performance Management Work.* Boston, MA; Harvard Business School Press

Hannan, R.L., **Krishnan, R.**, and **Newman, A.H.** (2008). The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. *Accounting Review*, 83(4): 893–913

Hannan, R.L., **McPhee, G.P.**, **Newman, A.H.**, and **Tafkov, I.D.** (2013). The effect of relative performance information on performance and effort allocation in a multi-task environment. *Accounting Review*, 88(2): 553–575

Hazels, B. and **Sasse, C.M.** (2008). Forced ranking: A review. *SAM Advanced Management Journal*, 73(2): 35–39

Heffetz, O. and **Frank, R.H.** (2011). Preferences for status: Evidence and economic implications. In J. Benhabib, M.O. Jackson, and A. Bisin, editors, *Handbook of Social Economics, Vol. 1A*, 69–91. Amsterdam: North Holland

Holmstrom, B. and **Milgrom, P.** (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, 7: 24–52

Huberman, B.A., **Loch, C.H.**, and **Önçüler, A.** (2004). Status as a valued resource. *Social Psychology Quarterly*, 67(1): 103–114

Iso-Ahola, S.E. and **Dotson, C.O.** (2014). Psychological momentum: Why success breeds success. *Review of General Psychology*, 18(1): 19–33

Jalava, N., **Joensen, J.S.**, and **Pellas, E.** (2015). Grades and rank: Impacts of non-financial incentives on test performance. *Journal of Economic Behavior & Organization*, 115: 161–196

Kosfeld, M. and **Neckermann, S.** (2011). Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics*, 3(3): 86–99

Kuhnen, C.M. and **Tymula, A.** (2012). Feedback, self-esteem, and performance in organizations. *Management Science*, 58(1): 94–113

Kuziemko, I., **Buell, R.W.**, **Reich, T.**, and **Norton, M.I.** (2014). "Last-place aversion": Evidence and redistributive implications. *Quarterly Journal of Economics*, 129(1): 105–149

Lazear, E.P. and **Rosen, S.** (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, 89(5): 841–864

Legge, S. and **Schmid, L.** (2015). Media attention and betting markets. *University of St. Gallen, Department of Economics Working Paper 2015-21*

Lount Jr., R.B. and **Wilk, S.L.** (2014). Working harder or hardly working? Posting performance eliminates social loafing and promotes social laboring in workgroups. *Management Science*, 60(5): 1098–1106

Malcomson, J.M. (1984). Work incentives, hierarchy, and internal labor markets. *Journal of Political Economy*, 92(3): 486–507

Mas, A. and **Moretti, E.** (2009). Peers at work. *American Economic Review*, 99(1): 112–145

Miller, G.A. (1995). Wordnet: a lexical database for English. *Communications of the ACM*, 38(11): 39–41

Moers, F. (2005). Discretion and bias in performance evaluation: The impact of diversity and subjectivity. *Accounting, Organizations and Society*, 30(1): 67–80

Murphy, R. and **Weinhardt, F.** (2014). Top of the class: The importance of ordinal rank. *CESifo Working Paper 4815*

Murthy, U.S. (2010). The effect of relative performance information under different incentive schemes on performance in a production task. *Mimeo, University of South Florida*

Murthy, U.S. and **Schafer, B.A.** (2011). The effects of relative performance information and framed information systems feedback on performance in a production task. *Journal of Information Systems*, 25(1): 159–184

Niederle, M. and **Vesterlund, L.** (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122(3): 1067–1101

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, 37(1): 7–63

Propper, C. and **Wilson, D.** (2003). The use and usefulness of performance measures in the public sector. *Oxford Review of Economic Policy*, 19(2): 250–267

Rosaz, J., **Slonim, R.**, and **Villeval, M.C.** (2012). Quitting and peer effects at work. *IZA Discussion Paper 6475*

Rustichini, A. (2008). Dominance and competition. *Journal of the European Economic Association*, 6(2-3): 647–656

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

Tafkov, I.D. (2013). Private and public relative performance information under different compensation contracts. *Accounting Review*, 88(1): 327–350

Tincani, M.M. (2015). Heterogeneous peer effects and rank concerns: Theory and evidence. *Mimeo, University College London*

Tran, A. and **Zeckhauser, R.** (2012). Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics*, 96(9): 645–650