ISSN 1471-0498

DEPARTMENT OF ECONOMICS
DISCUSSION PAPER SERIES

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

David Gill, Zdenka Kissova,

Jaesun Lee and Victoria Prowse

Number 783
March 2016

Manor Road Building, Oxford OX1 3UQ

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: March 2, 2016
First version: July 22, 2015

Abstract

Rank-order relative-performance evaluation, in which pay, promotion and symbolic awards depend on the rank of workers in the distribution of performance, is ubiquitous. Whenever firms use rank-order relative-performance evaluation, workers receive feedback about their rank. Using a real-effort experiment, we aim to discover whether workers 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 firms' 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.

*We thank the Nuffield Centre for Experimental Social Sciences for hosting our experiment. We also thank the Economic and Social Research Council and the British Academy for funding the experiment. Finally, we thank Jozef Dobos for research assistance in programming the experiment.

†Department of Economics, University of Oxford; david.gill@economics.ox.ac.uk.
‡PricewaterhouseCoopers; zdenka.kiss@gmail.com.
§Department of Economics, Cornell University; jl2634@cornell.edu.
¶Department of Economics, Cornell University; prowse@cornell.edu.

1 Introduction

Relative-performance evaluation (RPE) is ubiquitous in the workplace. Bonuses, promotions, performance appraisals and symbolic awards often depend on how well employees perform relative to other similar workers in the firm (Gibbons and Murphy, 1990; Prendergast, 1999; Kosfeld and Neckermann, 2011), while some companies base executive compensation on across-firm performance comparisons (Gibbons and Murphy, 1990).

Rank-order RPE, in which pay, promotion, employee appraisals and non-pecuniary awards depend on the rank of workers in the distribution of performance, is particularly popular. 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 workers 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: corporate sector 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 workers 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 firms use rank-order RPE, workers receive feedback about their specific rank in the distribution of performance. As we discuss in detail below, it is therefore important that firms understand how workers 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 workers' preferences over the specific rank that they achieve. Understanding how workers 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.

[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 standard in the experimental literature on effort provision, we use the term 'effort' to mean performance or output in a work task rather than the associated cost of producing the output (see, e.g., Gill and Prowse, 2012). Neither our analysis nor our results depend on this choice of terminology.

The effective design of transparency policies, such as public disclosure of official hospital, school and university league tables in the United Kingdom, 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 $r+1$. Such serial dependence in unobservables can take innumerable forms. Possible causes of serial dependence include permanent between-subject differences in ability, 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, working hard one period and resting the next, then the unobservables will be negatively autocorrelated, giving rise to non-causal negative correlation between rank in round r and effort in round $r+1$.[4] 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 $r+1$.

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

[3]Of course, in many firms 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, firms 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.

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 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.

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 12%. By contrast, being ranked in the middle of the pack, that is being ranked 9th or 10th, reduces effort by 11% 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). Furthermore, as rank increases from 17th to 10th, effort falls monotonically, while effort increases monotonically as rank increases from 9th to 1st (except for a small decrease when rank increases from 4th to 3rd). 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 firms might choose to design their performance feedback policies. In particular, it might be profitable for employers 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 firms need not worry about designing different feedback policies for groups of workers 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. Employers might also find it productive to organize workers into groups with similar abilities, so that all workers have a realistic prospect of obtaining top ranks. Finally, the U-shaped pattern of response also has implications for the optimal design of incentive schemes. Firms 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 workers 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.[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]

As well as discovering how workers 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-

[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] For example: Ehrenberg and Bognanno (1990); Fershtman and Gneezy (2011); Genakos and Pagliero (2012); and Delfgaauw et al. (2013). 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.

order feedback will be available; as far as we aware only Eriksson et al. (2015) fail to find an effect.[8] 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; Heffetz and Frank, 2011).[9]

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 workers respond to the content of rank-order feedback according to whether the feedback was provided publicly or privately. As described in Supplementary Web Appendix C, we also find no differences in the average level of effort by mode of feedback.[10]

The paper proceeds as follows: Section 2 describes the experimental design; Section 3 considers 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, 2004). For each session, invited students

[8] 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).

[9] 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 Clark et al. (2010) find that gift-exchange reciprocity is stronger for workers of high rank, and Kuziemko et al. (2014) find that people are more willing to gamble and less willing to give to less-fortunate others to avoid low rank in the distribution of money in a setting in which endowments are randomly allocated.

[10] The existing literature only considers whether public and private rank-order feedback have different effects on average performance. 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.

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. 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 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 × 15 grid of capital letters and scored one point for each valid English word that they correctly spotted.[11] In the numerical task, subjects added up pairs of 2-digit numbers and scored one point for each pair that they correctly added up.[12] 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.

[11] 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.

[12] 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.

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 standard in real-effort experiments, our use of the term 'effort' therefore corresponds to the behavior of the subject rather than the associated cost of achieving a given points score. Neither our analysis nor our results depend on this choice of terminology.

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.[13] In the Baseline, 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.[14] 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.

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.[15]

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.

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

[14] 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."

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

The subjects were also asked to self-report their degree of competitiveness.[16]

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

As shown in Figure 1 and described in detail in Supplementary Web Appendix C, we find that providing rank-order feedback increases effort substantially on average.[17] This increase 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.

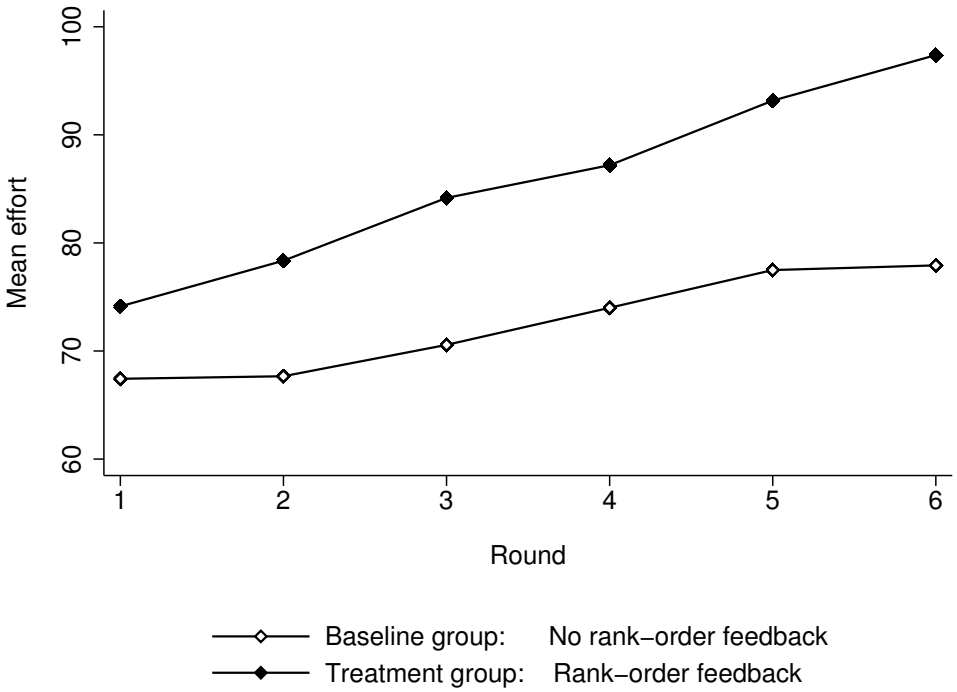


Figure 1: Round-by-round mean effort.

In this section, we seek to shed light on exactly how workers 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.

[16] In Supplementary Web Appendix D we describe the question on competitiveness and consider the interaction between competitiveness and rank-order feedback.

[17] As noted in Section 2.2, we use the term 'effort' to denote a subject's total points score across the two real-effort tasks in a given round. Thus, as is standard in real-effort experiments, our use of the term 'effort' corresponds to the behavior of the subject rather than the associated cost of achieving a given points score. Neither our analysis nor our results depend on this choice of terminology.

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.[18] 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 $\sum_{k=1}^{N} H(k) = 0$. Given this normalization, reported ranks that elicit subsequent effort provision above (below) that expected given the other drivers of effort correspond to positive (negative) values of the rank response function.[19] 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.

Figure 2 illustrates some forms that the rank response function might take when 4 subjects are given rank-order feedback. Panel 2(a) shows the flat rank response function that applies when subjects do not respond to the content of rank-order feedback, e.g., because they do not look at the feedback or because they see the feedback but do not adjust effort provision in response to its specific content. Panel 2(b) shows an increasing rank response function, where subjects are increasingly motivated to provide effort by higher rank-order feedback, and Panel 2(c) shows a reverse case, where subjects are increasingly motivated by lower rank-order feedback. Finally, Panel 2(d) shows a more complex case in which the rank response function is non-monotonic and subjects are most highly motivated after achieving the lowest or highest possible ranks. These figures illustrate just some of the many different forms that the rank response function might take. Next, we explain our strategy for identifying the shape of the rank response function.

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

[19] An alternative normalization would be to set $H(k) = 0$ for a particular value of k. Such alternative normalizations would simply shift the rank response function up or down.

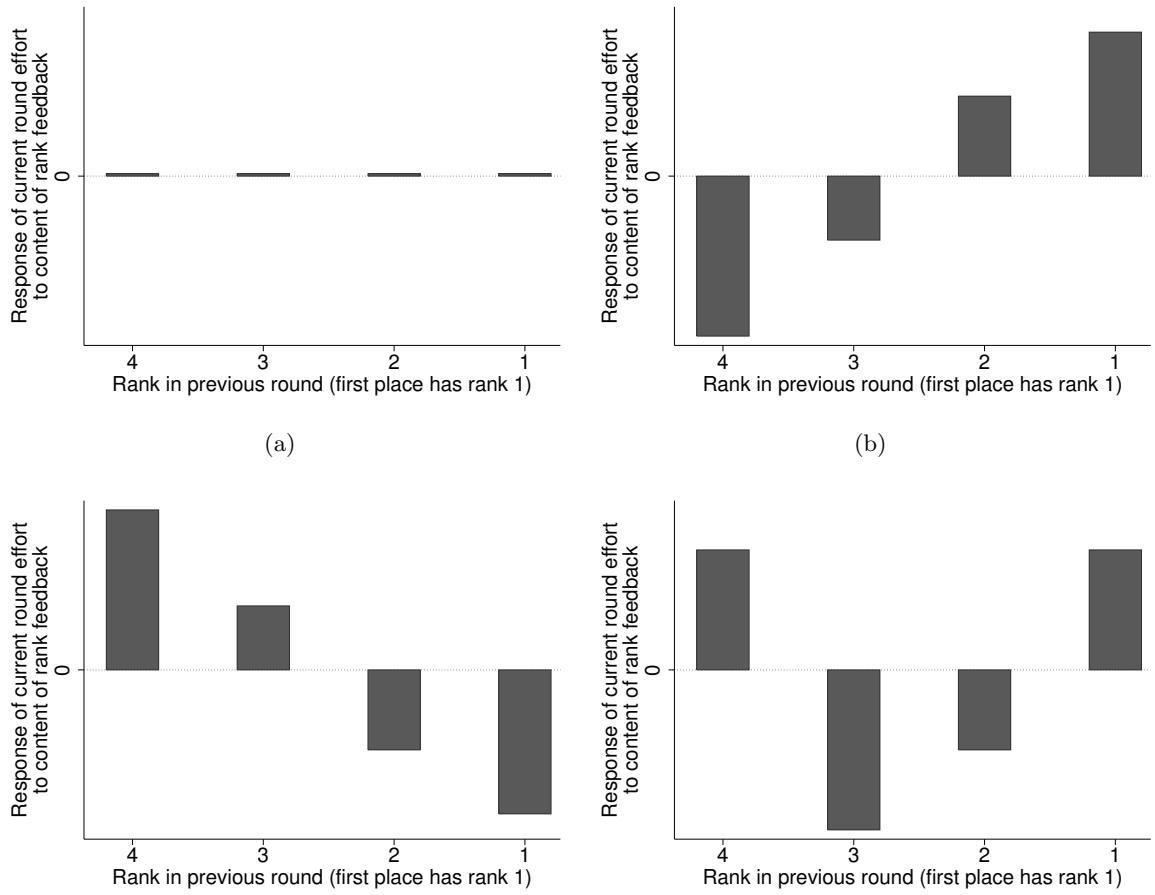


Figure 2: Example rank response functions when 4 subjects are given rank-order feedback.

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).[20]

Serial dependence in the unobserved drivers of effort can take innumerable forms. Possible causes of serial dependence include permanent between-subject differences in 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

[20]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.

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, if subjects learn at systematically different rates, then unobservables will be positively autocorrelated; 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 $r+1$. 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.[21] However, we need to ensure that the subjects 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).

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.[22] 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).

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,

[21] This argument 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. Supplementary Web Appendix B.1 provides evidence that this assumption holds for our sample.

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

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 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 OLS regression 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.[23]

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.[24] This sample contains 350 subject-session-round observations, divided between 151 tied groups, and includes 203 distinct 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.

[23] 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.

[24] 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 $\mathcal{G}$	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 OLS regression 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. After estimating the 17 coefficients on rank in the previous round, we smooth the estimated rank response function to reduce the noisiness of the estimation results.[25] Figure 3 shows the smoothed rank response function. Figure SWA.1 in Supplementary Web Appendix B.2 shows that the non-smoothed rank response function, obtained directly from the 17 coefficients on rank in the previous round, closely resembles the smoothed rank response function.

[25] We use a local linear regression to smooth the estimated rank response function (Fan, 1992, discusses local linear regression; for other applications of local linear regression in economics see, e.g., Blundell and Duncan, 1998, DiNardo and Tobias, 2001 and Becker et al., 2012). The local linear regression estimate of the rank response function at a particular previous rank, k, is given by the prediction from a weighted linear regression that uses the non-smoothed estimates of the rank response function as inputs, weighted according to distance from k, with weights chosen using a Gaussian kernel function with the bandwidth, or smoothing, parameter implied by the Rule of Thumb (ROT) method (see Fan and Gijbels, 1996, pp. 110–112).

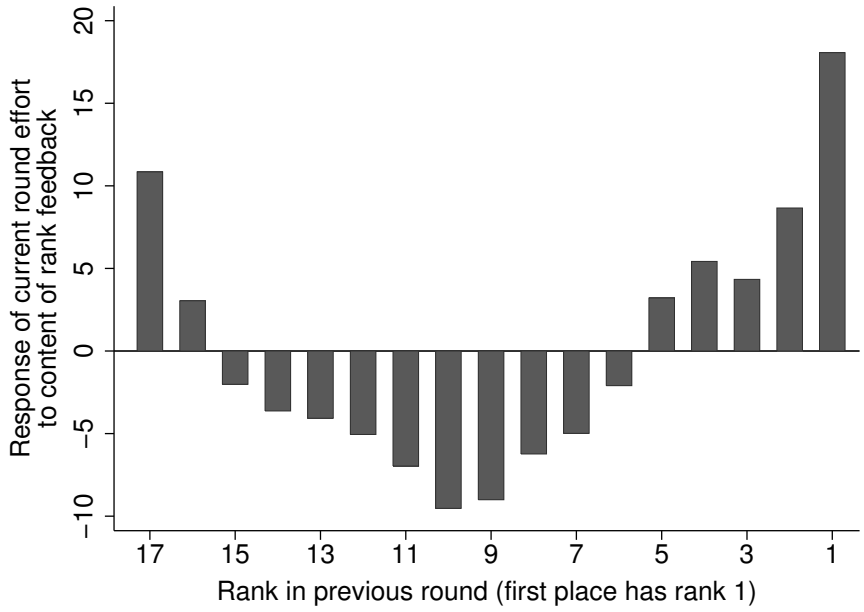


Figure 3: Smoothed fully flexible rank response function.

Figure 3 shows that subjects respond strongly to the specific rank that they achieve. 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 18.1 units, while being ranked last increases effort by 10.8 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 21.1%, while being ranked last increases effort by 12.6%. Figure 3 also shows that being ranked in the middle of the pack, that is being ranked 9th or 10th, reduces effort by about 9 units, a decrease of about 11% 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). Furthermore, we can see from Figure 3 that the response to the content of rank-order feedback exhibits a clear pattern. As rank increases from 17th to 10th, effort falls monotonically, while effort increases monotonically as rank increases from 10th to 1st (except for a small decrease when rank increases from 4th to 3rd). 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.3 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 3, indicating 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:

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

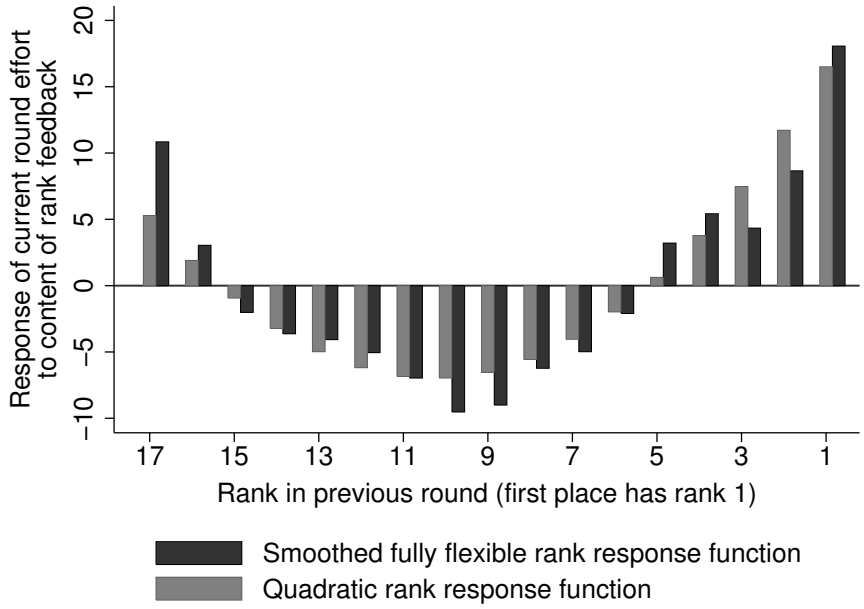


Figure 4: Smoothed fully flexible rank response function and quadratic rank response function.

Figure 4 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.

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 $\delta_2 = 0$ ($p = 0.015$). 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.[26] Together, these results provide statistical support for a U-shaped rank response function.

	(1)	(2)	(3)
Rank in the previous round (δ_1)	-5.604***	-4.478**	−5.186***
	[0.003]	[0.023]	[0.007]
	(1.913)	(1.969)	(1.913)
Squared rank in the previous round (δ_2)	0.272***	0.213**	0.254***
	[0.006]	[0.038]	[0.010]
	(0.099)	(0.103)	(0.098)
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	0.015	0.075	0.027

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 subject level to account for within-subject non-independence across rounds, are shown in round brackets. 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.

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 23, 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.

[26] A test of the joint significance of the interactions between the previous rank variables and sub-treatment indicators gives $p = 0.879$.

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 $p = 0.881$. 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 0.725, 0.906 and 0.539, 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

Firms frequently use rank-order relative-performance evaluation to motivate workers: bonuses, promotions, performance appraisals and symbolic awards often depend on workers' ranks in the distribution of performance. Whenever firms use rank-order relative-performance evaluation, workers receive feedback about their rank. As yet, there is little consensus about exactly how workers respond to rank-order relative-performance evaluation: there is an active debate about the effectiveness of rank-order feedback provided by firms (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).

In this paper, we have shown that workers 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. In particular, we find that workers respond to the specific rank that they achieve. Using a methodology that purges completely the effects of serially dependent unobserved drivers of effort, we find a U-shaped response pattern: workers who are told that they are among the best or worst performers respond by increasing effort provision substantially, relative to workers who are informed that they rank in the middle of the pack.

In the introduction we outline how this U-shaped rank response pattern might affect the design of effective performance feedback policies, workplace organizational structures and incentives schemes that take into account the implicit incentives generated by workers' preferences over rank. We close with a hope that these ideas will spur applied work that considers in more detail, from both a theoretical and an empirical perspective, the implications of workers' responses to the content of rank-order feedback for how firms should interact with and attempt to motivate their employees. The magnitude of the effects that we find are large, which points to the likelihood that workers' responses to the content of rank-order feedback have economically important consequences for how firms should interact with their employees.

References

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*

Becker, A., **Deckers, T.**, **Dohmen, T.**, **Falk, A.**, and **Kosse, F.** (2012). The relationship between economic preferences and psychological personality measures. *Annual Review of Econonics*, 4: 453–478

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

Blundell, R. and **Duncan, A.** (1998). Kernel regression in empirical microeconomics. *Journal of Human Resources*, 33(1): 62–87

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

DiNardo, J. and **Tobias, J.L.** (2001). Nonparametric density and regression estimation. *Journal of Economic Perspectives*, 15(4): 11–28

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

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American statistical Association*, 87(420): 998–1004

Fan, J. and **Gijbels, I.** (1996). *Monographs on Statistics and Applied Probability 66: Local Polynomial Modelling and its Applications*. Chapman & Hall / CRC Press

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*

Girard, Y. and **Hett, F.** (2013). Competitiveness in dynamic group contests: Evidence from combined field and lab data. *Gutenberg School of Management and Economics Discussion Paper 1303*

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

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer and V. Macho, editors, *GWDG-Bericht Nr. 63: Forschung und Wissenschaftliches Rechnen*, 79–93. GWDG: Göttingen

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

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

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

Supplementary Web Appendix

(Intended for Online Publication)

A Experimental instructions

The envelope that you collected on your way in contains the instructions for this session. Please now open this envelope.

[SUBJECTS OPEN ENVELOPE]

Please look at the first page of the instructions. I am reading from these instructions. Mobile phones and any other electronic devices must now be turned off. These must remain turned off for the duration of this session. Please do not use or place on your desk any personal items, including phones, calculators, pens, pencils, etc. Please do not look into anyone else's booth at any time.

Thank you for participating in this session on decision-making. You were randomly selected from the Nuffield Centre for Experimental Social Sciences' pool of subjects to be invited to participate in this session. I remind you that the Nuffield Centre for Experimental Social Sciences has a strict no deception policy. This means that I will not mislead, misinform or lie to you at any point during this session.

During this session there will be a number of pauses for you to ask questions. During such a pause, please raise your hand if you want to ask a question. After you raise your hand, you will be approached by my assistant who will provide you with a pen and paper. You should write your question on the paper and then pass the pen and paper back to my assistant, who will pass the paper to me. I will write my answer on the paper and my assistant will then pass the paper back to you. You should read my answer and then give the paper back to my assistant. Apart from asking questions in this way, you must not communicate with anybody in this room or make any noise.

Are there any questions?

This session will last around 90 minutes. There are 17 participants taking part in this session (excluding myself and my assistants). This session will be divided into 7 rounds. The first round will be a practice round, for which you will not be paid. The remaining 6 rounds will be paying rounds. You will receive a payment of £2.50 per paying round. Additionally, you will be paid a show-up fee of £5 for your attendance today. You will be paid in cash upon your departure today. You will be paid individually in a separate room by a laboratory assistant. ln each round, including the practice round, you will have 3 minutes in which you may attempt a verbal task followed by 3 minutes in which you may attempt a numerical task.

I will now describe the verbal task.

The verbal task is a "word spotting" task. A grid of letters will be displayed on your screen. You will be able to search for and select English words appearing in the grid of letters. Valid words may appear horizontally, vertically or diagonally, and may occur forwards or backwards. Valid words are taken to be those appearing in the "WordNet" or "iSpell" English dictionaries. Both British and American spellings will be recognized, and both singular and plural word forms will be permitted. However, names and abbreviations will not be recognized. Word selections can be made using your mouse. To propose a sequence of letters as a word, you should position your cursor over the first letter of the proposed word, hold down the cursor, drag the cursor to the end of the word, and release the cursor. If the selected sequence of letters is a valid word

then the selected letters will turn yellow. Please note that valid words may partly overlap and you may use the same letters in multiple words.

In each round, your performance in the verbal task will be measured by your points score in the verbal task. Your points score in the verbal task will be number of valid word selections made within the permitted time of 3 minutes. You will not be penalized for any attempts to select invalid words.

While you are completing the verbal task your screen will display the number of the current round, the remaining time for the verbal task in the current round and your current points score in the verbal task. After the 3 minutes allocated to the verbal task have elapsed, you will be moved automatically to the numerical task.

Are there any questions?

I will now describe the numerical task. The numerical task is an adding up task. This task consists of a number of questions. For each question, you will be presented with 2 2-digit numbers and you will be asked to add the numbers together. You may enter your answer using the mouse and the number-pad which will be displayed on your screen. Please note that the number keys on your keyboard have been disabled and hence you should enter your answer using your mouse. If your answer is correct then you will be moved to the next question and new numbers for adding up will appear on your screen. If your answer is incorrect then you will remain on the same question and you will be able to enter another answer. You will only be moved to the next question once you have answered the current question correctly.

In each round, your performance in the numerical task will be measured by your points score in the numerical task. Your points score in the numerical task will be the number of questions answered correctly within the permitted time of 3 minutes. You will not be penalized for any incorrect answers. While you are completing the numerical task your screen will display the number of the current round, the remaining time for the numerical task in the current round and your current points score in the numerical task.

Are there any questions?

In each round all participants in this room will be presented with the exact same versions of the verbal and numerical tasks. This means that in any given round everybody will see the exact same grid of letters in the verbal task and will be presented with the exact same questions, in the same order, in the numerical task. The verbal and numerical tasks will vary randomly from round to round. However, the difficultly of the tasks will not vary systematically over rounds.

After the practice round there will be a pause for questions and then the 6 paying rounds will start. In each of the 6 paying rounds, after completion of the 2 tasks, there will be a 4 minute break. During this 4 minute break, you will receive feedback about your total points score in that round. Note that by "total points score in that round" I mean the number obtained by adding together your points score in the verbal task in that round and your points score in the numerical task in that round.

The feedback that you will receive during the 4 minute break after completion of the 2 tasks will be as follows:

- {Baseline group and Sub-treatments 1-3}

  Your total points score in that round will be displayed on your screen for 1 minute.

- {Sub-treatment 1}

  During the remaining 3 minutes of the break, your rank among the participants in this room in that round will be displayed on your screen.

- {Sub-treatment 2}

  During the remaining 3 minutes of the break, I will personally and privately inform each of you about your rank in that round among the participants in this room. I will do this by handing you a card indicating your rank and I will point to where your rank is written on the card.

- {Sub-treatment 3}

  During the remaining 3 minutes of the break, all participants will be asked to stand up and I will publicly inform each of you about your rank in that round among the participants in this room.

- {Sub-treatments 1-3}

  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.

Are there any questions?

The practice round will start shortly. During this practice round, you will have 3 minutes to complete the verbal task, followed by 3 minutes to complete the numerical task. After completion on the 2 tasks you will be told your total points score. Because this is a practice round, you will not be paid for this round.

Are there any questions?

Please look at your screen now and press the start button. If you have a technical problem with your computer during the practice round then please raise your hand.

[PRACTICE ROUND]

The practice round is now over. Are there any questions?

We will shortly move to the paying rounds. Recall, you will receive a payment of £2.50 per paying round. Note that your payment will not depend on your points scores in the tasks. In each round, after completion of the 2 tasks there will be a 4 minute break during which you will receive feedback about your total points score. Specifically:

[INFORMATION ABOUT TREATMENT-SPECIFIC FEEDBACK REPEATED]

Once the first paying round commences there will be no further opportunities for questions. Are there any questions at this point?

Please look at your screen now. If you have a technical problem with your computer during any of the 6 paying rounds then please raise your hand. Reminder, you must not speak at any time.

[6 PAYING ROUNDS]

The session is now complete. Please fill in the questionnaire. When you have submitted it, please raise your hand and wait to be called to a separate room where you will be paid by a laboratory assistant. Once you have been paid you are free to leave the building. Thank you for participating.

B Appendix on response to the content of rank-order feedback

B.1 Response to session characteristics

As explained in Section 3.2, the tied-groups estimator uses across-session ties from round 1 that only include subjects from the same sub-treatment. As noted in footnote 21, it is valid to use these within-sub-treatment across-session ties if, within sub-treatment, subjects do not condition first-round 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. To increase precision, we also control for sub-treatment-specific effects of a subject's own demographic characteristics on effort provision.

Formally, we estimate the following equation for effort in round 1:

$$\begin{aligned} \text{Effort}_{n,s,1} &= \sum_{k\in\mathcal{S}} \varrho_k \mathbf{1}_{\{k\}}(s) + \sum_{m=1}^{3} \phi_m \mathbf{1}_{\{m\}}(\text{Sub–treatment}) + \sum_{m=1}^{3} \theta_m X_{n,s} \mathbf{1}_{\{m\}}(\text{Sub–treatment}) \\ &+ \xi_{n,s,1} \qquad \text{for } n = 1, ..., N;\ s = 1, ..., S, \end{aligned} \tag{5}$$

where the indicator function $\mathbf{1}_{\{\cdot\}}(x)$ takes the value 1 if $x = \cdot$ and zero otherwise. In the above, $\mathcal{S}$ is the set of sessions forming the Treatment group, excluding one arbitrarily chosen session per sub-treatment. The vector $X_{n,s}$ contains a dummy variable for each possible combination of the subject's own demographic characteristics (see Section 2.4 for a description of the demographic characteristics), excluding one arbitrarily chosen reference category. Finally, $\xi_{n,s,1}$ denotes the unexplained component of first-round effort.

As noted above, our aim is to test the joint significance of the effects of the session dummies on effort provision. The first column of Table SWA.1 shows the p value for the joint null hypothesis that $\varrho_k = 0$ for all $k \in \mathcal{S}$: we comfortably fail to reject the hypothesis that subjects do not condition effort on session characteristics. Further, if subjects do not condition effort on any characteristics of the session itself, then the coefficients on the session dummies should be jointly zero for any transformation of first-round effort. Columns 2-4 report p values for three such transformations. For all three transformations, again we comfortably fail to reject the hypothesis that subjects do not condition effort on session characteristics.

	(1) $\text{Effort}_{n,s,1}$	(2) $\text{Effort}_{n,s,1} \leq 50$	(3) $\text{Effort}_{n,s,1} \leq 75$	(4) $\text{Effort}_{n,s,1} \leq 100$
p value	0.346	0.554	0.301	0.472
Mean of dependent variable	74.129	0.110	0.549	0.910
Subjects	255	255	255	255

Notes: p values are for the joint test that $\varrho_k = 0$ for all $k \in \mathcal{S}$ in (5) and use heteroskedasticity-consistent standard errors. In Column 1 the dependent variable is effort in round 1. In Columns 2, 3 and 4 the dependent variable takes the value 1 if effort in round 1 is less than or equal to 50, 75 and 100, respectively, and is zero otherwise.

Table SWA.1: Tests for response to session characteristics.

B.2 Robustness to smoothing the fully flexible rank response function

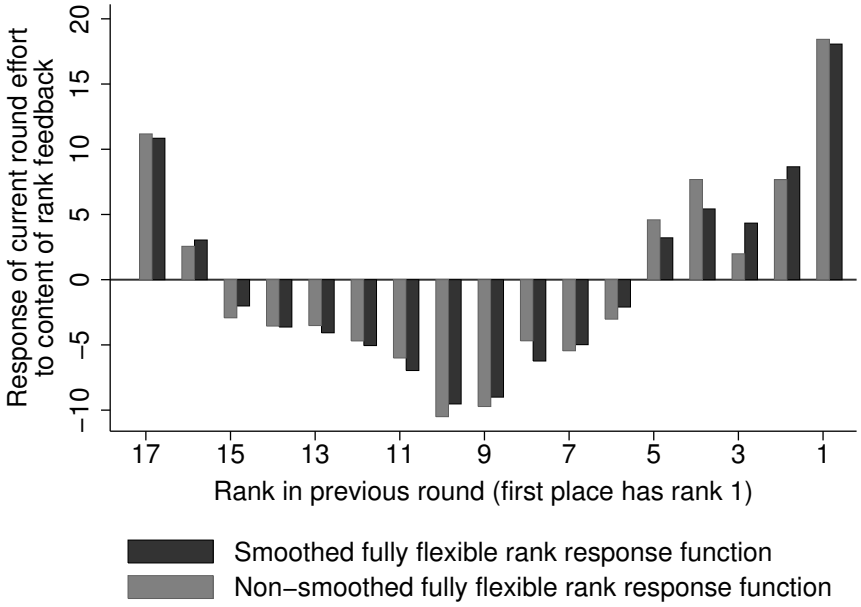


Figure SWA.1: Smoothed and non-smoothed fully flexible rank response functions.

B.3 Comparisons to standard panel data estimators

Figure SWA.2 compares the smoothed fully flexible rank response function, obtained in Section 3.4.1 using the tied-groups estimation procedure and illustrated in Figure 3, to the rank response functions obtained from standard random and fixed effects panel data estimators. To maintain comparability, we smooth the estimated rank response functions obtained from the random and fixed effects estimators using the procedure described in footnote 25 and we continue to normalize the average value of the rank response function over the $N = 17$ possible ranks to zero. Figure SWA.2 shows that the standard panel data estimators are unable to replicate the results obtained using the tied-groups estimator, indicating that the standard panel data estimators suffer from confounds discussed in Section 3.2.

We now describe formally the random and fixed effects panel data estimators that we use. First, we estimate the fully flexible rank response function (2) using the following panel regression of effort with a linear control for lagged effort and subject random effects:

$$\begin{aligned}\text{Effort}_{n,s,r} = \sum_{k=1}^{N} v_k \mathbf{1}_{\{k\}}(\text{Rank}_{n,s,r-1}) + g_r + bX_{n,s} + \psi \text{Effort}_{n,s,r-1} + \mu_{n,s} + e_{n,s,r} \\ \text{for} \quad n = 1, ..., N; \; s = 1, ..., S; \; r = 2, ..., R, \end{aligned} \tag{6}$$

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. In this model: g_r for $r = 2, ..., R$ are round fixed effects; $X_{n,s}$ consists of dummy variables for each combination of demographic characteristics; $\mu_{n,s}$ is a permanent subject-level unobservable, assumed to be a random effect; and $e_{n,s,r}$ denotes all further unobserved effort shifters.

Second, we estimate the rank response function using the following panel regression of first differenced effort with subject random effects:

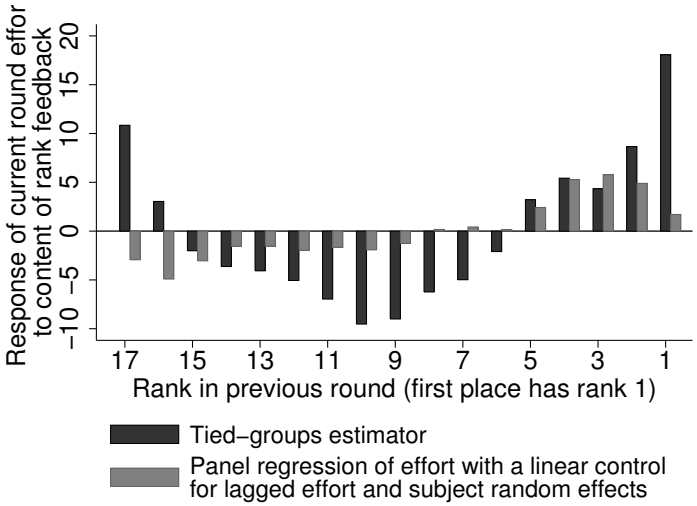
$$\begin{aligned}\text{Effort}_{n,s,r} - \text{Effort}_{n,s,r-1} = \sum_{k=1}^{N} v_k \mathbf{1}_{\{k\}}(\text{Rank}_{n,s,r-1}) + g_r + bX_{n,s} + \mu_{n,s} + e_{n,s,r} \\ \text{for} \quad n = 1, ..., N; \; s = 1, ..., S; \; r = 2, ..., R. \end{aligned} \tag{7}$$

This model was adopted by Charness et al. (2014, p.47), and is identical to that given by (6) except that here the coefficient on effort in the previous round is fixed at unity instead of being estimated.

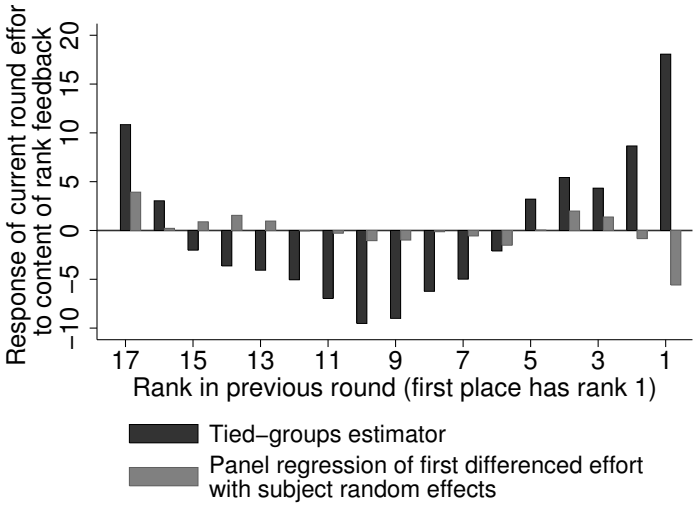
Third, we estimate the rank response function using the following panel regression of effort with subject fixed effects:

$$\begin{aligned}\text{Effort}_{n,s,r} = \sum_{k=1}^{N} v_k \mathbf{1}_{\{k\}}(\text{Rank}_{n,s,r-1}) + g_r + \mu_{n,s} + e_{n,s,r} \\ \text{for} \quad n = 1, ..., N; \; s = 1, ..., S; \; r = 2, ..., R. \end{aligned} \tag{8}$$

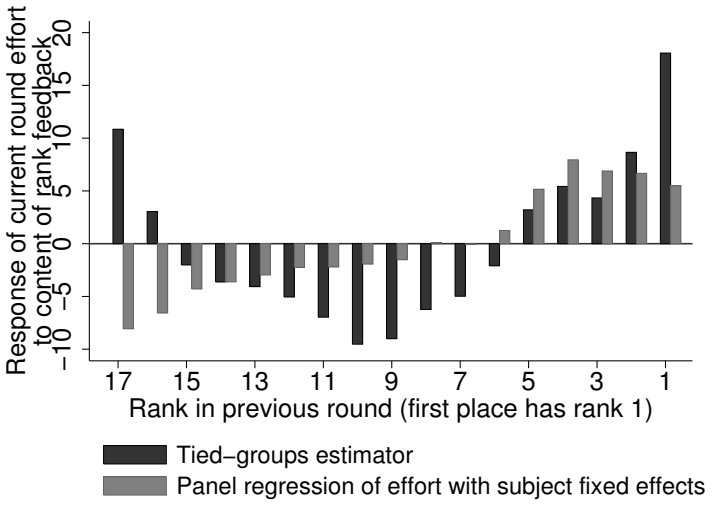
In this model: $\mathbf{1}_{\{k\}}(\text{Rank}_{n,s,r-1})$, g_r and $e_{n,s,r}$ are as in (6); and $\mu_{n,s}$ is a subject fixed effect. The subject fixed effects absorb all effects of demographic characteristics on the level of effort provision, and hence $X_{n,s}$ is absent from (8).



(a)



(b)



(c)

Figure SWA.2: Rank response functions obtained from the tied-groups estimator and from random and fixed effects panel data estimators.

C The impact of providing rank-order feedback on average effort

In this section, we start by considering how rank-order feedback affects average effort, and we then analyze the impact of rank-order feedback on the dynamics of how effort provision evolves over rounds. Figure SWA.3 shows round-by-round average effort for the Baseline group and each of the sub-treatments.

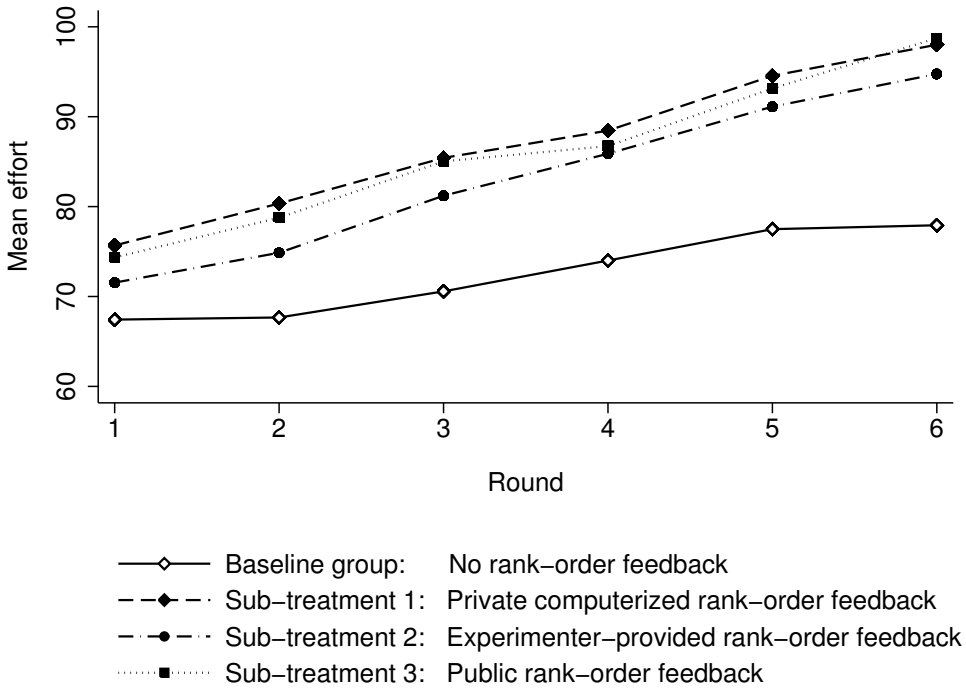


Figure SWA.3: Round-by-round mean effort by sub-treatment.

We can see from Figure SWA.3 that average effort is substantially higher when subjects are given rank-order feedback. The regressions reported in Table SWA.2 formalize this finding. Looking first at the left-hand-side column of Panel I, we see that rank-order feedback increases effort by about 20% on average, and the increase is statistically significant at the 1% level. The left-hand-side column of Panel II shows that public, private and experimenter-provided feedback all increase effort substantially and statistically significantly. Finally, the left-hand-side column of Panel III shows that the differences in effort provision according to the nature of the rank-order feedback are small and statistically insignificant. Thus, we find no evidence that public and private rank-order feedback motivate our subjects differently. The middle column of Table SWA.2 shows that the anticipation of rank-order feedback motivates our subjects to work harder even in the first round before they have received any rank-order feedback. The right-hand-side column of Table SWA.2 shows the effect of rank-order feedback after the first round, when subjects have been exposed to the feedback at least once. Table SWA.3 shows that the results in Table SWA.2 are robust to including demographic controls. Table SWA.4 provides the regression results round by round.

As described in Section 2.1, we used a flat-wage payment scheme: subjects were paid the

same fixed amount per round independent of performance. The flat wage allows us to isolate the pure effect of feedback about rank in the distribution of performance uncontaminated by any preference over rank in the distribution of earnings or a desire for monetary benefits associated with higher rank. 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 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.

Next, we consider the impact of rank-order feedback on the dynamic evolution of effort provision over rounds. We can see from Figure SWA.3 that effort increases over rounds and that this increase is particularly pronounced when subjects are given rank-order feedback. Table SWA.5, which reports the results of regressions of effort on a linear round trend, formalizes these findings (note that the estimated linear round trends are invariant to the inclusion of demographic controls, since demographics are round-invariant). The left-hand-side column of Panel I shows a statistically significant positive linear round trend for the subjects exposed to rank-order feedback: the per-round increase in effort is 5.5% of the average level of effort across all rounds for subjects in the Treatment group. The rest of Panel I shows that the round trend varies little according to the nature of the rank-order feedback.[27] Panel II shows that there is a more modest, but still statistically significant, positive round trend for subjects not exposed to rank-order feedback: the per-round increase in effort is 3.4% of the average level of effort across all rounds for subjects in the Baseline group. The left-hand-side column of Panel III (together with Panel II) shows that effort increases almost twice as fast over rounds for subjects exposed to rank-order feedback when compared to subjects that never receive such feedback, with the difference statistically significant at the 1% level. Finally, the rest of Panel III shows that effort increases faster over rounds with rank-order feedback, irrespective of whether the feedback is provided publicly or privately.

In summary, the path of effort over rounds is substantially steeper when subjects are given rank-order feedback. It seems that the desire to rank highly in the distribution of performance spurs the subjects to learn more quickly how to improve their performance over time in the real-effort tasks. Once again, a desire for self esteem rather than social esteem appears to drive behavior, since moving from private to public feedback has no influence on the evolution of effort provision over rounds.

[27] The differences are small and far from being statistically significant. Using the regressions underlying Panel I, the difference between the round trends in ST1 and ST3 has a two-sided $p = 0.623$; for ST1 and ST2, $p = 0.559$; and for ST2 and ST3, $p = 0.888$.

	All rounds	Round 1	Rounds 2–6
Panel I: Regressions of effort on rank-order feedback treatment indicator			
Treatment indicator	13.219***	6.698**	14.523***
	[0.000]	[0.023]	[0.000]
	(3.431)	(2.928)	(3.596)
Intercept	72.513***	67.431***	73.529***
	[0.000]	[0.000]	[0.000]
	(3.140)	(2.658)	(3.294)
Subject-round observations	1836	306	1530
Panel II: Regressions of effort on sub-treatment indicators			
Sub-treatment 1 (ST1) indicator	14.562***	8.245**	15.825***
	[0.000]	[0.013]	[0.000]
	(3.813)	(3.307)	(3.993)
Sub-treatment 2 (ST2) indicator	10.715***	4.113	12.035***
	[0.008]	[0.242]	[0.004]
	(3.984)	(3.505)	(4.171)
Sub-treatment 3 (ST3) indicator	13.610***	6.910**	14.951***
	[0.001]	[0.046]	[0.001]
	(4.053)	(3.445)	(4.250)
Intercept	72.513***	67.431***	73.529***
	[0.000]	[0.000]	[0.000]
	(3.142)	(2.667)	(3.296)
Subject-round observations	1836	306	1530
Panel III: Differences in average effort between sub-treatments			
ST1 minus ST3	0.952	1.335	0.875
	[0.777]	[0.649]	[0.803]
	(3.350)	(2.930)	(3.504)
ST1 minus ST2	3.847	4.132	3.790
	[0.240]	[0.169]	[0.267]
	(3.266)	(3.000)	(3.408)
ST2 minus ST3	-2.896	-2.797	-2.915
	[0.414]	[0.376]	[0.432]
	(3.543)	(3.152)	(3.706)

Notes: Panel I reports results from Ordinary Least Squares regressions of effort on an indicator for being in the Treatment group (the Baseline group forms the reference category). Panel II reports results from Ordinary Least Squares regressions of effort on indicators for being in Sub-treatment 1, Sub-treatment 2 and Sub-treatment 3 (the Baseline group forms the reference category). Panel III reports average effort differences between sub-treatments, computed from the regressions underlying Panel II. Two-sided p values are shown in square brackets and heteroskedasticity-consistent standard errors, with clustering at the subject level to account for within-subject non-independence across rounds, are shown in round brackets. *, ** and *** denote significance at the 10%, 5% and 1% levels (2-sided tests).

Table SWA.2: Effect of rank-order feedback on effort provision.

	All rounds	Round 1	Rounds 2–6
Panel I: Regressions of effort on rank-order feedback treatment indicator			
Treatment indicator	13.348***	6.521**	14.714***
	[0.000]	[0.041]	[0.000]
	(3.749)	(3.176)	(3.947)
Intercept	72.723***	67.571***	73.753***
	[0.000]	[0.000]	[0.000]
	(4.352)	(3.815)	(4.598)
Subject-round observations	1836	306	1530
Panel II: Regressions of effort on sub-treatment indicators			
Sub-treatment 1 (ST1) indicator	13.988***	7.516**	15.282***
	[0.001]	[0.038]	[0.000]
	(4.117)	(3.603)	(4.320)
Sub-treatment 2 (ST2) indicator	10.421**	4.040	11.698***
	[0.012]	[0.275]	[0.007]
	(4.117)	(3.695)	(4.319)
Sub-treatment 3 (ST3) indicator	15.027***	7.395**	16.554***
	[0.001]	[0.045]	[0.000]
	(4.393)	(3.667)	(4.636)
Intercept	72.913***	67.584***	73.979***
	[0.000]	[0.000]	[0.000]
	(4.278)	(3.784)	(4.523)
Subject-round observations	1836	306	1530
Panel III: Differences in average effort between sub-treatments			
ST1 minus ST3	-1.040	0.120	-1.272
	[0.760]	[0.969]	[0.721]
	(3.406)	(3.062)	(3.561)
ST1 minus ST2	3.566	3.475	3.584
	[0.268]	[0.270]	[0.284]
	(3.211)	(3.147)	(3.337)
ST2 minus ST3	-4.606	-3.355	-4.856
	[0.197]	[0.304]	[0.195]
	(3.563)	(3.256)	(3.735)

Notes: All results were obtained as described in the notes to Table SWA.2, except that the regressions additionally include a dummy variable for each combination of demographic characteristics with the reference category being female, born in the United Kingdom, aged below 22 and studying a STEMM subject (see Section 2.4 for a description of the demographic characteristics). Two-sided p values are shown in square brackets and heteroskedasticity-consistent standard errors, with clustering at the subject level to account for within-subject non-independence across rounds, are shown in round brackets. *, ** and *** denote significance at the 10%, 5% and 1% levels (2-sided tests).

Table SWA.3: Robustness of results in Table SWA.2 to including demographic controls.

	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6
Panel I: Regressions of effort on rank-order feedback treatment indicator						
Treatment indicator	6.698**	10.690***	13.600***	13.192***	15.675***	19.459***
	[0.023]	[0.002]	[0.000]	[0.000]	[0.000]	[0.000]
	(2.928)	(3.485)	(3.832)	(3.397)	(3.948)	(4.273)
Intercept	67.431***	67.667***	70.569***	74.000***	77.490***	77.922***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(2.658)	(3.153)	(3.527)	(3.088)	(3.617)	(3.889)
Subject-round observations	306	306	306	306	306	306
Panel II: Regressions of effort on sub-treatment indicators						
Sub-treatment 1 (ST1) indicator	8.245**	12.676***	14.853***	14.451***	17.059***	20.088***
	[0.013]	[0.001]	[0.001]	[0.000]	[0.000]	[0.000]
	(3.307)	(3.938)	(4.261)	(3.821)	(4.369)	(4.862)
Sub-treatment 2 (ST2) indicator	4.113	7.201*	10.637**	11.882***	13.627***	16.828***
	[0.242]	[0.083]	[0.015]	[0.002]	[0.003]	[0.001]
	(3.505)	(4.140)	(4.341)	(3.885)	(4.592)	(5.066)
Sub-treatment 3 (ST3) indicator	6.910**	11.098***	14.467***	12.729***	15.651***	20.808***
	[0.046]	[0.008]	[0.002]	[0.002]	[0.001]	[0.000]
	(3.445)	(4.164)	(4.533)	(4.124)	(4.707)	(4.974)
Intercept	67.431***	67.667***	70.569***	74.000***	77.490***	77.922***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(2.667)	(3.164)	(3.539)	(3.099)	(3.629)	(3.902)
Subject-round observations	306	306	306	306	306	306
Panel III: Differences in average effort between sub-treatments						
ST1 minus ST3	1.335	1.578	0.386	1.722	1.408	-0.720
	[0.649]	[0.660]	[0.917]	[0.625]	[0.716]	[0.865]
	(2.930)	(3.582)	(3.696)	(3.522)	(3.861)	(4.234)
ST1 minus ST2	4.132	5.475	4.216	2.569	3.431	3.260
	[0.169]	[0.124]	[0.224]	[0.428]	[0.357]	[0.453]
	(3.000)	(3.554)	(3.458)	(3.239)	(3.719)	(4.341)
ST2 minus ST3	-2.797	-3.897	-3.829	-0.847	-2.024	-3.979
	[0.376]	[0.306]	[0.313]	[0.814]	[0.623]	[0.374]
	(3.152)	(3.802)	(3.787)	(3.591)	(4.111)	(4.466)

Notes: All results were obtained as described in the notes to Table SWA.2. Two-sided p values are shown in square brackets and heteroskedasticity-consistent standard errors are shown in round brackets. *, ** and *** denote significance at the 10%, 5% and 1% levels (2-sided tests).

Table SWA.4: Round-by-round effect of rank-order feedback on effort provision.

Panel I: Regressions of effort on linear round trend variable for Treatment group and sub-treatments				
	Treatment group (T)	Sub-treatments ST1	ST2	ST3
Linear round trend	4.677***	4.495***	4.842***	4.765***
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.233)	(0.421)	(0.419)	(0.353)
Intercept	74.039***	75.838***	71.124***	74.212***
	[0.000]	[0.000]	[0.000]	[0.000]
	(1.306)	(2.117)	(2.287)	(2.358)
Subject-round observations	1530	612	408	510
Panel II: Regression of effort on linear round trend variable for Baseline group (B)				
Linear round trend	2.439***			
	[0.000]			
	(0.450)			
Intercept	66.416***			
	[0.000]			
	(2.804)			
Subject-round observations	306			
Panel III: Differences between Treatment group (T) or sub-treatment and Baseline group (B)				
	T minus B	ST1 minus B	ST2 minus B	ST3 minus B
Linear round trend	2.239***	2.056***	2.403***	2.326***
	[0.000]	[0.001]	[0.000]	[0.000]
	(0.503)	(0.614)	(0.612)	(0.569)
Intercept	7.623**	9.422***	4.708	7.795**
	[0.014]	[0.008]	[0.194]	[0.034]
	(3.070)	(3.498)	(3.604)	(3.649)

Notes: The results in Panels I and II were obtained from Ordinary Least Squares regressions of effort on a linear round trend variable that takes the value of 0 in round 1, and increases linearly to the value of 5 in round 6. The results reported in Panel III were computed from the regressions underlying Panels I and II. Two-sided p values are shown in square brackets and heteroskedasticity-consistent standard errors, with clustering at the subject level to account for within-subject non-independence across rounds, are shown in round brackets. *, ** and *** denote significance at the 10%, 5% and 1% levels (2-sided tests).

Table SWA.5: Trends in effort provision over rounds.

D The effect of rank-order feedback according to the competitiveness of subjects

At the end of each experimental session, we asked subjects to report their degree of competitiveness. The categories were as follows: (a) "I am strongly competitive. I am always interested in how my performance compares to the performance of others."; (b) "I am moderately competitive. I often take interest in how my performance compares to the performance of others."; and (c) "I am not competitive at all. I do not compare my performance to the performance of others." Fewer than 10% of subjects reported themselves to be not competitive at all. Thus, for the purposes of the analysis, we merge the second and third categories, giving a simple binary categorization of subjects as: (i) 'strongly competitive'; or (ii) 'not strongly competitive'.

Since our measure of competitiveness is based on a post-experimental self-report, we need to check that the proportion of subjects who reported themselves to be strongly competitive does not depend on whether the subjects were exposed to rank-order feedback during the experiment. We find that 33.7% of subjects in the Treatment group reported themselves to be strongly competitive, compared to 34.0% of subjects in the Baseline group.

The left-hand-side and middle columns of Table SWA.6 replicate the regression reported in the left-hand-side column of Panel I in Table SWA.2 for, respectively, strongly competitive subjects and not strongly competitive subjects. Looking first at the left-hand-side column of Table SWA.6, we see that for strongly competitive subjects rank-order feedback increases average effort across the six rounds substantially and statistically significantly. The middle column shows that rank-order feedback also increases average effort substantially and statistically significantly for the subjects that are not strongly competitive. The right-hand-side column shows that the effect of rank-order feedback on average effort varies little according to subject competitiveness: the difference in the effect of rank-order feedback is small and far from being statistically significant. The right-hand-side column further shows that when subjects are not exposed to rank-order feedback, strongly competitive subjects work about 20% harder than those that are not strongly competitive, with the difference statistically significant at the 5% level.

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, recall that 'strongly competitive' was equated 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.

	Strongly competitive	Not strongly competitive	Difference
Treatment indicator	12.642**	13.251***	-0.609
	[0.032]	[0.001]	[0.932]
	(5.823)	(4.090)	(7.099)
Intercept	81.618***	68.141***	13.476**
	[0.000]	[0.000]	[0.038]
	(5.264)	(3.781)	(6.467)
Subject-round observations	618	1212	1830

Notes: All results were obtained from Ordinary Least Squares regressions of effort on an indicator for being in the Treatment group (the Baseline group forms the reference category). One subject did not report her competitiveness. Two-sided p values are shown in square brackets and heteroskedasticity-consistent standard errors, with clustering at the subject level to account for within-subject non-independence across rounds, are shown in round brackets. *, ** and *** denote significance at the 10%, 5% and 1% levels (2-sided tests).

Table SWA.6: Effect of rank-order feedback on effort provision according to subject competitiveness.

Next, we consider how competitiveness interacts with the dynamic evolution of effort provision over rounds. The left-hand-side and middle columns of Table SWA.7 replicate the regressions reported in the left-hand-side column of Table SWA.5 for, respectively, strongly competitive subjects and not strongly competitive subjects. The left-hand-side column of Table SWA.7 shows that the qualitative findings reported in Table SWA.5 extend when we restrict attention to strongly competitive subjects: effort increases over rounds with and without rank-order feedback, but the path of effort over rounds is steeper when subjects are given rank-order feedback. The middle column of Table SWA.7 shows that the same qualitative findings also hold for subjects that are not strongly competitive. The right-hand-side column of Panel I in Table SWA.7 shows that when rank-order feedback is provided, effort increases faster over rounds for strongly competitive subjects compared to the rate of increase for subjects that are not strongly competitive. The right-hand-side column of Panel II shows that the same is true when rank-order feedback is not provided. Finally, the right-hand-side column of Panel III shows that the impact of rank-order feedback on the evolution of effort provision is a little weaker for strongly competitive subjects, but the difference is not close to being statistically significant.

In summary, the path of effort over rounds is substantially steeper for strongly competitive subjects, while both strongly competitive and not strongly competitive subjects increase their effort faster over rounds when rank-order feedback is provided. As discussed above, the more competitive subjects might be 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: this extra motivation might also spur the strongly competitive subjects to learn more quickly how to improve their performance over time.

Finally, we consider whether competitiveness interacts with how subjects respond to the content of rank-order feedback. Using the same methodology as for the demographic characteristics (Section 3.4.3), we find no statistically significant differences according to competitiveness: a test of the joint significance of the interactions between the previous rank variables and an indicator for being strongly competitive gives $p = 0.631$. Thus we find that both strongly competitive

and not strongly competitive subjects exhibit first-place loving and last-place loathing.

The existing literature on how competitiveness interacts with the provision of relative-performance feedback is sparse. Using a different measure of competitiveness that emphasizes the desire to win rather than performance comparisons more generally, Gerhards and Siemer (2014) find that more competitive subjects work harder only in one of two tasks, and then only with feedback about the winner of an award. Girard and Hett (2013) measured competitiveness by the willingness to enter a tournament rather than be paid piece-rate, and find that more competitive teams work less hard and respond more strongly to interim rank information during the course of a between-team competition.

	Strongly competitive	Not strongly competitive	Difference
Panel I: Treatment group			
Linear round trend	5.597***	4.209***	1.387***
	[0.000]	[0.000]	[0.004]
	(0.392)	(0.284)	(0.483)
Intercept	80.268***	70.869***	9.399***
	[0.000]	[0.000]	[0.001]
	(2.296)	(1.536)	(2.754)
Subject-round observations	516	1014	1530
Panel II: Baseline group			
Linear round trend	4.092***	1.771***	2.321**
	[0.000]	[0.001]	[0.014]
	(0.786)	(0.487)	(0.911)
Intercept	71.387***	63.713***	7.674
	[0.000]	[0.000]	[0.200]
	(4.847)	(3.514)	(5.903)
Subject-round observations	102	198	300
Panel III: Differences between Treatment group and Baseline group			
Linear round trend	1.504*	2.438***	-0.934
	[0.083]	[0.000]	[0.362]
	(0.859)	(0.558)	(1.022)
Intercept	8.881*	7.156*	1.725
	[0.093]	[0.060]	[0.789]
	(5.243)	(3.790)	(6.453)

Notes: All results were obtained from Ordinary Least Squares regressions of effort on a linear round trend variable that takes the value of 0 in round 1, and increases linearly to the value of 5 in round 6. The results reported in Panel III were computed from the regressions underlying Panels I and II. One subject did not report her competitiveness. Two-sided p values are shown in square brackets and heteroskedasticity-consistent standard errors, with clustering at the subject level to account for within-subject non-independence across rounds, are shown in round brackets. *, ** and *** denote significance at the 10%, 5% and 1% levels (2-sided tests).

Table SWA.7: Trends in effort provision over rounds according to subject competitiveness.