Feedback, Self-Esteem and Performance in Organizations*

Camelia M. Kuhnen†

Agnieszka Tymula‡

Forthcoming in Management Science

Abstract

We examine whether private feedback about relative performance can mitigate moral hazard in competitive environments by modifying the agents' self-esteem. In our experimental setting people work harder and expect to rank better when told they may learn their ranking, relative to cases when feedback will not be provided. Individuals who ranked better than expected decrease output but expect a better rank in the future, while those who ranked worse than expected increase output but lower their future rank expectations. Feedback helps create a ratcheting effect in productivity, mainly due to the fight for dominance at the top of the rank hierarchy. Our findings suggest that organizations can improve employee productivity by changing the likelihood of feedback, the reference group used to calculate relative performance, and the informativeness of the feedback message.

*We thank Terrance Odean (the editor), an anonymous referee, as well as Pierpaolo Battigalli, Stefano DellaVigna, Shimon Kogan, David Myatt, Jonathan Parker, Imran Rasul, Paola Sapienza, Nora Szech and participants at the 2008 ESA European Meeting, the 2008 La Pietra-Mondragone Theory Workshop, the 2008 International ESA Conference, the 2008 NBER Behavioral Finance Meeting, the 2009 IZA Workshop on Genes, Brains and Labor Markets, and seminar participants at New York University, Northwestern University and U.C. Berkeley for valuable comments and discussion. This paper was previously circulated as "Rank expectations, Feedback and Social Hierarchies".

†Department of Finance, Kellogg School of Management, Northwestern University, 2001 Sheridan Rd., Evanston, IL 60208-2001, c-kuhnen@kellogg.northwestern.edu. Kuhnen gratefully acknowledges financial support from the Zell Center for Risk Research.

‡New York University, Center for Neural Science, 4 Washington Place, New York, NY 10003, tymula@nyu.edu.

1 Introduction

Performance appraisals, such as the 360-degree feedback process, have become common features of the workplace over the last two decades (Prewitt (2007)). While the goal of these appraisals is to encourage employee development and improve performance, empirical evidence suggests that providing feedback does not always lead to better outcomes in organizations, as it can negatively impact the employees' self-esteem (Kluger & DeNisi (1996), Smither, London, & Reilly (2005)). Recently, companies such as GE, Yahoo and Whirpool have changed aspects of the appraisal process such as the frequency of feedback, the labels provided for particular performance levels (e.g., "successful" versus "middle 50%") and the benchmarks used to define performance (e.g., absolute criteria versus relative rankings), indicating that it is still unclear what constitutes effective feedback.[1] To shed light on this issue, in this paper we examine theoretically and empirically how feedback and self-esteem considerations interact and influence employee performance.

Self-esteem has long been thought of in the psychology literature as a strong motivator of human behavior (Maslow (1943), McClelland, Atkinson, Clark, & Lowell (1953)). People derive utility from thinking of themselves as good, productive or valuable according to social criteria, and their actions are shaped by the desire to maintain high levels of self-esteem. Recently, this concept has been introduced in theoretical models of economic choice in non-competitive settings as "ego utility" (Benabou & Tirole (2002), Koszegi (2006)). However, ego utility may also affect strategic interactions, where self-esteem is determined by an individual's perception of his relative standing among peers, and not necessarily by beliefs about absolute measures of his ability. In such settings, as in the workplace, the existence of relative performance feedback implies that ego utility is influenced not only by an individual's own actions, but also by those of other players. While these strategic considerations are similar to those studied in the tournaments literature, existing theory models do not capture the behavior of agents in settings where the benefit of being the most productive player is simply ego utility, or self-esteem. Moreover, there are no empirical or experimental accounts of behavior in such settings. We seek to address these gaps in the literature.

Specifically, our goal is to understand the impact of ego utility on productivity in competitive settings where participants receive private feedback about their relative standing. The theoretical framework we develop and the experimental results imply that private feedback about relative ranking has ex-ante and ex-post effects on the productivity of workers and on the dynamics of social hierarchies. As predicted by the model, in our experimental setting agents work harder and expect to rank better when they are told they may learn their ranking, relative to cases when they are told feedback will not be provided. After receiving feedback, individuals who learn that they have ranked better than expected decrease their output but expect an even better rank in the future, while those who were told they ranked worse than expected increase their output and at the same time lower their rank expectations going forward. These effects are stronger in earlier rounds of the task, while subjects learn how they compare to their peers in terms of output produced. This rank hierarchy is established early on, and it remains relatively stable later in the task. Private information regarding relative standing helps create a ratcheting effect in the group's

[1] See the Business Week article "The Struggle to Measure Performance" (January 9, 2006) and the Wall Street Journal article "For Relevance, Firms Revamp Worker Reviews" (July 17, 2006).

average output. This increase in output over time is mainly due to the fight for dominance at the top of the hierarchy. Moreover, increasing the heterogeneity in the ability of peers leads to lower output from low ability individuals, but has no impact on the output of high ability workers.

In the model and the experimental setting we isolate the ego utility effect from other reasons why feedback about rank may change behavior. For instance, feedback may influence productivity if compensation is performance-based, since people may care more about their relative, rather then objective level of wealth (Clark & Oswald (1996), Easterlin (1995), Luttmer (2005)). Feedback may also change behavior if it provides information about the nature of the project (Seta (1982), Bandura (1986)). Moreover, if feedback is public, and thus the relative ranking is common knowledge among participants, peer monitoring or concerns for social status and reputation may influence the behavior of participants (Kandel & Lazear (1992), Falk & Ichino (2006), Mas & Moretti (2009)). Therefore, to minimize the influence of these other channels through which relative rank information may impact actions, we use a setting where participants receive a flat wage, the task that they work on does not involve changes in strategy, and feedback is private and anonymous.

Our premise that people's self-esteem depends on their relative standing among peers is supported by a large body of evidence. Research from social psychology shows that when effort is unobservable people work harder when they are provided with a social comparison criterion, for example with the average productivity of past participants (Szymanski & Harkins (1987), White, Kjelgaard, & Harkins (1995)), suggesting that people dislike falling behind the average. Moreover, Falk, Huffman, & Sunde (2006) show that low productivity subjects are more likely than high productivity ones to choose not to learn their rank in the group at the end of an experimental task, while Burks, Carpenter, Goette, & Rustichini (2010) find that individuals who are confident that they have high ability are interested in learning information about their relative performance. This evidence is consistent with the idea that utility is influenced by learning about one's relative ranking. Furthermore, recent neuroeconomics evidence shows that the mere fact of outperforming other workers generates activation in the brain's reward centers, and therefore is a pleasant experience (Dohmen, Falk, Fliessbach, Sunde, & Weber (2011)).

This paper contributes to the two strands of economics literature focused on ego utility, and, respectively, on feedback provision. Benabou & Tirole (2002) argue that self-confidence is valuable because it enhances the motivation to act, and investigate a variety of strategies people may use to enhance their self-image. They show that people may handicap their performance by exerting low effort and use self-deception through selective memory in order to maintain high self-perception about their ability. This keeps them motivated to undertake profitable endeavors in the future. Weinberg (1999) and Bandiera, Larcinese, & Rasul (2009) treat self-esteem as a consumption good by assuming that an individual's utility is increasing with his perception of his own ability. Koszegi (2006) also incorporates perceptions about one's ability in the utility function and shows how ego-motivated individuals manage their self-image and how this later influences their effort choice. Ertac (2005) and Ederer (2010) study optimal feedback provision in settings where information about relative performance is used to learn about one's own ability but has no effect on the utility function. In these two papers there are no self-esteem considerations and behavior changes only when new information about output becomes available. As a result such settings preclude the existence of the ex-ante feedback effects which we study in our model and also document empirically.

Since prior models of ego-utility do not account for the possibility that in settings such as the workplace one's self-esteem is not shaped in isolation but is also influenced by the actions of others, they have ambiguous implications for the effect of relative rank information on behavior. When feedback is likely to be provided, ex-ante concerns for self-image may increase effort, since agents seek to learn that they rank high, as in Weinberg (1999). However, the prospect of receiving feedback may also lead to lower ex-ante effort, because agents with positive beliefs about themselves will avoid competing, in order to preserve their self-esteem, as in Koszegi (2006).[2] Ex-post effects of feedback are also difficult to predict based on existent theories. For instance, after receiving bad feedback about relative performance, people with self-image concerns could use deception strategies, as in Benabou & Tirole (2002), in order to discard this information or interpret it to their advantage. They may give up competing if the perceived chances of winning in the future are low, or may engage in the task again because it is the only way to regain self-confidence, as in Koszegi (2006). Complementing these prior models, our theoretical framework applies to multi-agent settings and makes clear predictions about the ex-ante and ex-post effects of feedback.

In our model we focus on the consumption value of self-esteem and assume that the agents' utility is increasing in the level of their own output and decreasing in the output of their peers, to capture the fact that self-esteem is often determined by relative performance comparisons (Szymanski & Harkins (1987)). Unlike in models where ego utility is determined by absolute performance or ability, in our setting we can account for the possibility that the same individual may behave differently depending on whether he is surrounded by less or more productive peers. We also allow agents to endogenously set the reference standard against which they compare themselves, to account for the possibility that goals are chosen by each agent as a function of their beliefs and environment (White et al. (1995)). In the model we show that goals (or the importance that people place on relative comparisons) indeed depend endogenously on the beliefs about one's relative ability in the reference group, and this dependence has implications for the agent's choice of effort. We are also able to characterize the optimal feedback frequency and show that it depends on the agent's relative standing in the peer group. Therefore the novelty of the model is that it captures the effect of feedback on purely ego-motivated individuals in a competitive setting. Furthermore, the theoretical framework we develop – although quite stylized for tractability reasons – is helpful in understanding the experimental results that we document in the paper.

Related to the work on self-esteem is a large literature on status and peer effects. People care about social status as defined by their relative income (Frank (1984), Frank (1985)), they value public recognition independently of any monetary consequence (Delfgaauw, Dur, Sol, & Verbeke (2009)) and are willing to trade off material gains to obtain it (Huberman, Loch, & Onculer (2004)). The quest for status has labor market implications, for instance regarding incentives and promotion schemes, or job search and sorting (Cowen & Glazer (2007), Clark, Masclet, & Villeval (2006) and Neckermann & Frey (2008)). Peer monitoring has also been proposed as an effective incentive

[2]Contrary to this prediction, the evidence in Burks et al. (2010) indicates that people with positive beliefs about their own ability actually seek to learn information about their relative performance, suggesting that relative performance, and not the level of ability per se, enters the utility function. Our model and experimental data point in the same direction that relative performance matters for the agents' utility.

mechanism (Major, Testa, & Bylsma (1991), Kandel & Lazear (1992)), even when output does not have an impact on monetary payoffs (Falk & Ichino (2006), Mas & Moretti (2009)).

In contrast to the streams of work on status seeking and peer monitoring effects, our focus is on the internal drive of individuals to rank well relative to others, and not on people's need for public recognition or reputation among peers. In line with prior evidence, we assume that people enjoy performing well relative to others even in situations when performance is private information, or when there are no future consequences via reputation or career concerns channels. A related driver of behavior to the one studied here is intrinsic motivation: people enjoy effortful endeavors, even in the absence of incentive pay, because completing such endeavors generates a sense of personal growth and fulfillment (e.g. Deci (1975)). Benabou & Tirole (2003) formalize the concepts of intrinsic and extrinsic motivation and show under which conditions the latter will "crowd out" or "crowd in" the former. There is extensive empirical evidence that external intervention (for example output-based pay or monitoring) crowds out intrinsic motivation and undermines productivity (Frey (1997), Deci, Koestner, & Ryan (1999), Frey & Jegen (2001)). For instance, Gneezy & Rustichini (2000) show that piece rates lead to increased performance only if they are substantial and even piece rates as high as 10% may lead to a decrease in output as compared to a situation where no incentive pay is used. Since extrinsic motivators often turn out to have detrimental effects, finding the optimal level of incentive pay that would improve rather than impair productivity is not trivial. We are therefore considering an alternative incentive device - private information about one's relative performance in the group - that can potentially reinforce intrinsic motivation in ego-driven individuals.

An important question left for future work is whether in environments where monetary incentives are strong enough to actually motivate people to work hard, they may crowd out the effects of feedback driven by self-esteem that we demonstrate here in a flat-wage environment. The evidence so far is mixed. On the one hand, Eriksson, Poulsen, & Villeval (2009) find in an experimental setting that releasing information about relative performance does not significantly influence the subjects' average effort when they face piece-rate or tournament pay. On the other hand, Blanes i Vidal & Nossol (2009) and Azmat & Iriberri (2009) find that when piece-rates incentives are used to induce effort, providing individuals with relative performance feedback increases productivity.

Our results suggest that in settings where monetary incentives are weak or non-existent, moral hazard can be mitigated by optimally providing feedback to agents regarding their relative performance. Ego utility, or self-esteem can be used as a motivator for productivity. In light of these findings, it is possible that by changing the reference peer group, the timing and the recipients of feedback, organizations can benefit from the dynamics of social hierarchy effects on productivity.

2 Theoretical framework and implications

In our model two agents, i and j, work on similar tasks. For agent i (and similarly for j) output is given by $y_i = a_i + e_i + \widetilde{\varepsilon}_i$, where a_i represents the agent's innate ability level, e_i is the agent's effort and $\widetilde{\varepsilon}_i \sim N(0, \sigma^2)$ is an exogenous shock independently and identically distributed across agents. The agent does not know his own ability, nor the ability of the other worker.

Each agent's utility is increasing in his own output and decreasing in the output of the opponent, since people enjoy performing better relative to others (Szymanski & Harkins (1987)). Agents work for a fixed wage. Similar to Falk & Knell (2004), we assume that agents can choose how much to care about the feedback about the opponent's output, and therefore about their relative rank. This choice is captured by the variable $s_i \geq 0$, which we call the agent's standard.[3]

At the end of the working period each agent knows how much he produced and he may also learn how much the other agent produced. In the beginning of the period each agent is told the probability with which he will get information about the output of his opponent. We will denote this probability as p for agent i, and q for agent j. Agent i knows p but not q, and the opposite is true for agent j. It is common knowledge that random variables p and q are independent, and have a probability distribution function f such that, unconditionally, $E(p) = E(q) = \frac{1}{2}$.

We assume that the agent's utility after he observes only his own output is equal to the level of his output y_i. If he also observes the output of his opponent his utility is equal to $y_i - y_j \ln\left(\frac{k}{k-s_i}\right)$, where $k > s_i$ is a parameter. Since expression $\ln\left(\frac{k}{k-s_i}\right)$ is increasing in s_i, this means that all else equal, the higher a standard s_i the agent sets, the more he needs to produce to achieve a given level of utility from comparing his output to that of the competitor. If the standard s_i played no further role in our model, it would be optimal for the agent to always choose the lowest standard possible. In the psychology literature, such behavior is attributed to a "self-enhancement" motive: in order to feel good about themselves, people compare themselves downward, i.e., to those who are less productive. At the same time, people who set high standards have been shown to perform better. We capture this "self-improvement" motive by using a cost of effort function that is decreasing in the standard s_i, which implies that a given level of effort is less costly when one works on ambitious and demanding tasks. Concretely, we assume that agent i experiences the following cost of effort while working: $c_i(a_i, e_i) = \alpha - \gamma a_i \ln(\beta - e_i + ps_i)$ where $\alpha, \beta > 0$ and $0 < \gamma < 1$ are parameters. These assumptions ensure that effort is less costly if agents set a higher standard s_i for themselves. Further, because $\gamma > 0$, effort is less costly for a more able worker, that is, being better skilled to do a task makes the job more enjoyable or less stressful.[4]

Therefore, an agent who does not know his own or his opponent's ability and expects to get feedback about the opponent with probability p must choose effort and standard levels to maximize the following expected utility function: $E_i(u_i) = (1-p) E_i(y_i) + p\left(E_i(y_i) - E_i(y_j) \ln\left(\frac{k}{k-s_i}\right)\right) - \alpha + \gamma E_i(a_i) \ln(\beta - e_i + ps_i)$. In the Appendix we solve for the general equilibrium of the model and prove several propositions, highlighted below:

Proposition 1. If the agent believes that his ability is relatively high (low) compared to the ability of the competitor then he will produce more (less) output and expect better (worse) relative performance when the likelihood of feedback increases.

Proposition 2. After receiving good (bad) feedback about one's own ability, i.e., after the agent learns that he is better (less) skilled than he expected, the agent's output will decrease (increase) if $p < (\geq) \frac{3\gamma}{2(1-\gamma)}$ (sufficient condition is that $\gamma > (\leq) \frac{2}{5}$).

Proposition 3. If the agent learns that his competitor is better (less) skilled than he expected,

[3]This variable can be interpreted as a measure of how much the individual would be hurt by an increase in the output of the other player, how frequently he decides to compare himself to the other, or how ambitious and motivated he is. The higher s_i is, the more ambitious is the agent's goal.

[4]The dependence of the cost function on p is purely technical. It ensures that when $p = 0$, the standard s_i does not change the cost function, since then the agent can not compare himself with the competitor.

he will decrease (increase) his future output.[5]

Proposition 4. When the agent's beliefs about relative performance are revised upwards (downwards), he expects better (worse) relative performance in the future.

Proposition 5. When the agent's beliefs about relative ability are revised upwards (downwards), he will choose a higher (lower) standard.

Proposition 6. If abilities a_i and a_j and feedback probabilities p and q are common knowledge, then for a given q if agent i is good enough relative to agent j (that is, if $a_i \geq \frac{1}{q}\left(a_j - \frac{(k-\beta)(1-q)}{(1-\gamma)}\right)$) it is optimal for the principal to increase the frequency of feedback for the high ability worker i.

Therefore, the model predicts that the feedback policy can influence productivity and beliefs before and after rank information is revealed to the agents. Agents with different likelihoods of receiving information about their opponents' output will (all else equal) expect to rank differently and will produce different levels of output. Agents who initially do not know their relative position in the group adjust effort and beliefs about future rank as they change their perceptions of relative ability. Different patterns in behavior and beliefs will occur after good and bad feedback, that is, after the subject learns that he ranked better or worse than he expected. We use these theoretical implications to guide our interpretation of the patterns observed in the experimental data.

3 Experimental design

We use a simple task to understand the role of private feedback regarding relative rank on productivity and to examine whether self-esteem considerations are important. In the experiment we ask subjects to solve multiplication problems (i.e., to multiply one-digit numbers by two-digit numbers) during several identically structured rounds. We use this task for several reasons. First, no previous task knowledge is required and it is easy to explain. Second, task learning effects, which we would like to avoid, should be minimal. In other words, we expect that participants know how to solve multiplication problems before they came to the lab, and their ability to solve these problems does not change much during the duration of the task. Third, the score on this task depends on the subjects' ability as well as on their effort choice. Therefore, different subjects will end up with different scores, which will lead to dispersed rankings. Fourth, the subjects' ranks depend not only on their own (possibly unknown to them) abilities but also on the unknown skills and effort decisions of other participants. As a result, we are likely to find situations where the subjects' expectations are not confirmed by the received feedback. This allows us to study how this mismatch between expectations and reality affects future expectations and productivity.

In the task it is necessary to control for the difficulty level of the multiplication problems. We therefore follow Cromer (1974) and generate 206 multiplication problems of the same difficulty level (for example, $89*4$, $76*9$ or $73*8$). Problems are presented to each subject on a computer screen. Each time the subject solves the multiplication correctly one point is added to his score and the next problem is presented. If the subject provides a wrong answer, the score remains unchanged

[5]Propositions 2 and 3 imply that an agent will change his output based on feedback about his own or his opponent's ability. An agent who learned that his own ability is higher or the ability of his opponent is lower will increase his output if $p \geq \frac{3\gamma}{2(1-\gamma)}$. For $p < \frac{3\gamma}{2(1-\gamma)}$, the direction of change in the output depends on the strength of the effect of own ability relative to that of the competitor's ability.

and he is asked to solve the same problem again until answered correctly. By not allowing subjects to move on to the next question unless the previous one is solved, we avoid a situation where participants may strategically skip difficult problems looking for easy ones.

The experiment consists of 18 identically-structured rounds. In each round and for each subject, three feedback conditions are possible. The conditions differ with respect to the probability with which the subject receives feedback about his relative rank at the end of the round. This probability is either 0, 0.5 or 1. We refer to these as the "No", "Maybe" or "Sure" treatments, respectively. The feedback condition is determined randomly and independently for each subject at the beginning of every round. Therefore, in the same round different subjects face different feedback conditions. Importantly, each subject knows only his own feedback probability.

Figure 1 shows the sequence of events in each round. First, subjects are told which feedback condition they are in. This allows us to study the ex-ante effect of feedback probability on rank, expected rank, and output. Then subjects are asked to report their expected rank in that round.[6] Following this, participants have 90 seconds to work on multiplication problems. For each subject, their score is displayed on the screen throughout the round and is updated after every correct answer (the score is reset to zero at the beginning of every round). Therefore subjects are always informed about their own output, independent of their feedback condition. After the 90 seconds pass, subjects are asked to asses how much effort they have put into the task in the current round. Answers are provided using a six point scale ranging from "no effort at all" to "a lot of effort".[7]

In the final stage of each round, that lasts for fifteen seconds, each subject either sees the performance ranking or not, depending on the feedback condition they have been assigned to in that round. The ranking is determined by the current period scores of all subjects in the group. The subject who solves the highest number of problems ranks as number one, the person who solves the second highest number of problems ranks as number two, and so forth. Participants who solve the same number of problems in a round receive the same rank. Each subject can see the scores and ranks of all the participants but he can identify *only* his rank and score. Therefore each subject knows that nobody else can associate his identity to his rank and score.

Given our design, the baseline condition in the experiment refers to trials where subjects know that the feedback probability is zero. An alternative baseline condition could have been one where subjects participate in the same problem-multiplication task as in our experiment but are never told that it is possible to learn their relative productivity and never receive relative rank feedback. We chose to design the experiment as described above in order to match the theoretical setting, where agents are always informed about the value of the feedback probability, and also to capture

[6] We do not pay subjects if their rank expectations turn out to be correct at the end of the round, because doing so would distort behavior: all subjects would declare that they rank last, solve zero problems, and achieve the last rank indeed. We understand the importance of incentive compatibility, and in other tasks where final compensation depends on output – and is not a flat wage like in the current experiment – paying people if they make the correct rank guess would certainly be desirable. However, as explained earlier, to understand how ego utility (i.e., liking to believe that we rank higher that others) changes behavior we are confined to a flat-wage environment.

[7] As a caveat, it is possible that the self-declared effort is not a perfect measure of the true effort the subject put into solving problems that round, as participants may strategically choose what level of effort to report. Since the predictions of Propositions 1 through 4 are about output, not effort levels, we test them by using the output measure, which is the actual number of problems solved. We only use the self-reported effort variable for one robustness check at the end of the empirical analysis. Also, since the correlation coefficient between self-reported effort and actual outcome produced is 0.34 ($p < 0.001$), this indicates that the effort variable is related to how hard people work.

the fact that in most corporate settings employees know that feedback is possible.

The experiment was programmed using the z-Tree software (Fischbacher (2007)). Subjects were given a written copy of the instructions (see Appendix). The task was also described verbally by the experimenter. Subjects then practiced the task for one period, but feedback about relative rank was not provided during that time. This practice period therefore served a dual purpose: it helped subjects understand how to use the software, and it also provided them individually with noisy information about their own ability on the multiplication task, since each person observed how many problems they solved in those 90 seconds. No communication or external aids (calculators, scratch paper, etc.) were allowed during the experiment. Subjects were recruited from Northwestern University using standard procedures. We conducted eight sessions, but one of them had to be excluded due to technical problems. We therefore present data from the 54 subjects (24 men and 30 women), participating in the remaining seven sessions. Each subject group consisted of six to nine people. Subjects received a fixed fee of $23 for their participation, independent of performance.

4 Experimental Results

4.1 Ex-ante effects of feedback

As predicted by Proposition 1, ex-ante information about the likelihood of receiving feedback at the end of the period about one's rank has a significant impact on both the subjects' expected relative performance, as well as on their actual output, measured as the number of multiplication problems solved correctly.[8] These effects are illustrated in Fig. 2.

Output is 12.20% higher (11.31 vs. 10.08 solved problems per round, $p < 0.001$ in a one-sided mean comparison test), and the expected rank is better (3.98 vs. 4.80, $p < 0.001$ in a one-sided mean comparison test) for participants who are in the "Maybe" feedback condition, than for those in the "No" feedback condition. In the "Sure" condition the average output is also significantly higher (10.78 vs. 10.08, $p < 0.04$ in a one-sided mean comparison test) and the expected rank better (4.12 vs. 4.80, $p < 0.001$ in a one-sided mean comparison test) than in the "No" feedback condition. There is no significant difference between the output or expected rank of subjects in the "Maybe" feedback condition versus "Sure" feedback condition. These effects characterize most subjects, as 61% of participants produce a higher output on average in trials when the feedback probability is 0.5 compared to when it is 0. Similarly, 69% of subjects produce on average higher output when the feedback probability is 1, compared to when it is 0.

Fig. 3 reveals significant gender effects on output and rank expectations, in each of the three feedback likelihood conditions. Men solve significantly more problems than women. Across all treatments, the average number of problems solved is 13.13 for men, and 8.80 for women ($p < 0.001$ in a one-sided mean comparison test), in line with the prior literature on gender and competitiveness

[8] While Proposition 1 has implications for beliefs about relative output ($E_i[y_i] - E_i[y_j]$), it is difficult to elicit these beliefs in a setting with more than two participants. We could have asked subjects in the beginning of each round to state how many more problems they expected to solve relative to, say, the median participant in the room. Nonetheless, given the heterogeneity in the number of people across all sessions of the experiment, the answers of participants in different sessions would have been difficult to compare. Hence we instead proxied for these beliefs about relative output by eliciting the subjects' beliefs about their relative rank (e.g., "This round I expect to rank first.").

(e.g. Gneezy, Niederle, & Rustichini (2003)). Also, men expect to rank better than women do (i.e., men report lower values for $ExpectedRank_t$). Across all conditions, men expect to receive a rank of 3.40, while women expect to receive a rank of 5.01. The difference is statistically significant ($p < 0.001$). This is consistent with prior findings. For instance, Huberman et al. (2004) observe that males seek status more than women, and Falk & Knell (2004) find that women have significantly lower aspiration levels than men regarding college education accomplishments.

While these results suggest an unconditional effect of the feedback likelihood on output and rank expectations, Proposition 1 specifically implies that only subjects who believe they have high relative ability will increase their output, and expect a better rank, when they are told that they are in a condition in which relative performance feedback is more likely to be received. The evidence shown in Figure 4 is consistent with its prediction. We document that subjects who have better than median rank expectations in the prior round (and hence believe that they have high relative ability) are those individuals who respond more to the ex-ante feedback condition in the current round. If told that feedback is possible (that is, when they are in the "Sure" or "Maybe" feedback conditions), these individuals increase output and expect a better rank, relative to situations when they are told rank feedback will not be provided. Specifically, output increases from an average of 11.88 in the "No" feedback condition to 12.91 in the "Sure" and "Maybe" conditions. Similarly, expected rank decreases (i.e., people expect to do better) from 3.72 in the "No" feedback condition to 3.02 in the "Sure" and "Maybe" conditions. Both effects are significant at $p < 0.01$. Compared to these individuals who think they have high relative ability, those subjects with poor rank expectations in the prior round do not react as much, in terms of output and future rank beliefs, to information about the feedback condition they face. For these subjects, output does not differ significantly depending on whether rank feedback is possible or not, and rank expectations only improve by 0.36 if feedback is possible, an effect half the size of that observed among participants who think they are relatively more able than their peers. These results support to the predictions of Proposition 1 regarding the role of the ex-ante feedback condition on the agents' output and beliefs.

Additionally, we find that the subjects' rank expectation and their actual rank are positively correlated, and this relationship becomes stronger in later periods. The Spearman rank correlation between $ExpectedRank_t$ and $Rank_t$ is 0.62 in the first six periods, 0.81 in periods seven through twelve, and 0.82 in periods thirteen through eighteen ($p < 0.001$ in all cases). The difference between the rank correlation measured in the first six periods and that measured in last six or twelve rounds is significant at $p < 0.001$. Therefore, as the task progresses, people get better at guessing their actual rank in the hierarchy.

4.2 Ex-post effects of feedback

Propositions 2, 3 and 4 imply that the feedback received regarding one's relative standing in the group has effects on the expectations of future rank and on the actual output produced in future rounds. We find evidence consistent with these predictions.

At the end of each round, subjects can receive one of three types of feedback regarding their relative ranking, depending on the relationship between their actual rank and the rank they expected to get. If $Rank_t > ExpectedRank_t$, feedback is negative, since subjects did worse than they expected. If $Rank_t < ExpectedRank_t$, feedback is positive, and if $Rank_t = ExpectedRank_t$, it is neutral. We use three indicator variables, $BadFeedback_t$, $GoodFeedback_t$ and $NeutralFeedback_t$ to capture these three types of events.

The regression models in Table 1 show the role of received feedback on future output, expectations of rank, and actual rank. The reported effects are measured relative to getting neutral feedback (but similar effects are obtained if measured relative to not getting any feedback at all). Relative to getting neutral feedback, doing better than expected in round $t-1$ (i.e., $GoodFeedback_{t-1}$=1) leads the subjects to expect a rank better by 0.50 in round t. Doing worse than expected (i.e., $BadFeedback_{t-1} = 1$) has the opposite effect, leading subjects to declare a worse expected rank, that is, a value higher by 0.54 for the variable $ExpectedRank_t$.

As predicted by Propositions 2, 3 and 4, subjects who received good feedback believe they will rank even better in the future (compared to their initial expectations) but end up ranking worse, while those who received bad feedback think they will rank worse (compared to their initial expectations), but in fact will rank better. After receiving negative feedback, people solve 0.74 more problems and achieve a rank better by 0.38. After receiving positive feedback, output is lower by 0.76 problems and the actual rank worsens by 0.63. In all regression models we include group fixed effects since unknown common factors may drive the effort and beliefs of all subjects in the same experimental session, and also cluster standard errors by subject. In unreported robustness checks we cluster them by session and the results remain statistically significant. We also control for the round number to account for possible time trends in output production or beliefs, and find that on average subjects solve 0.05 more problems in each additional round, but there is no significant effect of round on average rank expectations. We return to this output increase effect in the next section and examine its potential causes.

An important concern is that the effect of good or bad rank feedback on output may simply be driven by a mechanical mean-reversion process and therefore may not be caused by self-esteem considerations. For instance, after performing very well in round t by solving many problems a subject is likely to get good feedback, that is, to learn he ranked better than expected, but in the next round his performance will decrease because of mean reversion (perhaps because people get tired after working particularly hard in that round). We would then mistakenly attribute the lower subsequent output to the fact that the person was informed that they exceeded their own rank expectations in the prior round (i.e., we would think of this as a feedback/self-esteem driven effect). To isolate the relative rank feedback effect from a potential mechanical mean reversion effect, in the regression models in Table 1 we control for the prior values of output produced by the subject. Therefore the estimated coefficient on the $GoodFeedback_{t-1}$ variable in the first column in Table 1, for instance, tells us the difference in output produced in period t by two individuals who solved the same number of problems in period $t-1$ but differ in that one received no feedback about his relative rank at the end of round $t-1$, whereas the other was informed he exceeded his rank expectations. If there exist mean-reversion in output it should influence both these individuals equally, since they solved the same number of problems during period $t-1$ (and each person knows perfectly their output), and should not depend on whether they were told how many problems other participants solved that period. In other words, mean reversion effects are orthogonal to the effects of the $GoodFeedback_{t-1}$ variable once we control for prior output. The same argument applies for the interpretation of the effect of the $BadFeedback_{t-1}$ variable. Similarly, when predicting expected rank and rank in the current period (see Table 1) we control for the values of these variables in the prior period to account for the mechanical effect that people who are top ranked can only stay put or move lower in the rankings, whereas people who are already at the bottom of the hierarchy can not rank any lower.

To further examine whether mean-reversion is at play, we conduct the following analysis: For participants who do not receive rank feedback in a particular round, we sort them into those whose (unreported) rank was better than estimated (i.e., those who would have received positive feedback had they received feedback) and those whose (unreported) rank was worse than estimated (i.e., those who would have received negative feedback had they received feedback). We then calculate for each group the number of problems solved in the next round and the average rank. We do the same calculations for participants who receive feedback.

As you can see from the results in Table 2, for individuals who did not receive feedback at the end of round t-1, whether or not in that round they ranked better than expected has no significant impact on how much output they produce in round t, or on their rank in round t, indicating that we do not observe mean reversion in productivity. For the sample of people who received feedback at the end of round $t-1$, output (as well as the rank) in round t is dependent on whether these people were told that in round $t-1$ they did better or worse than they had expected. Specifically, for this subsample of people who got rank feedback, output in round t is on average 11.41 problems if the feedback in the prior round was negative, and 10.24 if the feedback was positive. This difference in output of 1.17 problems is significant statistically ($p = 0.01$) and economically, since it represents about 9% of the average output produced per round in this task. The evidence in the table therefore indicates that the difference in output produced by those who did better or worse than expected in the prior round is not driven by mechanical mean reversion, but it depends critically on whether people actually received feedback regarding their rank in that prior round.

The regression models in Table 1 also indicate that the likelihood of receiving feedback in the current round and the gender of the subject have similar effects on output and expected rank as shown earlier in the univariate analysis, and illustrated in Figures 2 and 3. If feedback is likely to be received – that is, the probability of seeing the ranking at the end of the period is not zero, as captured by the indicator variable $FeedbackLikely_t$ – then subjects expect and achieve better ranks, and the output is larger (however, the last effect is no longer statistically significant). Males expect better ranks than females, and solve more problems.

We also find evidence suggesting that the **ex-ante** difference in expected ability influences the agents' beliefs about relative rank, and their actual output, in the direction predicted by the model. Propositions 3 and 4 imply that the better agent i believes his competitor j is, the worse is the rank expected by i, and the lower is the output produced by i. In our experiment, the number of men in the group is an exogenous manipulation of the beliefs of women participants regarding their relative ability at solving the task. This argument is suggested by the results in Gneezy et al. (2003) and Niederle & Vesterlund (2007) who show that women are less effective than men in competitive environments, and this effect is stronger in settings where women compete against men than in single-sex competitive environments. Stereotypes about men being better at solving mathematical problem can also contribute to women competitors being more pessimistic about their relative ability when more of the session participants are male. Hence, we measure the difference in the agents' expected ability by the gender composition of our subject groups. As shown by the results in Table 3, we find that the number of men in the group matters for the productivity of women, but not for that of men. Women's expected and actual ranks are worse, and their output is lower, the more men there are in the group, as predicted by Propositions 3 and 4.

4.3 Hierarchies and the fight for dominance

The experimental evidence so far indicates that feedback about rank can impact the dynamics of rankings. But these effects should be less important once the performance hierarchy is established. Indeed, as shown in Table 4, when we estimate the same regression models as in Table 1 for rounds 1-9 and 10-18 separately, we find that $GoodFeedback_{t-1}$ and $BadFeedback_{t-1}$ influence strongly the subjects' rank expectations in the early rounds, but these effects are no longer statistically significant during later rounds. In other words, feedback about relative performance in a particular round does not influence a subject's expectations about where he will stand in the hierarchy in the future, once the hierarchy is determined.

In light of this suggestive evidence, we examine further whether stable hierarchies do get formed, and whether this influences output and beliefs. First, the data indicate that output grows over time: the average number of problems solved in rounds 1 and 18 are 9.48 and 12.5, respectively, and the difference is statistically significant ($p < 0.01$). This could in part be due to learning effects (i.e., participants find better ways to do multiplications), and in part due to a competition or ratcheting effect that is caused by people's desire not to lose their status in the hierarchy. We revisit these two effects at the end of this section. Moreover, we find that the standard deviation of output increases over time, from 3.31 problems in round 1 to 5.32 problems in round 18, consistent with subjects expending the appropriate effort levels needed to maintain their rank (i.e., high effort for top-ranked individuals, and low effort for bottom-ranked ones). The standard deviation of expected rank also increases in later rounds, from 1.87 in round 1 to 2.31 in round 18, suggesting that people's expectations "fan out" as they learn about their relative performance. Early on, subjects have similar priors about their relative ability, but as they get feedback regarding their output level, posterior beliefs about rank became more heterogeneous, in accordance with the group's diversity in abilities.

Another way to illustrate that hierarchies form early on and remain relatively stable is to see whether people who were at the bottom of the ranking in the early rounds of the task tend to stay at the bottom in later rounds, while people who started by being at the top of the ranking will stay at the top. For each participant we calculate their average rank in the first six, middle six and last six rounds of the task. We will refer to these as the early, middle and late stages of the task. For each of theses three stages, we assign subjects to one of three rank performance bins: low, middle and high, depending on their average rank during the six rounds that comprised the stage. Thus, subjects in the low rank performance bin in a particular stage are those in the bottom third of the performance distribution, as determined by how their average rank compares to the average rank of the others in their peer group. Subjects in the high rank performance bin are those in the top third of the performance distribution as measured by their average rank during that stage.

Figures 5 and 6 show how people transition across rank performance bins as the task progresses. Fourteen of the seventeen (82%) of the individuals who are in the bottom third of the rank hierarchy during rounds 1 through 6 end up in the same low rank performance bin during rounds 7 through 12, and also during rounds 13-18. Of the twenty-one subjects who are in the top third of the rank hierarchy during the first six rounds, eighteen (86%) are still top performers during rounds 7 through 12, and fifteen (71%) remain at the top during rounds 13-18. Thus, while there are instances where subjects move up and down the hierarchy, most people stay in the same rank performance bin they are in during the first six rounds of the task. This indicates that by the end of the first six rounds the hierarchy is already established.

While people's ranks do not change much once the hierarchy is formed, the average output of the group increases over time, as shown above and in the results in Table 1. Does this increase come from top performers working harder to maintain their top rank, or by people in the middle or low end of the hierarchy who want to get better rankings? The answer to this question is relevant for optimal team formation and dynamics. If the increase in output comes from people at the top of the ranking fighting for dominance, and not from people at the bottom trying to get a better rank, then it may be efficient to reshuffle peer groups by assigning bottom performers to new teams. There, they have a chance to be higher up in the ranking, and will expend effort to preserve their newly-acquired position, thus increasing the total output produced. The evidence we find is consistent with this hypothesis.

Figure 7 shows that the ratcheting effect observed in average output comes mainly from subjects who were at the top or in the middle of the hierarchy in the first six rounds. Individuals who ranked in the bottom third of the hierarchy early on have a slower rate of productivity increase relative to the other participants. Therefore, the increase in output over time comes mainly from high productivity subjects who fight to maintain or improve their rank.

An alternative interpretation of the increase in output over time seen in Figure 7 is that people simply get better at solving multiplication problems as the task progresses, and those that had better performance earlier on learn faster. This interpretation is unrelated to ego utility or to the ratcheting effect (that is, strategically choosing to work harder in order to obtain a good rank). To investigate this alternative explanation, we obtain a measure of how difficult it is for subjects to solve multiplication problems. We calculate the cost of effort ($CostOfEffort_t$) per multiplication problem as the ratio of declared effort to output produced by each subject in each round. We average this quantity across the three performance categories (early top, middle and bottom performers). For learning to explain the patterns in Figure 7, it should be the case that the rate of change in output and the rate of change in the cost of effort over time are negatively related. In other words, early top performers will increase their output at a faster pace relative to bottom performers because their cost of effort decreases at a faster pace over time. As the data in Table 5 show, we do not find this to be the case.

The output of early top performers increases at twice the rate over time as that of early bottom performers ($\frac{\Delta Output}{\Delta Round}$ is 0.21 and 0.11 for these two categories, respectively). The cost of effort, however, decreases faster over time for bottom performers ($\frac{\Delta CostOfEffort}{\Delta Round}$ is -0.01 for bottom performers and -0.004 for top performers). These average rates of change in output, and in the cost of effort over time are estimated by regressing $Output_t$ and $CostOfEffort_t$ on $Round_t$, for participants in each of the three early performance categories, and are statistically significant at conventional levels ($p < 0.05$). Therefore, learning effects (i.e., the task getting easier over time) can not be the sole explanation for the increase in output of those ranking well early on, since the task seems to get easier faster for early bottom performers. Hence, ego utility – as shown by our model and previous empirical results – can be a driver of output and lead to ratcheting at the top of the hierarchy, a pattern illustrated by the data in Figure 7 and Table 5. Consistent with the fight for dominance interpretation, we observe that throughout the task early top performers declare higher effort levels relative to early bottom performers (4.40 versus 4.02, on a scale from 1 to 6) and produce higher output (14.90 versus 6.65 multiplication problems per round) They also have a lower average cost of effort (0.30 versus 0.61). All of these differences are statistically significant ($p < 0.01$).

Since the above results come with the caveat that self-reported effort may not represent the true effort spent by participants, we use an additional approach for testing the learning-based alternative hypothesis. We invited twenty-six new participants, ages 18-24 (8 men and 18 women), all students at Northwestern University to take part in a "no rank feedback" version of the experiment. In this version, each participant knew his or her absolute performance, namely, the number of problems solved correctly during each round. However, the possibility of seeing one's relative performance was never mentioned, and relative rank information was never provided. There were between 6 and 7 participants in each of four experimental groups, working concurrently on solving the same multiplication problems as in the main experiment, for the same fixed pay of \$23. Using this sample, we wanted to understand whether better initial performers improve faster than poor initial performers, when no feedback about relative rank was provided. If that happened in the data, it would have shed doubt on our assertion that the increase over time in the output of the top performers in the feedback experiment was due to fighting for the top ranks, and it would have suggested our "ratcheting effect" could have been simply driven by faster learning by initial top performers, and not by competition.

We present the data from the "no rank feedback" experiment in Figure 8. We assign each participant to one of three early performance groups – high, medium and low performers – based on the output produced in the first six rounds of the experiment. In the figure we show the output produced by each of the 26 participants in each of the 18 rounds. We also plot a regression line for each of the three early performance groups, indicating the rate of growth of output after the rounds during which initial performance is measured (i.e., rounds 7 to 18). In other words, we regress output on round number, for each of the three subsamples of high, medium and low performers, and plot the line of best fit. The figure shows that those labeled as high performers early on do not improve faster than those who are medium or low early performers. Specifically, looking at the improvement in output during rounds 7-18, the slope of output with respect to round is 0.08 for the high group, 0.20 for the middle group and 0.23 for the low initial performance group. Therefore, the data from the "no rank feedback" experiment indicate that, while people tend to improve with time during our experiment, it is not the case that those who did well in the beginning will increase their output faster than the other participants. This suggests that learning effects are not the driver of the result we documented in the main experiment (where relative rank feedback was possible) that high initial performers improved faster than the low performers. Therefore, this supports our interpretation that in the main experiment the ratcheting of output observed among the best participants was related to their ego-utility and interest in maintaining a top rank.

5 Implications for Corporate Feedback Policies

In light of our findings, it is natural to ask what are the characteristics of optimal feedback policies. It is important to know whether organizations can increase their total output through optimal feedback provision, perhaps by changing the timing and content of information released to workers or by revealing information to certain individuals only. Even though the current experimental setup does not allow us to compare such complex feedback policies, our results have several implications for improving productivity that match the observed actions undertaken by firms to improve their performance appraisals (e.g., changes in frequency, benchmarks or performance labels).

For instance, firms could take advantage of the ex-ante effect of feedback likelihood on effort provision. Proposition 1 suggests that an organization could produce more if it tailored the feedback probability to the agent's expected or true relative ability. In particular, by providing more frequent feedback to agents who believe that they have relatively high ability, managers can take advantage of the positive ex-ante effect that the likelihood of feedback has on output. This result, as shown by Proposition 6, continues to hold in well-established teams where workers' abilities and feedback probabilities are common knowledge. A manager of a well-established team can increase the overall output simply by providing feedback more frequently to workers who posses relatively high ability.

Note, though, that the firm needs to commit to a feedback policy ex-ante. The model does not allow for the strategic choice of when to reveal to workers their rank. That is, the likelihood of receiving feedback in the model is not conditional on the employee having achieved a bad or good rank. Thus, Proposition 2, which states that those who get bad feedback will increase output in the future relative to its level in the absence of feedback, does not imply that providing only bad feedback (i.e., telling people their rank only when the underperform compared to others) is an optimal policy. This is because the equilibrium derived in the model assumes the firm commits to a probability of rank feedback provision that is independent of the agents' actual output.

The principal could also manipulate the beliefs of the agents to use both the ex-ante and ex-post effects of feedback. Proposition 3 suggests that if competitors appear to be too tough, agents will decrease their output. Moreover, Proposition 1 implies that the possibility of receiving rank feedback only has motivational effects for agents who believe they are better skilled than their competitors. Therefore, if the firm can improve workers' beliefs about their relative ability, these optimistic beliefs will have a positive impact on productivity.

As shown in Table 4, however, the ex-post effects of feedback wear off as time goes by and people learn their true standing among their peers. To prolong the effectiveness of relative rank information, organizations could provide noisy feedback to slow down the learning of one's rank in the hierarchy, or reshuffle work groups once the hierarchy is established.

Lastly, our empirical results show that output increases over time and that this effect is possibly due to the competition for top ranks among the best performing individuals (as suggested by Figure 7). The data thus indicate that agents in heterogeneous groups will split into top performers who keep fighting for high ranks, and bottom performers who compete much less. Similarly, we also observe that women produce significantly less output when there are more men in the group (as shown in Table 3). Making teams more homogenous may therefore provide otherwise low-rank workers with an opportunity to climb the hierarchy, and as a result, may restore their incentives to generate more output.

6 Conclusion

We propose that individuals' utility is influenced by private information regarding their relative performance. This hypothesis implies that feedback about rank has effects on both productivity and on the dynamics of the rank hierarchy in groups of workers doing similar tasks. These predictions are supported by experimental evidence. To separate our hypothesis from alternative explanations as to why rank information may change behavior, we employ an experimental setting where subjects receive a flat wage for working on a simple problem solving task, and where there can not exist reputation, strategy learning or peer monitoring effects.

We find that agents who believe they have relatively high ability will increase output, and expect to rank better, if told feedback is likely. After receiving feedback, those who got better ranks than expected will decrease output, but expect even better ranks in the future, while the opposite is true of people who ranked lower than expected. The productivity hierarchy is established early on in the task, and there is a ratcheting effect in output. People at the top of the hierarchy early on work harder over time, while people at the bottom do not improve their productivity as much.

These results suggest that in competitive settings productivity and beliefs are influenced by privately observed information about relative rank. The effects of private rank feedback on output are comparable to those of peer monitoring mechanisms documented in prior work. For example, Mas & Moretti (2009) find that a 10% increase in average co-worker productivity is associated with 1.7% increase in a worker's effort. By optimally arranging the mix of workers in each shift, the firm in their sample could improve productivity by 0.2%. Similarly, Falk & Ichino (2006) find that a 10% increase in a peer's output results in a 1.4% increase in a given individual's effort. We find that giving people an opportunity to privately compare themselves to others raises individual output on average by 12.20%, an effect comparable to that of peer monitoring.

Our results suggest that relative rank feedback can be strategically used to improve employee performance in organizations. Firms can vary the frequency and content of information released to workers, they can provide feedback to subsets of employees, depending on their actual abilities or perceptions of their standing among their peers, or can reshuffle teams to encourage all workers to work hard to achieve a good rank. Therefore, the value that individuals assign to learning that they compare well to their peers makes relative rank feedback a useful tool for increasing firm productivity.

References

Azmat, Ghazala, & Nagore Iriberri, 2009, The importance of relative performance feedback information: Evidence from a natural experiment using high school students, *Working paper* .

Bandiera, Oriana, Valentino Larcinese, & Imran Rasul, 2009, Blissful ignorance? Evidence from a natural experiment on the effect of individual feedback on performance, *Working paper* .

Bandura, A., 1986, *Social foundations of thought and action: A social cognitive theory* (Englewood Cliffs, NJ: Prentince-Hall).

Benabou, Roland, & Jean Tirole, 2002, Self confidence and personal motivation, *Quarterly Journal of Economics* 117, 871–915.

Benabou, Roland, & Jean Tirole, 2003, Intrinsic and extrinsic motivation, *Review of Economic Studies* 70, 489–520.

Blanes i Vidal, Jordi, & Mareike Nossol, 2009, Tournaments without prizes: Evidence from personnel records, *Working paper* .

Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, & Aldo Rustichini, 2010, Overconfidence is a social signaling bias, *Working paper* .

Clark, A. E., & A. J. Oswald, 1996, Satisfaction and comparison income, *Journal of Public Economics* 61, 359–381.

Clark, Andrew, Davis Masclet, & Marie-Claire Villeval, 2006, Effort and comparison income: Survey and experimental evidence, *Working Paper* .

Cowen, Tyler, & Amihai Glazer, 2007, Esteem and ignorance, *Journal of Economic Behavior & Organization* 63, 373–383.

Cromer, Fred Eugene, 1974, Structural models for predicting the difficulty of multiplication problems, *Journal for Research in Mathematics Education* 5, 155–166.

Deci, E., 1975, *Intrinsic motivation* (New York, Plenum Press).

Deci, E., R. Koestner, & R.M. Ryan, 1999, A meta-analytic review of experiments examining the role of extrinsic rewards on intrinsic motivation, *Journal of Applied Psychology* .

Delfgaauw, Josse, Robert Dur, Joeri Sol, & Willem Verbeke, 2009, Tournament incentives in the field: Gender differences in the workplace, *Working paper* .

Dohmen, Thomas, Armin Falk, Klaus Fliessbach, Uwe Sunde, & Bernd Weber, 2011, Relative versus absolute income, joy of winning, and gender: Brain imaging evidence, *Journal of Public Economics* 95, 279–285.

Easterlin, Richard A., 1995, Will raising the incomes of all increase the happiness of all?, *Journal of Economic Behavior & Organization* 27, 35–47.

Ederer, Florian, 2010, Feedback and motivation in dynamic tournaments, *Journal of Economics and Management Strategy* 19, 733–769.

Eriksson, Tor, Anders Poulsen, & Marie Claire Villeval, 2009, Feedback and incentives: experimental evidence, *Labor Economics* 16, 679–688.

Ertac, Seda, 2005, Social comparisons and optimal information revelation: Theory and experiments, *Working paper* .

Falk, Armin, David Huffman, & Uwe Sunde, 2006, Self-confidence and search, *IZA Discussion Paper* .

Falk, Armin, & A. Ichino, 2006, Clean evidence on peer effects, *Journal of Labor Economics* 24, 39–57.

Falk, Armin, & M. Knell, 2004, Choosing the Joneses: Endogenous goals and reference standards, *Scandinavian Journal of Economics* 106, 417–435.

Fischbacher, Urs, 2007, Z-tree: Zurich toolbox for ready-made economic experiments, *Experimental Economcs* 10, 171–178.

Frank, Robert H., 1984, Interdependent preferences and the competitive wage structure, *RAND Journal of Economics* 15, 510–520.

Frank, Robert H., 1985, *Choosing the right pond: Human behavior and the quest for status* (New York: Oxford University Press).

Frey, Bruno, 1997, *Not just for the money – An economic theory of personal motivation* (Cheltenham, U.K.: Edward Elgar).

Frey, B.S., & R. Jegen, 2001, Motivation crowding theory: A survey of empirical evidence, *Journal of Economic Surveys* 15, 589–611.

Gneezy, Uri, Muriel Niederle, & Aldo Rustichini, 2003, Performance in competitive environments: Gender differences, *Quarterly Journal of Economics* 118, 1049–1074.

Gneezy, Uri, & Aldo Rustichini, 2000, Pay enough or don't pay at all, *The Quarterly Journal of Economics* 115, 791–810.

Huberman, Bernardo A., Christoph H. Loch, & Ayse Onculer, 2004, Status as a valued resource, *Social Psychology Quarterly* 103–114.

Kandel, Eugene, & Edward P. Lazear, 1992, Peer pressure and partnerships, *Journal of Political Economy* 100, 801–817.

Kluger, Avraham N., & Angelo DeNisi, 1996, The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory, *Psychological Bulletin* 119, 254–284.

Koszegi, Botond, 2006, Ego utility, overconfidence and task choice, *Journal of the European Economic Association* 4, 673–707.

Luttmer, Erzo, 2005, Neighbors as negatives: Relative earnings and well-being, *Quarterly Journal of Economics* 120, 963–1002.

Major, B., M. Testa, & W. H Bylsma, 1991, Responses to upward and downward social comparisons: The impact of esteem-relevance and perceived control, *in: Suls, J. and Wills, T. A. (eds.), Social comparison: Contemporary theory and research, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers* 237–260.

Mas, Alexander, & Enrico Moretti, 2009, Peers at work, *American Economic Review* 99, 112–145.

Maslow, Abraham H., 1943, A theory of human motivation, *Psychological Review* 370–396.

McClelland, D.C., J.W. Atkinson, R.A. Clark, & E.L. Lowell, 1953, *The achievement motive* (Princeton: Van Nostrand).

Neckermann, Susanne, & Bruno S. Frey, 2008, Awards as incentives, *Working paper* .

Niederle, Muriel, & Lise Vesterlund, 2007, Do women shy away from competition? Do men compete too much?, *Quarterly Journal of Economics* 122, 1067–1101.

Prewitt, Edward, 2007, *Managing performance to maximize results*, chapter Should you use 360 feedback for performance reviews? (Harvard Business School Press).

Seta, J., 1982, The impact of comparison processes on coactors' task performance, *Journal of Personality and Social Psychology* 42.

Smither, James W., Manuel London, & Richard R. Reilly, 2005, Does performance improve following multisource feedback? a theoretical model, meta-analysis, and review of empirical findings, *Personnel Psychology* 58, 33–66.

Szymanski, Kate, & Stephen G. Harkins, 1987, Social loafing and self-evaluation with a social standard, *Journal of Personality and Social Psychology* 891–897.

Weinberg, B., 1999, A model of overconfidence, *Working paper, Ohio State University* .

White, Paul H., Margaret M. Kjelgaard, & Stephen G. Harkins, 1995, Testing the contribution of self-evaluation to goal-setting effects, *Journal of Personality and Social Psychology* 69, 69–79.

Wood, J. V., & K. L. Taylor, 1991, *Social comparison: Contemporary theory and Research*, chapter Serving self-relevant goals through social comparison, 23–49 (New Jersey: Lawrence Erlbaum Associates, Inc.).

Figure 1: Sequence of events in a round.

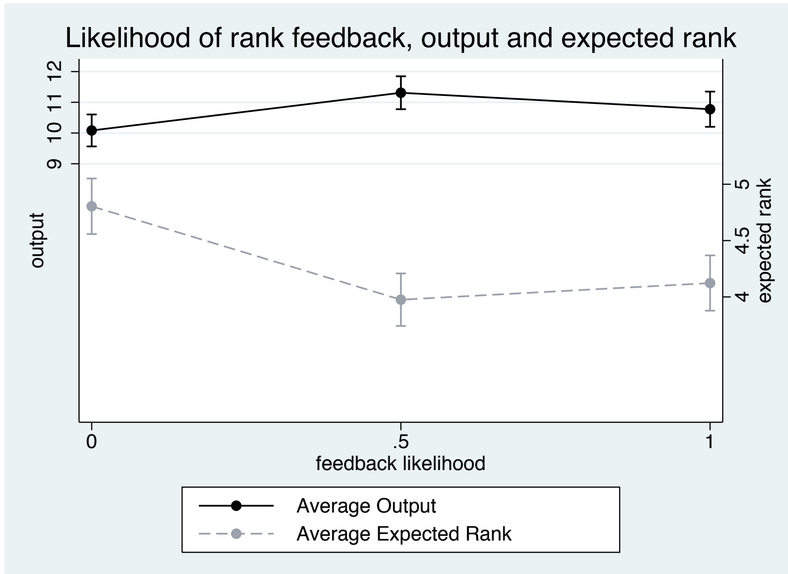

Figure 2: Feedback likelihood, output and expected rank

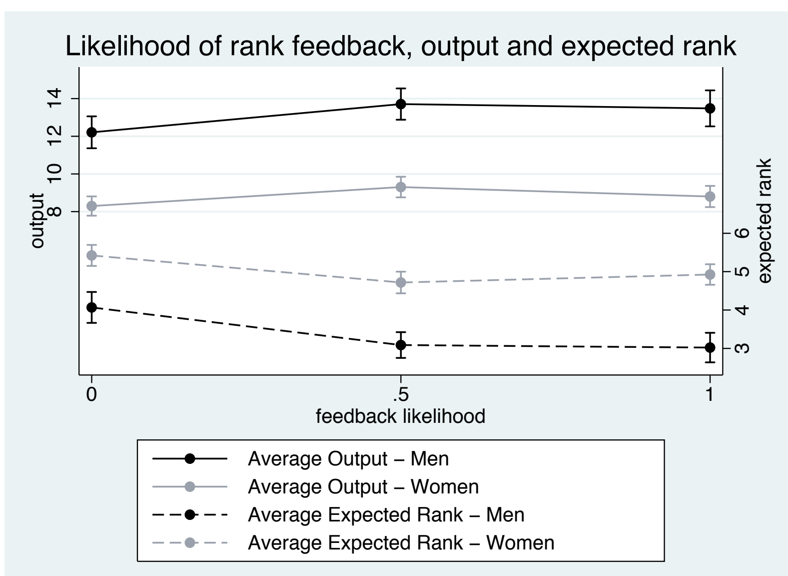

Figure 3: Feedback likelihood, output and expected rank, by gender

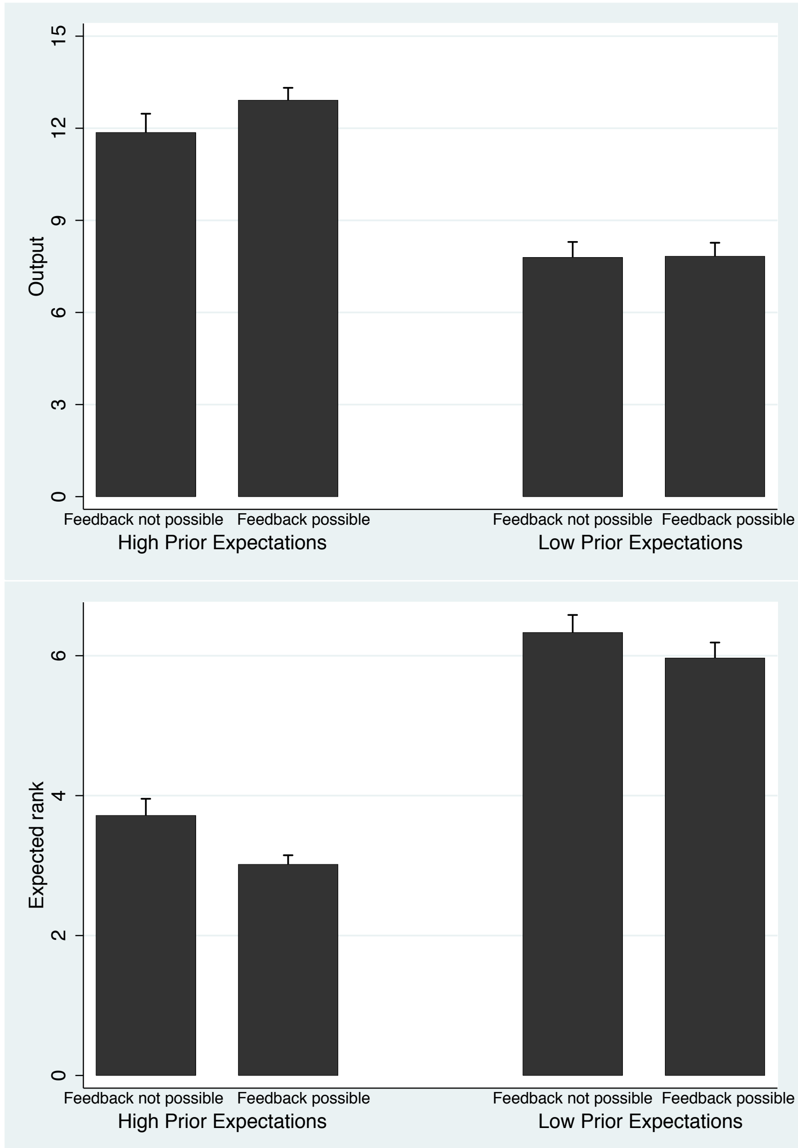

Figure 4: The effect of increasing the likelihood of relative rank feedback on output and expected rank as a function of the subjects' beliefs about their relative ability.

	Low$^{7\text{-}12}$	Medium$^{7\text{-}12}$	High$^{7\text{-}12}$
Low$^{1\text{-}6}$	13/17	4/17	0/17
Medium$^{1\text{-}6}$	4/18	12/18	2/18
High$^{1\text{-}6}$	0/19	2/19	17/19

Figure 5: Transitions across ranks: rounds 1-6 to rounds 7-12.

	Low$^{13\text{-}18}$	Medium$^{13\text{-}18}$	High$^{13\text{-}18}$
Low$^{1\text{-}6}$	13/17	4/17	0/17
Medium$^{1\text{-}6}$	4/18	11/18	3/18
High$^{1\text{-}6}$	0/19	3/19	16/19

Figure 6: Transitions across ranks: rounds 1-6 to rounds 13-18.

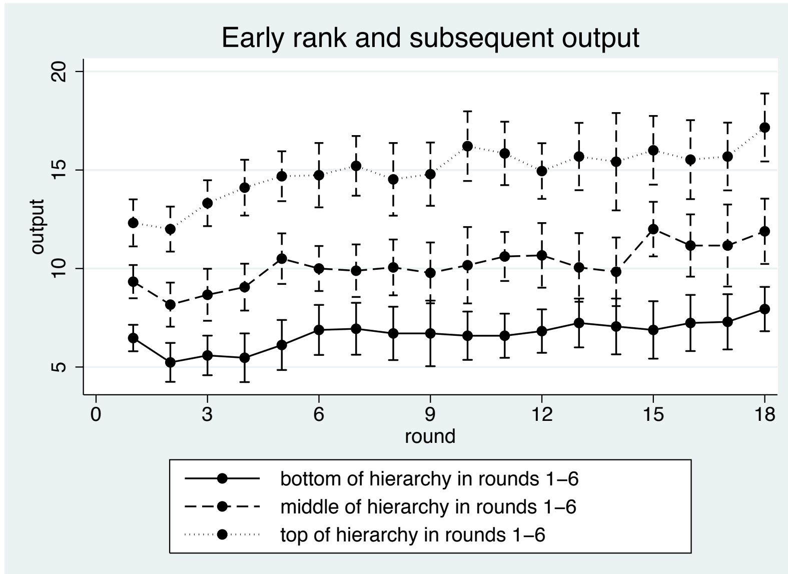

Figure 7: The average output produced each round by subjects who were at the top, in the middle or at the bottom of the rank hierarchy during the first six rounds.

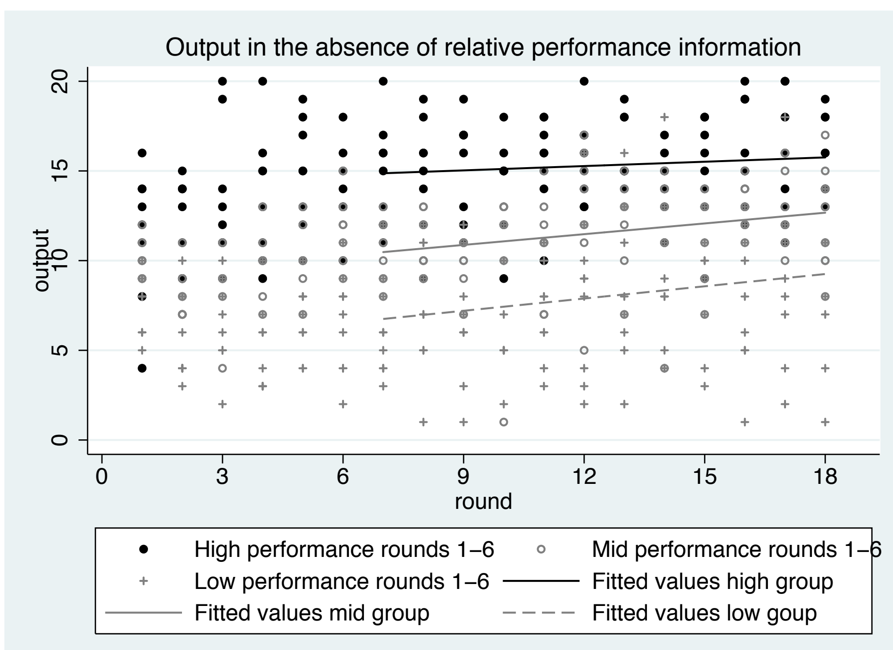

Figure 8: Output produced each round by top, middle and low early performers, classified based on their output during the first six rounds. Lines of best fit from the regression of output on round, using data from rounds 7 to 18, are also shown, for observations belonging to high, medium or low initial performers.

Table 1: The ex-post impact of feedback on expected rank, actual rank and output
$Output_t$ is the number of multiplication problems solved correctly by the subject in round t. $ExpectedRank_t$ is the rank that the subject expects to get in round t, as declared in the beginning of the round. $Rank_t$ is the actual rank achieved by the subject in round t. Low values for *ExpectedRank* and *Rank* indicate better rank expectations, and actual rank, respectively (e.g. the top performing subject has $Rank = 1$). $ExPostFeedback_t$ is an indicator variable equal to 1 if the subject received relative ranking feedback at the end of round t. $GoodFeedback_t$ is an indicator variable equal to 1 if the subject received positive feedback at the end of round t, i.e., when $Rank_t < ExpectedRank_t$. $BadFeedback_t$ is an indicator variable equal to 1 if the subject received negative feedback at the end of round t, i.e., when $Rank_t > ExpectedRank_t$. $FeedbackLikely_t$ is an indicator variable equal to 1 if the probability the subject will receive feedback on relative ranking is 0.5 or 1 (i.e., if the subject is in the "Maybe" or "Sure" feedback treatment). *Male* is an indicator variable equal to 1 if the subject is male. $Round_t$ is the round number. The reference category is given by observations where subjects received neutral rank information at the end of the prior round ($NeutralFeedback_{t-1} = 1$). T-statistics are in parentheses.

	$Output_t$ Coef./t	$ExpectedRank_t$ Coef./t	$Rank_t$ Coef./t
$GoodFeedback_{t-1}$	–0.76	–0.50	0.63
	(–2.55)**	(–4.56)***	(3.54)***
$BadFeedback_{t-1}$	0.74	0.54	–0.38
	(2.19)**	(4.17)***	(–2.26)**
$ExPostFeedback_{t-1}$	–0.12	0.23	–0.07
	(–0.36)	(1.49)	(–0.44)
$FeedbackLikely_t$	0.56	–0.55	–0.31
	(1.52)	(–2.81)***	(–1.79)*
$Output_{t-1}$	0.75		
	(13.43)***		
$ExpectedRank_{t-1}$		0.79	
		(13.52)***	
$Rank_{t-1}$			0.66
			(10.15)***
Male	1.35	–0.39	–0.69
	(3.33)***	(–2.19)**	(–2.63)**
$Round_t$	0.05	–0.01	–0.00
	(3.79)***	(–1.49)	(–0.20)
Adj. R^2	0.664	0.701	0.540
No. of obs	918	918	918

Robust standard errors clustered by subject
Session fixed effects, constant term included
$^{*}p < .10, ^{**}p < .05, ^{***}p < .01$

Table 2: Mean reversion or feedback effect?

For participants who do not receive rank feedback in a particular round, we sort them into those whose (unreported) rank was better than estimated (i.e., those who would have received positive feedback had they received feedback) and those whose (unreported) rank was worse than estimated (i.e., those who would have received negative feedback had they received feedback). We then calculate for each group the number of problems solved in the next round and the average rank. We do the same calculations for participants who receive feedback.

	Did not receive relative rank feedback at end of round $t-1$	Received relative rank feedback at end of round $(t-1)$
Got better rank than expected in round $t-1$	Output in round $t = 10.44$ Rank $= 4.36$ Observations $= 170$	Output in round $t = 10.24$ Rank $= 4.32$ Observations $= 146$
Got worse rank than expected in round $t-1$	Output in round $t = 10.60$ Rank $= 4.44$ Observations $= 281$	Output in round $t = 11.41$ Rank $= 4.07$ Observations $= 321$
	Diff. in output$= 0.16$ ($p = 0.4$) Diff. in rank$= 0.08$ ($p = 0.4$)	Diff. in output$=1.17^{***}$ ($p = 0.01$) Diff. in rank$=-0.25^{*}$ ($p = 0.1$)

Table 3: Heterogeneity in subjects' competitive abilities, output and expectations

Heterogeneity in the ability to compete is proxied by the gender mix in each subject group. The sample is split by the subjects' gender (Panel A: Women, Panel B: Men). $MenInGroup_t$ and $GroupSize_t$ are the number of male subjects, and the total number of subjects in the group, respectively. $Round_t$ is the round number. $Output_t$ is the number of multiplication problems solved correctly by the subject in round t. $ExpectedRank_t$ is the rank that the subject expects to get in round t, as declared in the beginning of the round. $Rank_t$ is the actual rank achieved by the subject in round t. Low values for $ExpectedRank$ and $Rank$ indicate better rank expectations, and actual rank, respectively (e.g. the top performing subject has $Rank = 1$). T-statistics are in parentheses.

	Panel A: Women Only			Panel B: Men Only		
	$Output_t$	$ExpectedRank_t$	$Rank_t$	$Output_t$	$ExpectedRank_t$	$Rank_t$
	Coef./t	Coef./t	Coef./t	Coef./t	Coef./t	Coef./t
$MenInGroup_t$	–1.10	0.37	0.60	0.25	–0.20	–0.10
	(–2.59)**	(1.69)*	(2.68)**	(0.32)	(–0.50)	(–0.25)
$GroupSize_t$	0.42	0.29	0.19	–1.33	0.66	0.63
	(0.97)	(1.25)	(0.86)	(–1.13)	(1.20)	(1.18)
$Round_t$	0.11	0.01	0.01	0.22	–0.02	0.00
	(4.57)***	(0.72)	(0.88)	(5.53)***	(–1.00)	(0.17)
$Constant$	7.85	1.47	1.56	20.75	–0.92	–1.33
	(2.97)***	(1.06)	(1.10)	(2.46)**	(–0.32)	(–0.47)
Adj. R^2	0.157	0.192	0.240	0.096	0.067	0.064
No. of obs	540	540	540	432	432	432

Robust standard errors clustered by subject

$^{*}p < .10, ^{**}p < .05, ^{***}p < .01$

Table 4: Diminishing effects of feedback over time

The table illustrates the ex-post impact of feedback on estimated rank, actual rank and effort, for rounds 1-9 (Panel A) and 10-18 (Panel B). $Output_t$ is the number of multiplication problems solved correctly by the subject in round t. $ExpectedRank_t$ is the rank that the subject expects to get in round t, as declared in the beginning of the round. $Rank_t$ is the actual rank achieved by the subject in round t. Low values for $ExpectedRank$ and $Rank$ indicate better rank expectations, and actual rank, respectively (e.g. the top performing subject has $Rank = 1$). $ExPostFeedback_t$ is an indicator variable equal to 1 if the subject received relative ranking feedback at the end of round t. $GoodFeedback_t$ is an indicator variable equal to 1 if the subject received positive feedback at the end of round t, i.e., when $Rank_t < ExpectedRank_t$. $BadFeedback_t$ is an indicator variable equal to 1 if the subject received negative feedback at the end of round t, i.e., when $Rank_t > ExpectedRank_t$. $FeedbackLikely_t$ is an indicator variable equal to 1 if the probability the subject will receive feedback on relative ranking is 0.5 or 1 (i.e., if the subject is in the "Maybe" or "Sure" feedback treatment). $Male$ is an indicator variable equal to 1 if the subject is male. $Round_t$ is the round number.

	Panel A: Rounds 1-9			Panel B: Rounds 10-18		
	$Output_t$	$ExpectedRank_t$	$Rank_t$	$Output_t$	$ExpectedRank_t$	$Rank_t$
	Coef./t	Coef./t	Coef./t	Coef./t	Coef./t	Coef./t
$GoodFeedback_{t-1}$	0.08	–0.73	0.25	–1.70	–0.22	1.03
	(0.19)	(–4.19)***	(0.94)	(–3.82)***	(–1.12)	(4.26)***
$BadFeedback_{t-1}$	1.20	0.89	–0.48	0.40	0.24	–0.32
	(2.71)***	(4.98)***	(–1.84)*	(0.79)	(1.23)	(–1.36)
$ExPostFeedback_{t-1}$	–0.39	0.09	0.05	0.17	0.31	–0.18
	(–1.08)	(0.70)	(0.24)	(0.34)	(1.39)	(–0.76)
$FeedbackLikely_t$	0.41	–0.31	–0.26	0.75	–0.75	–0.37
	(1.39)	(–1.82)*	(–1.52)	(1.43)	(–3.10)***	(–1.61)
$Output_{t-1}$	0.81			0.70		
	(16.60)***			(10.42)***		
$ExpectedRank_{t-1}$		0.82			0.77	
		(16.74)***			(10.64)***	
$Rank_{t-1}$			0.69			0.63
			(10.35)***			(8.64)***
$Male$	0.85	–0.35	–0.59	1.80	–0.42	–0.79
	(2.20)**	(–2.63)**	(–2.07)**	(3.73)***	(–1.88)*	(–2.91)***
$Round_t$	0.08	–0.01	–0.02	0.06	0.00	0.00
	(1.66)	(–0.73)	(–0.94)	(1.78)*	(0.04)	(0.15)
Adj. R^2	0.665	0.717	0.526	0.657	0.697	0.547
No. of obs	432	432	432	486	486	486

Robust standard errors clustered by subject

Session fixed effects, constant term included

$^{*}p < .10, ^{**}p < .05, ^{***}p < .01$

Table 5: Ratcheting effect or learning?

Subjects are divided into three categories (top, middle and bottom performers) depending on their rank in the hierarchy during the first six rounds of the task, as in Figure 7. $Effort_t$ is an input provided by each subject at the end of each round, before the ranking information is shown. $Output_t$ is the number of multiplication problems solved correctly by each subject in each round. $Cost\ of\ effort_t$ is calculated as $\frac{Effort_t}{Output_t}$. The average rate of change in output and in the cost of effort over time are captured by variables $\frac{\Delta Output}{\Delta Round}$ and $\frac{\Delta CostOfEffort}{\Delta Round}$, respectively, and are estimated by regressing $Output_t$ and $Cost\ of\ effort_t$ on $Round_t$ for subjects in each of the three early performance categories. The reported estimates are significantly different from zero at conventional levels ($p < 0.05$).

Ranking in rounds $1-6$	*Average declared effort per round*	*Average output per round*	*Average cost of effort per round*	$\frac{\Delta Output}{\Delta Round}$	$\frac{\Delta CostOfEffort}{\Delta Round}$
Top of hierarchy	4.40	14.90	0.30	0.21	-0.004
Middle of hierarchy	4.38	10.17	0.44	0.16	-0.01
Bottom of hierarchy	4.02	6.65	0.61	0.11	-0.01

A Appendix: Model

As indicated in the main text, an agent who does not know his own or his opponent's ability and expects to get feedback about the opponent with probability p has the following expected utility function:

$$E_i(u_i) = (1-p)\,E_i(y_i) + p\left(E_i(y_i) - E_i(y_j)\ln\left(\frac{k}{k-s_i}\right)\right) - \alpha + \gamma E_i(a_i)\ln(\beta - e_i + ps_i) \tag{1}$$

Agent i, therefore, faces the following problem:

$$\max_{e_i, s_i} E_i(a_i) + e_i - pE_i(y_j)(\ln k - \ln(k - s_i)) - \alpha + \gamma E_i(a_i)\ln(\beta - e_i + ps_i) \tag{2}$$

and the resulting first order conditions yield:

$$e_i^* = \beta - \gamma E_i(a_i) + p\left(k - E_i\left(y_j^*\right)\right) \tag{3}$$

$$s_i^* = k - E_i\left(y_j^*\right) \tag{4}$$

Equation (3) shows that effort is decreasing in expectations of one's own ability. This effect comes from the fact that the marginal cost of effort is increasing in $E_i(a_i)$. Since the cost function we use is decreasing in the ability level, being more able makes performing the task less unpleasant. This cost specification means that the benefits (coming from the cost function) of being of high ability are highest for low levels of effort and decrease as the effort level increases. This assumption captures the idea that there are limits to human effort and as we continue to work harder or longer, the ability-driven advantages wear off and at some level of effort everybody is equally exhausted. Also, the equilibrium level of effort increases in the level of the chosen standard. Being more motivated for the task makes a given level of effort seem less costly and allows the agent to take on extra work. Further, the more agent i expects agent j to produce, the lower a standard he sets and the lower a level of effort he puts into his task.

We now proceed to the general equilibrium solution of the model. For simplicity, to avoid infinite hierarchies of beliefs, we restrict attention to the first order beliefs, that is, to beliefs about one's own ability ($E_i(a_i)$ and $E_j(a_j)$) and beliefs about ability of the other player ($E_i(a_j)$ and $E_j(a_i)$). Second order beliefs – that is, beliefs of player i about the beliefs of player j – are such that $E_i(E_j(a_j)) = E_i(a_j)$ and $E_i(E_j(a_i)) = E_i(a_i)$.

Also, $E_i(E_j(p)) = E_j(p) = E_j(E_i(q)) = E_i(q) = \frac{1}{2}$. Therefore agent i expects agent j to produce:

$$E_i\left(y_j^*\right) = E_i(a_j) + \beta - \gamma E_i(a_j) + \frac{k}{2} - \frac{1}{2}E_i a_i - \frac{1}{2}E_i\left(E_j\left(e_i^*\right)\right) \tag{5}$$

Let $\overline{e}_i^* \equiv E_i\left(E_j\left(e_i^*\right)\right) = \int_0^1 e_i^*(p)f(p)dp$. Then combining equations (5) and (3) we get that:

$$e_i^* = \beta + \frac{pk}{2} - p\beta + \left(\frac{p}{2} - \gamma\right)E_i(a_i) - p(1-\gamma)E_i(a_j) + p\frac{1}{2}\overline{e}_i^* \tag{6}$$

After taking expectations with respect to probability p in equation (6) we obtain:

$$\overline{e}_i^* = \frac{1}{3}\left(2\beta + k + (1 - 4\gamma)E_i(a_i) - 2(1-\gamma)E_i(a_j)\right) \tag{7}$$

Combining equations (7) and (6) we obtain the formula for the equilibrium level of effort of agent i who has beliefs $E_i(a_i)$ and $E_i(a_j)$:

$$e_i^* = \frac{2p(1-\gamma) - 3\gamma}{3}E_i(a_i) - \frac{4p(1-\gamma)}{3}E_i(a_j) + \beta - \frac{2\beta p}{3} + \frac{2kp}{3} \tag{8}$$

Using equation (4) and (8) we obtain the equilibrium level of standard:

$$s_i^* = \frac{2(1-\gamma)}{3}E_i(a_i) - \frac{4(1-\gamma)}{3}E_i(a_j) + \frac{2}{3}(k - \beta) \tag{9}$$

In equilibrium agent i produces the following amount of output:

$$y_i^* = a_i + \frac{2p(1-\gamma) - 3\gamma}{3}E_i(a_i) - \frac{4p(1-\gamma)}{3}E_i(a_j) + \beta - \frac{2\beta p}{3} + \frac{2kp}{3} + \widetilde{\varepsilon}_i \tag{10}$$

and ex-ante expects to produce:

$$E_i(y_i^*) = \frac{(2p+3)(1-\gamma)}{3} E_i(a_i) - \frac{4p(1-\gamma)}{3} E_i(a_j) + \beta - \frac{2\beta p}{3} + \frac{2kp}{3} \tag{11}$$

We use equations (8) - (11) to derive the main propositions in the paper.

Our model predicts that the feedback policy can influence both productivity and beliefs even before any rank information is revealed to the agents. In particular, agents with different likelihoods of receiving information about their opponent's output will (all else equal) expect to rank differently and will produce different levels of output.

Proposition 1 *If the agent believes that his ability is relatively high (low) compared to the ability of the competitor then he will produce more (less) output and expect better (worse) relative performance when the likelihood of feedback increases.*

Proof. Using equation (10) we get that $\frac{dy_i^*}{dp} = \frac{2}{3}\left((1-\gamma)(E_i(a_i) - 2E_i(a_j)) + k - \beta\right)$.
Since $\gamma < 1$, $\frac{dy_i^*}{dp} > 0 \Leftrightarrow E_i(a_i) > 2E_i(a_j) - \frac{k-\beta}{1-\gamma}$ and $\frac{dy_i^*}{dp} \leq 0 \Leftrightarrow E_i(a_i) \leq 2E_i(a_j) - \frac{k-\beta}{1-\gamma}$.

We measure expected relative performance using the difference in agents' expected outputs, $E_i(y_i^*) - E_i(y_j^*)$, and say that agent expects better relative performance when this difference increases. The probabilities with which the agents receive feedback are not correlated, and thus we get:

$\frac{d\left(E_i(y_i^*)-E_i(y_j^*)\right)}{dp} = \frac{dE_i(y_i^*)}{dp} = \frac{2}{3}\left((1-\gamma)(E_i(a_i) - 2E_i(a_j)) + k - \beta\right)$

Since $\gamma < 1$, $\frac{d\left(E_i(y_i^*)-E_i(y_j^*)\right)}{dp} > 0 \Leftrightarrow E_i(a_i) > 2E_i(a_j) - \frac{k-\beta}{1-\gamma}$ and $\frac{d\left(E_i(y_i^*)-E_i(y_j^*)\right)}{dp} \leq 0 \Leftrightarrow E_i(a_i) \leq 2E_i(a_j) - \frac{k-\beta}{1-\gamma}$. ■

This proposition implies that giving subjects the opportunity to compare themselves to others makes the sufficiently self-confident ones (i.e., individuals for whom $E_i(a_i)$ is sufficiently high relative to $E_i(a_j)$) more productive and more optimistic about their relative position in the group, which is highly desirable for the principal.

Comparative statics allow us to predict how agents who initially do not know their relative position in the group adjust effort and beliefs about future rank as they change their perceptions of relative ability (but note that we do not explicitly model the belief updating process). Different patterns in behavior and beliefs will occur after good and bad feedback, that is, after the subject learns that he ranked better or worse than he expected.

Proposition 2 *After receiving good (bad) feedback about one's own ability, i.e., after the agent learns that he is better (less) skilled than he expected, the agent's output will decrease (increase) if $p < (\geq) \frac{3\gamma}{2(1-\gamma)}$ (sufficient condition is that $\gamma > (\leq) \frac{2}{5}$).*

Proof. From equation (10) we get $\frac{dy_i^*}{dE_i(a_i)} = \frac{2p(1-\gamma)-3\gamma}{3}$ and

$\frac{dy_i^*}{dE_i(a_i)} < 0 \Leftrightarrow p < \frac{3\gamma}{2(1-\gamma)}$ and $\frac{dy_i^*}{dE_i(a_i)} \geq 0 \Leftrightarrow p \geq \frac{3\gamma}{2(1-\gamma)}$

Note that since $p \leq 1$, if $\gamma > \frac{2}{5}$ then $\frac{dy_i^*}{dE_i(a_i)} < 0$. ■

There are two channels through which changes in expected own ability exert influence on output. First, for an agent who believes to be better skilled than initially thought, the same level of effort, other things equal, is less costly because the cost of effort function is assumed to decrease in ability. Nonetheless, since the marginal cost of effort increases in ability, while the marginal benefit is constant (and equal to 1), in equilibrium the agent whose beliefs about own ability have improved will choose a lower effort level. Second, he increases the level of standard which in turn leads to higher effort level. This is caused by the fact that the cross partial derivative of the cost of effort with respect to standard and ability is negative, which means that the more skilled agent i is, the more he enjoys difficult tasks. For high γ or low feedback probability p the overall effect of increased own ability expectations on future output is negative. This happens because a high value of γ implies that the marginal cost of effort is increasing in ability at a faster pace, and a low probability of feedback p weakens the positive motivational effect of high standard on effort.

Proposition 3 *If the agent learns that his competitor is better (less) skilled than he expected, he will decrease (increase) his future output.*

Proof. From equation (10) we get $\frac{dy_i^*}{dE_i\left(a_j\right)} = -\frac{4p(1-\gamma)}{3} < 0$. ■

If $E_i(a_j)$ increases agent i expects to be more hurt by the relative comparison with his peer and in order to protect himself he sets a lower standard (which makes the relative comparison less important in the utility function) and as a result works less.

From Propositions 2 and 3 we learn that an agent will change his future output when the feedback he receives about his own and/or his opponent's ability is not in accordance with his current beliefs. For example, an agent who learned that he is higher in the talent hierarchy (i.e., his own ability is higher and the ability of his opponent is lower) will increase his future output if $p \geq \frac{3\gamma}{2(1-\gamma)}$. For $p < \frac{3\gamma}{2(1-\gamma)}$, the direction of change in the output will depend on the strength of the effect of own ability relative to that of the competitor's ability.

The next proposition establishes formally how agent's beliefs change after he receives feedback about his relative position in the group.

Proposition 4 *When the agent's beliefs about relative performance are revised upwards (downwards), he expects better (worse) relative performance in the future.*

Proof. As in Proposition 1, we measure relative performance using the difference in agents' outputs, $E_i\left(y_i^*\right) - E_i\left(y_j^*\right)$, and say that agent expects better relative performance when this difference increases.

$\frac{d\left(E_i\left(y_i^*\right)-E_i\left(y_j^*\right)\right)}{dE_i(a_i)} = \frac{dE_i\left(y_i^*\right)}{dE_i(a_i)} - \frac{dE_i\left(y_j^*\right)}{dE_i(a_i)}$

$E_i\left(y_j^*\right) = \frac{4}{3}(1-\gamma) E_i(a_j) - \frac{2}{3}(1-\gamma) E_i(a_i) + \frac{2\beta+k}{3}$

$\Rightarrow$

$\frac{d\left(E_i\left(y_i^*\right)-E_i\left(y_j^*\right)\right)}{dE_i(a_i)} = \frac{(2p+3)(1-\gamma)}{3} - \frac{2(1-\gamma)}{3} = \frac{(2p+1)(1-\gamma)}{3} > 0$

$\frac{d\left(E_i\left(y_i^*\right)-E_i\left(y_j^*\right)\right)}{dE_i\left(a_j^*\right)} = -\frac{4p(1-\gamma)}{3} - \frac{4(1-\gamma)}{3} = -\frac{4(1-\gamma)(1+p)}{3} < 0$ ■

Furthermore, feedback also influences the agents' motivation. An individual who receives good feedback will become more ambitious in the future, in the sense that he will set more demanding goals for himself. This result comes through two separate channels. First, when agent i learns that agent j is less skilled he expects to do better in future relative comparisons, and as a result it is optimal for him to put greater importance (that is, choose a higher s_i) on such comparisons. Second, learning that agent i's own ability is higher translates to higher benefits of choosing a high standard, since the effort cost function decreases in the standard s_i. This is summarized in the following proposition:

Proposition 5 *When the agent's beliefs about relative ability are revised upwards (downwards), he will choose a higher (lower) standard.*

Proof. Using equation (9) we obtain

$\frac{ds_i^*}{dE_i(a_i)} = \frac{2(1-\gamma)}{3} > 0$ and $\frac{ds_i^*}{dE_i(a_j)} = -\frac{4(1-\gamma)}{3} < 0$ ■

The previous propositions indicate that feedback about relative rank has ex-ante and ex-post effects on beliefs and productivity in setting where agents still learn where they stand in the rank hierarchy. It is therefore natural to ask what would be the effect of feedback in well-established teams, that is, in settings where workers' abilities and feedback policies are common knowledge. We address this question below. Recall equation (3) and assume that there is common knowledge of abilities and feedback probability. We then get that in equilibrium:

$$e_i^* = \frac{\beta(1-p) + pk(1-q) + p\gamma a_j + pqa_i - pa_j - \gamma a_i}{(1-pq)} \quad (12)$$

$$y_i^* = \frac{\beta(1-p) + pk(1-q) + (1-\gamma)(a_i - pa_j)}{(1-pq)} + \widetilde{\varepsilon}_i \quad (13)$$

Therefore, the principal who hires a pair of workers (i,j) and provides worker i with feedback with probability p and worker j with probability q expects the following level of total output

$$Y^*(i,j) = y_i^* + y_j^* = \frac{\beta(2-p-q) + qk(1-p) + pk(1-q)}{1-pq} + \frac{(1-\gamma)}{(1-pq)}\left(a_i(1-q) + a_j(1-p)\right) \quad (14)$$

Proposition 6 *For a given q, if agent i is good enough relative to agent j (that is, if* $a_i \geq \frac{1}{q}\left(a_j - \frac{(k-\beta)(1-q)}{(1-\gamma)}\right)$ *) it is optimal for the principal to increase the frequency of feedback for worker i.*

Proof. $\frac{dY^*}{dp} = \frac{(1-q)}{(1-pq)^2}\left((k-\beta)(1-q) + (1-\gamma)(a_i q - a_j)\right)$. $\frac{dY^*}{dp} \geq 0 \Leftrightarrow a_i \geq \frac{1}{q}\left(a_j - \frac{(k-\beta)(1-q)}{(1-\gamma)}\right)$ ■

This proposition implies that a principal can extract more output from agents if he provides more frequent feedback to high ability workers. Feedback about relative rank is a cheap way to motivate the high types to work harder, since they enjoy learning that they did better than the competition.

A few comments regarding the modeling choice are in order. (1) An alternative specification would be to make the utility dependent on relative ability, not output. The problem with having a model where agents learn about their ability (a constant trait) is that this learning should take place relatively quickly. In the absence of other output-based rewards, there is no incentive to continue production after ability is known and therefore such model could not explain the increase in output over time observed in the data from our experiment. (2) In the economic literature standards or goals against which agents compare themselves have been traditionally assumed to be exogenous variables. Nonetheless, findings from social psychology (e.g. Wood & Taylor (1991), White et al. (1995)) as well as economics (Falk & Knell (2004)) suggest that goals are to some extent chosen by individuals. Therefore we model the standard s_i as a choice variable. An exogenous component of goal setting could easily be incorporated in our framework without changing the predictions of the model. However, since in the experimental setup we did not manipulate goals exogenously, we focus the model on the endogenously chosen standard. (3) As suggested by a reviewer, it is plausible that even in the absence of feedback people may still enjoy believing that they probably performed well in relative terms. In such a case, the likelihood of feedback should not affect effort decisions, contrary to Proposition 1 and our empirical findings. The relationship between effort and the likelihood of feedback could be restored by assuming a different utility function of the form $u_1 = y_1 + v(y_1 - y_2)$, with appropriate assumptions on the function v. (4) In many real life situations, as well is in our experimental setup, people compare themselves with a larger number of peers, and the model can be easily extended to allow for more than two agents. In such a setting where agent i receives feedback about the output of multiple opponents, his utility equals $y_i - \sum_{j \neq i} \frac{y_j}{n-1} \ln\left(\frac{k}{k-s_i}\right)$, where n is the number of agents performing the task. All propositions, with only minor modifications, continue to hold for this specification. For tractability and ease of exposition, we restrict our attention to the two-agent setting. (5) Note also that in a two-agent setting where each agent receives perfect feedback about their competitor's output, and where agents have the opportunity to interact after the task, social status concerns are likely to arise (i.e., people want to look good in front of the other person by having produced high output). However, our experiment was designed to eliminate social status concerns or peer pressure, since the large number of participants in the room made it impossible for a subject to attribute a particular level of output to specific individuals around them. Since each person's performance was only known to themselves, we believe concerns for social status are minimal. Given this feature of the experimental setup, and the model's aim to facilitate the interpretation of the data, we choose not to incorporate the social status aspect in the theoretical framework.

B Appendix: Experiment Instructions

Welcome to our experiment on economic decision making! The study will last about 60 minutes, during which you will participate in a 45-minute experiment, and will fill out some questionnaires. Your task during the experiment is to solve multiplication problems. Each time you provide a correct answer one point is added to your score. Your score is refreshed in each period and you are going to play for 18 periods. In each of the periods:

1)You will be told what information you will receive at the end of the period regarding your rank in the group. Your rank is based on the number of correct answers provided by you and the other participants. You will see one of the following three statements on the screen, selected at random for each one of the participants in each period: **"You WILL see the ranking this period."** In this case, you will see the rank information at the end of the period. **"You MAY see the ranking this period."** In this case, there is an equal chance that you will or will not see the rank information at the end of the period. **"You WILL NOT see the ranking this period."** In this case, you will not see the rank information at the end of the period.

2)You will be asked to estimate your rank in the group, before seeing any of the multiplication problems. Your rank is determined by your score in the current period. If you have the highest score (i.e., nobody solved more multiplication problems than you did), you will rank as number 1. If there is only one person who solved more problems you will rank as number 2, and so on. Therefore, if you expect that x people will have higher score than yours, please type in a number equal to $x + 1$ as your expected rank and press the "Submit" button. *Example: You expect that 5 people will do better than you. Type in 6 and press "Submit".*

3) You will be presented with multiplication problems to solve. In each period you will have 90 seconds during which you can work on the multiplication problems. To provide an answer, type it in the box and press "Submit". If your answer is correct a point will be added to your score and you will see another multiplication problem. If your answer is incorrect, your score will remain unchanged and you will see the message "Incorrect. Please try again". You will be asked to solve the same problem again. Only after you provide correct answer the program will move on to the next multiplication problem.

4) You will be asked to report the level of effort you have put into doing the task during that period. Check the appropriate field that reflects how much effort you have put into doing the task, ranging from "no effort at all" to "a lot of effort", then press "Submit".

5)You may see how you have ranked relative to others during the period, depending on what you were told in the beginning of the period (see (1)) If the ranking information is provided to you this round, you will have 15 seconds to see it. The ranking is presented in such a way that every participant can identify only his/her own score. In other words, your exact ranking for that period will be known to you only. No other participant can see how you ranked that period. *Example: There are 10 participants. You solved 3 problems and five people did better than you. The screen that you will see may look like this:*

This period is over! Ranking in this period:

Rank	Name	Score
1	.	10
1	.	10
3	.	9
4	.	8
5	**You**	**3**
5	.	3
5	.	3
5	.	3
9	.	1
9	.	1

In case you do not see the ranking you will be asked to wait for 15 seconds for the experiment to continue. Then, the experiment moves on to the next period and all the stages are repeated. In the end of the experiment we will ask you to fill in a short questionnaire.

Payment You will receive a total of **$23** in cash for your participation in our study.

Practice periods You will have a chance to practice this task for one period. We encourage you to type in at least one correct and one incorrect answer so that you know how to behave in both cases. You will not see any ranking information in the practice period.