

Rank Expectations, Feedback and Social Hierarchies

Camelia M. Kuhnen*

Agnieszka Tymula[†]

May 2008 (Preliminary version)

Abstract

We investigate the role of self-esteem, generated by private feedback regarding relative performance, on the behavior of agents working on a simple effort provision task for a flat wage. We isolate the impact of learning one's rank in the group on one's output from any reputation, strategy-updating or peer monitoring effects.

Feedback has both ex-ante and ex-post effects on the productivity of workers and on the dynamics of social hierarchies. 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 remains relatively stable afterwards. Private relative rank information helps create a ratcheting effect in the group's average output, which is mainly due to the fight for dominance at the top of the hierarchy.

These results suggest that in environments where monetary incentives are weak, moral hazard can be mitigated by providing feedback to agents regarding their relative performance, and by optimally choosing the reference peer group. Therefore, social hierarchy effects on productivity are important for optimal team formation and dynamics.

*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.

[†]Department of Economics, Bocconi University, agnieszka.tymula@phd.unibocconi.it, via Sarfatti 25, 20 136 Milano, Italy, agnieszka.tymula@phd.unibocconi.it

1 Introduction

Self-esteem has long been thought of in the psychology literature as a strong motivator of human behavior (Maslow (1943), McClelland, Atkinson, Clark, and Lowell (1953)). Recently, this concept has been introduced in theoretical models of economic choice (Benabou and Tirole (2002), Koszegi (2006)) as “ego utility”. People derive utility from thinking of themselves as good, skilled or valuable according to some social criteria, and their actions are shaped by the desire to maintain high levels of self-esteem.

So far, the economics literature on ego utility has focused on understanding the role of self-esteem on behavior in non-competitive settings. However, ego utility may also affect strategic interactions, where self-esteem is determined by an individual’s beliefs about his relative standing among his peers, and not necessarily by beliefs about absolute measures of his ability. In such settings, beliefs about relative rank are modified by the feedback that individuals receive about their relative performance. Therefore, 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 literature¹, it still remains to be seen which predictions from tournament theory apply to settings where the prize is simply ego utility, or self-esteem. This is at the core of the issue we seek to address.

Specifically, our goal is to understand the role of ego utility on productivity in competitive settings where participants receive feedback about their relative standing. We isolate the ego utility effect from any other reasons why feedback about rank may change behavior. For instance, feedback may influence productivity if compensation is performance-based², since people seem to care more about their relative, rather than objective level of wealth (Clark and Oswald (1996), Easterlin (1995), Luttmer (2005)). Feedback may also change behavior if it provides information about the optimal course of action in subsequent stages of the project (Seta (1982), Bandura (1986), Major, Testa, and Bylsma (1991), Kluger and DeNisi (1996)). 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 participants’ behavior going forward (Kandel and Lazear (1992), Knez and Simester (2001), Falk and Ichino (2006), Mas and Moretti (2007)). To minimize the influence of these other channels through which relative rank

¹See Prendergast (1999) for a review of that literature.

²Whether or not this holds true for all compensation schemes is unclear, as suggested by Eriksson, Poulsen, and Villeval (2008).

information may influence actions, we employ a setting where participants receive a flat wage, the task that they work on does not involve changes in strategy or learning, 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.³ 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 and Harkins (1987), White, Kjelgaard, and Harkins (1995)). Thus, individuals are willing to exert more costly effort to avoid falling behind the average or to be better than the average. In the context of a search experiment, Falk, Huffman, and 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 the task, consistent with the idea that a low rank decreases utility. There is also recent neurological evidence that social comparison matters for well-being. Fliessbach, Weber, Trautner, Dohmen, Sunde, Elger, and Falk (2007) show that the brain encodes as rewarding instances when an individual learns that he has received higher pay than another participant in the experiment, for solving the same task.

Our experimental results show that private feedback about relative ranking has both ex-ante and ex-post effects on the productivity of workers and on the dynamics of social hierarchies. 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 average output. This ratcheting effect (working harder over time) is mainly due to the fight for dominance at the top of the hierarchy. Moreover, increasing the heterogeneity in the ability of members of the peer group leads to lower output from low ability individuals, but has no impact on the output of high ability

³In other words, relative rank has “consumption value”, a term for ego utility coined by Benabou and Tirole (2002).

workers.

In a related paper, Eriksson, Poulsen, and Villeval (2008) measure output and effort levels across subject groups that face one of two variable compensation schemes — piece-rate or tournament pay— and find that releasing information about relative performance does not significantly influence the subjects’ average output or effort in either pay condition. These findings may indicate that strong monetary incentives crowd out the effect of feedback that we demonstrate using a flat-wage environment.

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, is a motivator for productivity. In light of these findings, it is possible that by changing the reference peer group, a social planner or principal can benefit from the dynamics of social hierarchy effects on productivity.

Rankings are commonly used in many environments – for example, in the labor market for corporate executives or fund managers, in educational institutions or sales departments. Institutions that publish rankings are usually concerned with the performance of their members. Therefore, understanding what impact ranking may have on performance is of key importance to the motivational politics of a modern firm.

2 Literature review

This paper contributes to several strands of literature that we review below: ego utility, intrinsic motivation, peer effects, tournaments and gender differences in competitive environments.

Ego utility/self-esteem

Benabou and Tirole (2002) focus on the effect of self-esteem on the behavior of people with time inconsistent preferences (hyperbolic discounters). They argue that self-confidence is valuable because it enhances motivation to act, and investigate a variety of intrapersonal strategies people may use to enhance their self-image. They show that people may handicap their performance (for example by exerting low effort), and use self-deception through selective memory or awareness management in order to maintain high self-perception about their ability. This keeps them motivated to undertake profitable endeavors in the future. Weinberg (1999) and Koszegi (2006), on the other hand, treat self-esteem as a consumption good and incorporate it directly in the utility function.

They assume that individuals' utility is increasing with their perception of their own ability, which is updated in a Bayesian manner after receiving new relevant information.

It is worth to point out that none of the above mentioned models takes into account the fact that in most real life situations people exert externalities on one another. Usually, one's self-perception is not shaped in isolation but is also influenced by the actions of others. If one person works harder he increases his probability of success (and thus of having high self-esteem) and at the same time decreases the probability of success of people around him. Therefore, it is not clear what would happen when people compete for high self-esteem. Moreover, subjects may differ with respect to how much they value their self-image in a particular task and also with respect to their ability to solve the task.⁴

Thus, the predictions of extant theoretical models regarding people's reaction to relative rank information in the absence of monetary incentives are not clear. When feedback is provided, ex-ante concerns for self-image can increase effort, as agents seek to learn that they rank high. However, the prospect of receiving feedback can also lead to lower ex-ante effort, because of disappointment avoidance. As suggested by Koszegi (2006), agents with positive beliefs about themselves wish to preserve their self-esteem and may decide to avoid competing, because doing so reduces the informativeness of signals about ability obtained during the task.

Ex-post effects of feedback are also difficult to predict based on existent theories. After receiving bad feedback, high anxiety and stress levels could undermine subsequent performance (Rauh and Seccia (2006)). Moreover, people with self-image concerns could employ deception strategies as suggested in Benabou and Tirole (2002) in order to discard or interpret this information to their advantage. However, it remains unclear if they would continue to engage in the task in the future. The literature on biased tournaments would suggest that after receiving bad feedback people may give up if the perceived chances of winning in the future are minimal. Koszegi (2006), on the other hand, suggests people could engage in the task again because it is the only way to regain self-confidence. The impact of positive feedback on future actions is also unclear, from a theoretical standpoint. People may be more likely to continue to work hard because they believe their chances of winning are high. They may also give up and not compete anymore, in order to reduce the informativeness of future rankings, and thus conserve their currently high self-esteem. Individuals may also lower their future efforts if they learn they are far

⁴The literature on tournaments with heterogeneous agents will help us model this setting.

better than their competitors, as suggested by the literature on biased tournaments.

Related to the work on self-esteem is a large literature on the value of public recognition, or status. People care about social status as defined by their relative income (Frank (1984), Frank (1985)). Moreover, people value public recognition independently of any monetary consequence and are willing to trade off material gains to obtain it (Huberman, Loch, and Onculer (2004)).

Cowen and Glazer (2007) consider the role of self-esteem in behavior in the labor market. They assume that people care about evaluations in an asymmetric way, as a loss of esteem pains an individual more than a gain in esteem satisfies him, and suggest implications for wage and promotion schemes, and job search and sorting. Importantly, in their model, changes in a person's self-esteem require that others learn about that individual's type or ability.

Clark, Masclet, and Villeval (2006) use survey data and experimental evidence from a gift-exchange game to determine the effect of status or relative income on work effort, and find a strong effect of others' incomes on individual effort decisions in both datasets. The individual's rank in the income distribution has a more powerful effect on effort than does others' average income, suggesting that comparisons are more ordinal than cardinal. Clark, Masclet, and Villeval (2006) further show that, controlling for own income and income rank, the width of the relevant income distribution matters, with effort increasing in the distance from the bottom of the income distribution. Last, effort is also affected by comparisons over time: those who received higher income offers or had higher income rank in the past exert lower levels of effort for a given current income.

Intrinsic and extrinsic motivation

Our focus is on the internal drive of individuals to rank well relative to others. People enjoy winning (or learning they rank high in some social hierarchy) even in situations when this 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)). Importantly, this effect is independent of the social context. Benabou and 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. A reduced-form approach to this topic is presented by Frey (1997). The interplay between the two types of incentives is generated by the assumption that extrinsic motivators (e.g.

bonuses) convey information about the agent's ability or about the difficulty of the task, and hence influence the agent's intrinsic interest in making the project successful.

Peer effects

The social environment is an important determinant of individuals' actions, as shown by the recent and growing literature on peer effects⁵. The existing evidence shows that typically individuals with lower ranks improve in the presence of peers. Mas and Moretti (2007) find out that most of peer effects come from the increased output of low productivity workers, who benefit from the presence of high ability workers. Seta (1982) and Bandura (1986) show that performance on a particular task is improved in the presence of someone who does the task better or is more successful. Major, Testa, and Bylsma (1991) in their review on self-improvement conclude that having peers who perform better may serve as reference standard that serves as a self-improvement device.

Falk and Ichino (2006) find evidence that peer effects increase productivity in an experiment where subjects are asked to fill letters into envelopes, and receive pay independent of output. When subjects work in randomly-assigned pairs, average output is higher than when subjects work alone. Moreover, the standard deviations of output are smaller within pairs than between pairs of subjects, and low-productivity workers are the most sensitive to the behavior of peers. Peer pressure (monitoring) has been proposed to be an effective incentive mechanism by Kandel and Lazear (1992), and evidence in support of this idea is presented in Mas and Moretti (2007) and Knez and Simester (2001). Moreover, studies like Falk and Ichino (2006) and Mas and Moretti (2007) suggest that relative performance matters even when it does not have an impact on monetary payoffs, since flat wages are employed in both studies.

Bandiera, Barankay, and Rasul (2007) use field data to show that social incentives, namely the presence of friends, affect worker's performance. Friends conform to a common productivity norm that lies between the typical performances of the most and least able friends. Falk, Fischbacher, and Gächter (2004) find experimental evidence suggesting that people belonging to the same group often behave similarly (in this case, when deciding contributions towards a public good), even when group assignments are random.

Tournaments

If individuals get more utility from ranking higher in a hierarchy, then we can view their actions as a tournament, where the external incentives are zero (i.e. flat-wage

⁵For evidence on the endogeneity of peer group choice, see Evans, Oates, and Schwab (1992) and Foster (2006).

compensation). The prize is simply the private knowledge that one is better than the rest. This prize may be increasing in the number of competitors, since individuals may value ranking first in a small group less than ranking first in a large group. If this assumption is correct, then certain predictions from tournament theory may apply in our setting: people may put in more effort when faced with a larger number of competitors, and effort levels may be lower when group members are more heterogeneous in their abilities. However, it is possible that tournament theory does not apply in our setting, since the theory requires that the prize is the same for all participants, and it is known ex-ante by everybody. In our experiment, participants may be heterogeneous in how much they value ranking high, and the distribution of their preferences over high rank is not common knowledge.

Gender differences related to self-esteem and ability to compete

Huberman, Loch, and Onculer (2004) observe that males seek status more than women. Falk and Knell (2004) asked students from the University of Zurich to indicate their aspired diploma degree as a measure of their reference standard. High school grades were taken as a proxy for their abilities. They show that, controlling for ability and socioeconomic characteristics, female students have significantly lower reference levels than male students and that more able individuals set higher standards for themselves than less able agents.

Gneezy, Niederle, and Rustichini (2003) propose and find experimental evidence in support of the idea that women may be less effective than men in competitive environments, even if they are able to perform similarly in noncompetitive environments. They find that increasing the competitiveness of the environment leads to a significant increase in performance for men, but not for women. This effect is stronger when women have to compete against men than in single-sex competitive environments. Niederle and Vesterlund (2007) give subjects the choice of getting compensated for solving an effort task via either a noncompetitive piece rate or a competitive tournament incentive scheme. They find that although there are no gender differences in performance, men select competitive settings twice as much as women when choosing their compensation scheme.

In light of these results, we use the gender composition of the subject group as a proxy for the heterogeneity in the subjects' ability to compete in the experimental task.

3 Experimental design

The ideal dataset for understanding the role of private feedback regarding relative rank on productivity would allow us to compare workers' output when such feedback is provided and when it is not provided, all other things being equal. It would also describe the workers' personal characteristics and rank expectations. It is hard, if not impossible, to obtain such data from the field. First, feedback must be privately observed. Second, performance must not be influenced by other confounding factors such as incentive schemes, peer pressure and career or status concerns. Hence, we use a controlled experimental setting to test our theory.

In our experiment we ask subjects to solve simple multiplication problems (multiply one-digit numbers by two-digit numbers) during multiple, identically structured rounds. This means that subjects make "real effort" choices (as opposed to other experiments where subjects are asked to choose a number at a cost). We use this task because it has a number of advantages. First of all, no previous knowledge is required and it is easy to explain. Secondly, task learning effects, which we would like to avoid, should be minimal. Moreover, the score on this task depends on the subject's ability as well as on the effort choice. Therefore, different subjects will end up with different scores, which will lead to a dispersed ranking. Also, 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. We are also able to assess whether this response differs when feedback is positive (i.e., the subject learns that he did better than expected) and negative (i.e., the subject learns that he did worse than expected).

In order for our behavior measures to be meaningful, it is necessary to control for the difficulty level of the multiplication problems. If randomly generated numbers were used and the worker solved more problems in round two than in round three this could mean two things. Either the worker's effort remained the same across the rounds but the problems in period two were easier, or he/she worked harder in round two while the problems were equally difficult in both rounds. We generated 206 multiplication problems of the same difficulty level, as in Cromer (1974)⁶ in order to avoid this possible confound.

⁶Examples of problems used are: $89 * 4$, $76 * 9$, $73 * 8$.

The problems were presented to the subjects on the computer screen. Each time the subject solved the multiplication correctly one point was added to his/her score and the next problem was presented. Each time the subject provided a wrong answer the score remained unchanged and the subject was asked to solve the same problem again until he/she answered correctly. By not allowing subjects to move on to the next question unless the previous one was solved, we avoid a situation where a part of the subjects strategically skips more difficult problems looking for easier ones. Subjects received a fixed fee of 23 US dollars for their participation independent of performance.

The experiment consisted of 18 rounds. Each of them had the same structure but there were three possible feedback conditions. The conditions differed with respect to the probability with which the subject received feedback about his/her relative rank at the end of the round. This probability was either 0, 0.5 or 1. We will refer to these treatments as the “No”, “Maybe” or “Sure” feedback conditions, respectively. The feedback condition was determined randomly and independently for every subject at the beginning of each round.⁷ Therefore, in the same round different subjects faced different feedback conditions.

In the beginning of each round (i.e. stage one) subjects were informed which feedback condition they were in. This information was consistent with what happened at the end of the period and subjects were aware of that. We informed subjects about the feedback condition in the beginning of each round because it allows us to study the ex-ante effect of feedback probability on rank, expected rank, output and effort choices.

In stage two subjects were asked to report their expected rank in that round.⁸ After that, in stage three subjects had 90 seconds to work on multiplication problems. For each subject, their score was displayed on the screen throughout the round and was updated after every correct answer. The score was reset to zero at the beginning of every round. In stage four subjects were asked to assess how much effort they had put into the task that round. Answers were provided using a six point scale ranging from “no effort at all” to “a lot of effort”.

⁷Using alternating messages within one session allows us to control for session effects.

⁸We did not pay subjects if their rank expectations turned out to be correct at the end of the round, because doing so would have distorted behavior: all subjects would have declared that they rank last, solved zero problems, and achieved 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 made 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 than others) changes behavior we are confined to a flat-wage environment.

In the final stage of each round, that lasted for fifteen seconds, each subject would either see the relative ranking feedback or not, depending on the feedback condition they had been assigned to for that round. The ranking was determined by the current period scores of all subjects in the group. The subject that solved the highest number of problems would rank as number one, the one whose score was lower than scores of two other subjects would rank as number three, etc. Each subject could see the scores and ranks of all the participants but he could identify *only* his rank and score. Therefore each subject knew that nobody else could associate his identity to his actual rank and score.

At the end of the experiment subjects were asked to fill out a short questionnaire about their socioeconomic characteristics (age, gender) and education history (years of schooling, courses taken) and completed two personality tests: The Big Five Inventory and Achievement Motivation Scale, to measure traits such as neuroticism, extraversion, openness, conscientiousness, agreeableness and the need for achievement.

The experiment was programmed using the z-Tree software (Fischbacher (2007)). Subjects were given a written copy of the instructions (see Appendix) which they were asked to read before the experiment started. The task was also described verbally by the experimenter. Subjects practiced the task for one period, but feedback was not provided during that time. No external aids (calculators, scratch paper, etc.) were allowed. 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 remaining 54 subjects (24 male and 30 female), in seven sessions. Each of these subject groups consisted of six to nine people.

4 Results

4.1 Ex-ante effects of feedback

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 rank, as well as on their actual output, measured as the number of multiplication problems solved correctly. These effects are illustrated in Fig. 1.

Output is on average higher (11.35 vs. 10.58, $p < 0.07$ in a one-sided mean comparison test), and the expected rank is better (4.16 vs 4.90, $p < 0.001$ in a one-sided mean

comparison test) for participants who are in the “Maybe” feedback condition, than for those in the “No” feedback condition. There is no significant difference between the output or expected rank of subjects in the “Maybe” feedback condition versus “Sure” feedback condition.

Fig. 2 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 12.91 for men, and 8.69 for women ($p < 0.0001$ in a one-sided mean comparison test). This result is in line with the prior literature on gender and competitiveness (e.g. Gneezy, Niederle, and 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 3.53, while women expect to receive a rank of 5.53. The difference is statistically significant ($p < 0.001$ in a one-sided mean-comparison test).

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 0.58 in the first six periods, 0.82 in periods seven through twelve, and 0.84 in periods thirteen through eighteen ($p < 0.0001$ in all cases).

4.2 Ex-post effects of feedback

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.

Table 1 shows the role of received feedback on future output, expectations of rank, and actual rank. Doing better than expected in round $t - 1$ (i.e. $GoodFeedback_{t-1}=1$) leads the subjects to expect a better rank 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 (i.e. a higher value for $ExpectedRank_t$). Both of these effects are measured relative to receiving neutral feedback.

Interestingly, while ranking information seems to make well-performing subjects think

they will rank even better in the future, and badly-performing subjects think they will rank worse, the opposite actually happens. After receiving negative feedback, people solve more problems, and achieve a better rank. After receiving positive feedback, output is lower and the actual rank worsens. These effects are measured relative to receiving neutral feedback. I control for the prior values of expected rank, output and actual rank to account for the mechanical effect that people who are top ranked can only move higher in the rankings, whereas people who are already at the bottom of the hierarchy can not rank any lower.

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 1 and 2. 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.

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 3, 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 test more formally whether stable hierarchies do get formed, and if so, how soon it happens. Fig. 3 shows evidence that hierarchies indeed emerge, and that effort is sustained even after the social dominance order is established. First, the data indicate that output grows over time. 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. As we argued earlier, we believe learning effects are

minimal given the nature of the task.

Moreover, we find that the standard deviation of output increases over time, 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, 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 calculated 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 these three stages, we assigned 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 compared 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 4 and 5 show how people transition across rank performance bins as the task progresses. Fourteen of the seventeen (82%) of the individuals who were 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 were 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 had in 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, as shown in Figure 3. 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.

Figure 6 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 productivity that is shown in Figure 3 comes from high productivity subjects who fight to maintain or improve their rank. A recent quote⁹ by Vijay Singh, who was the number one player in the Official World Golf Rankings in 2004 and 2005, illustrates this ratcheting effect: "I'm playing pretty good now, but my ranking doesn't say that. I'm number two."

We also find that while group size does not matter for the average output level, the ex-ante dispersion in skill does. We proxy heterogeneity in abilities by the gender composition of our subject groups. As shown by the results in Table 4, 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. These findings complement the results of Gneezy, Niederle, and Rustichini (2003) and Niederle and 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.

5 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

⁹<http://www.brainyquote.com/quotes/quotes/v/vijaysingh183223.html>

theory from alternative explanations as to why rank information changes behavior, we employ an experimental setting where subjects receive a flat wage for working on a simple multiplication problem solving task, and where there can not exist status, reputation, strategy-learning or peer monitoring effects.

We find that agents increase output, and expect to rank better, if they think 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 of rankings on output. People at the top of the hierarchy early on work harder over time to maintain that position, while people at the bottom do not change their productivity level over time.

These findings suggest that private rank information can be used as a motivator to increase firm productivity. Teams, or peer groups, can be reshuffled to allow low-rank workers to potentially climb the hierarchy in another group, and as a result, to generate more output.

References

- Bandiera, O., I. Barankay, and I. Rasul (2007). Social incentives in the workplace. *Working paper, London School of Economics*.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Benabou, R. and J. Tirole (2002). Self confidence and personal motivation. *Quarterly Journal of Economics*.
- Benabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies* 70, 489–520.
- Clark, A., D. Masclet, and M.-C. Villeval (2006). Effort and comparison income : Survey and experimental evidence. *Working Paper*.
- Clark, A. E. and A. J. Oswald (1996). Satisfaction and comparison income. *Journal of Public Economics* 61(3), 359–381.
- Cowen, T. and A. Glazer (2007). Esteem and ignorance. *Journal of Economic Behavior & Organization* 63, 373–383.
- Cromer, F. E. (1974, May). Structural models for predicting the difficulty of multiplication problems. *Journal for Research in Mathematics Education* 5(3), 155–166.
- Deci, E. (1975). *Intrinsic motivation*. New York, Plenum Press.
- Easterlin, R. A. (1995, June). Will raising the incomes of all increase the happiness of all? *Journal of Economic Behavior & Organization* 27(1), 35–47.
- Eriksson, T., A. Poulsen, and M. C. Villeval (2008, April). Feedback and incentives: experimental evidence. *Working Paper*.
- Evans, W. N., W. E. Oates, and R. M. Schwab (1992, October). Measuring peer group effects: a study of teenage behavior. *The Journal of Political Economy* 100(5), 966–991.
- Falk, A., U. Fischbacher, and S. Gächter (2004, November). Living in two neighborhoods - social interactions in the lab. *IZA Discussion Paper No. 1381*.
- Falk, A., D. Huffman, and U. Sunde (2006, December). Self-confidence and search. *IZA Discussion Paper*.

- Falk, A. and A. Ichino (2006). Clean evidence on peer effects. *Journal of Labor Economics* 24, 39–57.
- Falk, A. and M. Knell (2004). Choosing the joneses: Endogenous goals and reference standards. *Scandinavian Journal of Economics* 106(3), 417–435.
- Fischbacher, U. (2007). Z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2), 171–178.
- Fliessbach, K., B. Weber, P. Trautner, T. Dohmen, U. Sunde, C. Elger, and A. Falk (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science*, 1305–1308.
- Foster, G. (2006). It’s not your peers, and it’s not your friends: some progress toward understanding the educational peer effect mechanism. *Journal of Public Economics* 90(8-9), 1455–1475.
- Frank, R. H. (1984). Interdependent preferences and the competitive wage structure. *RAND Journal of Economics* 15, 510–520.
- Frank, R. H. (1985). *Choosing the right pond: Human behavior and the quest for status*. New York: Oxford University Press.
- Frey, B. (1997). *Not just for the money – An economic theory of personal motivation*. Cheltenham, U.K.: Edward Elgar.
- Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: gender differences. *Quarterly Journal of Economics* 118(3), 1049–1074.
- Huberman, B. A., C. H. Loch, and A. Onculer (2004). Status as a valued resource. *Social Psychology Quarterly*, 103–114.
- Kandel, E. and E. P. Lazear (1992). Peer pressure and partnerships. *Journal of Political Economy* 100(4), 801–817.
- Kluger, A. N. and A. 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.
- Knez, M. and D. Simester (2001). Firm-wide incentives and mutual monitoring at continental airlines. *Journal of Labor Economics*, 743–772.
- Koszegi, B. (2006). Ego utility, overconfidence and task choice. *Journal of the European Economic Association*.

- Luttmer, E. (2005). Neighbors as negatives: relative earnings and well-being. *Quarterly Journal of Economics* 120(3), 963–1002.
- Major, B., M. Testa, and 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, A. and E. Moretti (2007). Peers at work. *Working paper.*
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 370–396.
- McClelland, D., J. Atkinson, R. Clark, and E. Lowell (1953). *The achievement motive.* Princeton: Van Nostrand.
- Niederle, M. and L. Vesterlund (2007, August). Do women shy away from competition? do men compete too much? *Quarterly Journal of Economics* 122(3), 1067–1101.
- Prendergast, C. (1999, March). The provision of incentives in firms. *Journal of Economic Literature* 37, 7–62.
- Rauh, M. T. and G. Seccia (2006, May). Anxiety and performance: an endogenous learning-by-doing model. *International Economic Review* 47(2).
- Seta, J. (1982). The impact of comparison processes on coactors' task performance. *Journal of Personality and Social Psychology* 42.
- Szymanski, K. and S. 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, P. H., M. M. Kjelgaard, and S. G. Harkins (1995, July). Testing the contribution of self-evaluation to goal-setting effects. *Journal of Personality and Social Psychology* 69, 69–79.

Appendix: 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, for sure 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, for sure 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.

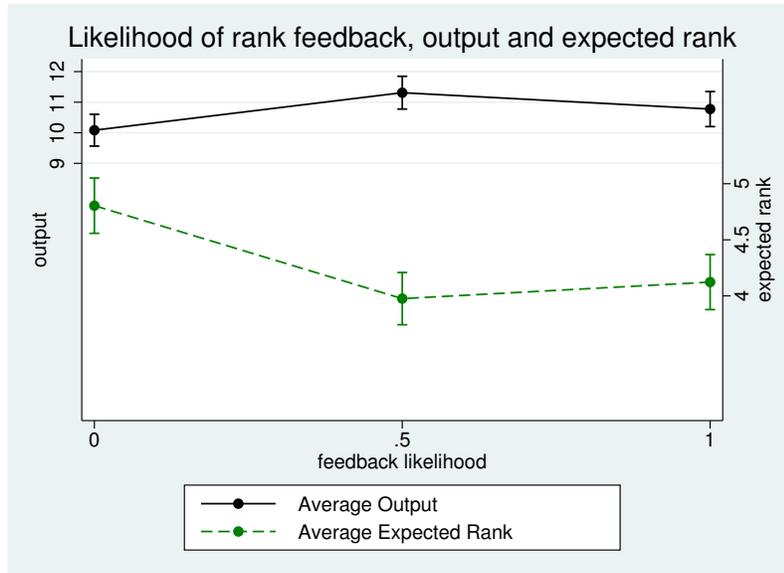


Figure 1: Feedback likelihood, output and expected rank

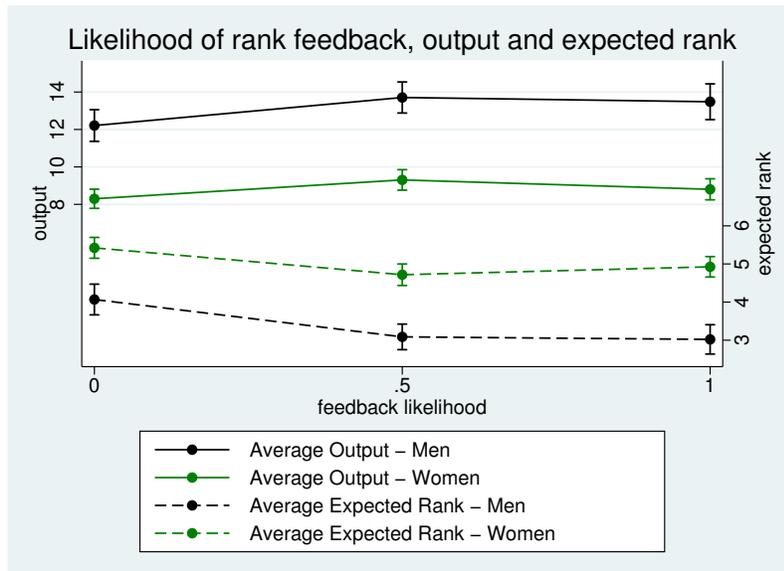


Figure 2: Feedback likelihood, output and expected rank, by gender

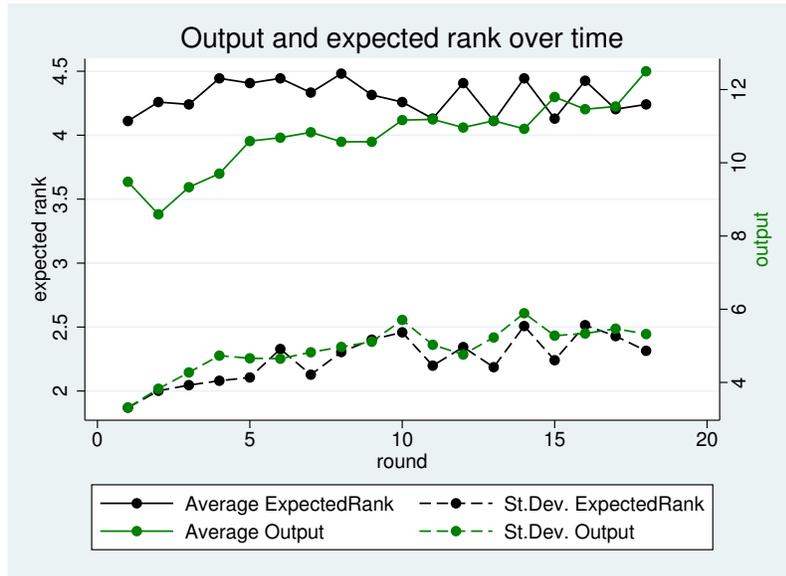


Figure 3: Output and expected rank, by round

	Low ⁷⁻¹²	Medium ⁷⁻¹²	High ⁷⁻¹²
Low ¹⁻⁶	13/17	4/17	0/17
Medium ¹⁻⁶	4/18	12/18	2/18
High ¹⁻⁶	0/19	2/19	17/19

Figure 4: Transitions across ranks: rounds 1-6 to rounds 7-12.

	Low ¹³⁻¹⁸	Medium ¹³⁻¹⁸	High ¹³⁻¹⁸
Low ¹⁻⁶	13/17	4/17	0/17
Medium ¹⁻⁶	4/18	11/18	3/18
High ¹⁻⁶	0/19	3/19	16/19

Figure 5: Transitions across ranks: rounds 1-6 to rounds 13-18.

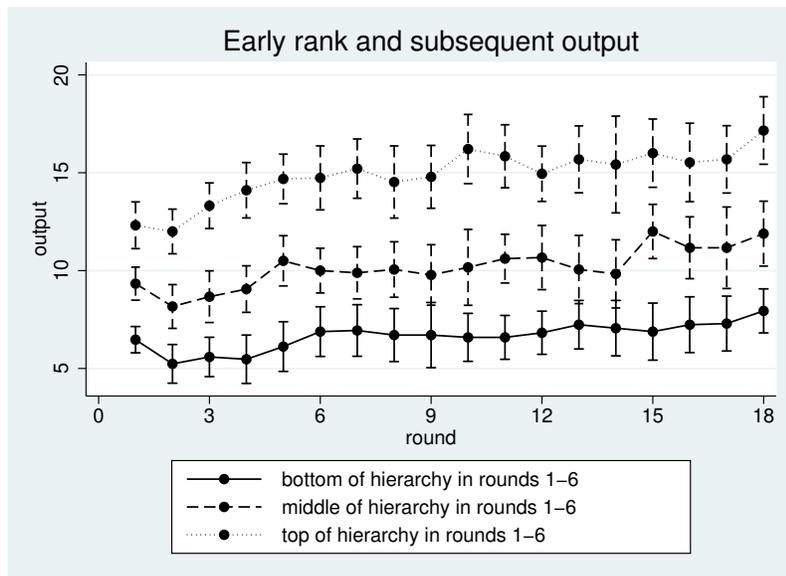


Figure 6: The average output produced each round by subjects who were at the top, in the middle or at the bottom of the rank hierarchy during the first six rounds.

Table 1: The ex-post impact of feedback on estimated rank, actual rank and effort. $Output_t$ 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.

	$Output_t$	$ExpectedRank_t$	$Rank_t$
	Coef./t	Coef./t	Coef./t
$GoodFeedback_{t-1}$	-0.76 (-2.55)**	-0.50 (-4.56)***	0.63 (3.54)***
$BadFeedback_{t-1}$	0.74 (2.19)**	0.54 (4.17)***	-0.38 (-2.26)**
$ExPostFeedback_{t-1}$	-0.12 (-0.36)	0.23 (1.49)	-0.07 (-0.44)
$FeedbackLikely_t$	0.56 (1.52)	-0.55 (-2.81)***	-0.31 (-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 (3.33)***	-0.39 (-2.19)**	-0.69 (-2.63)**
$Round_t$	0.05 (3.79)***	-0.01 (-1.49)	-0.00 (-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 included

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 2: Ex-post impact of feedback on estimated rank, actual rank and effort – alternative specification.

	<i>Output_t</i> Coef./t	<i>ExpectedRank_t</i> Coef./t	<i>Rank_t</i> Coef./t
<i>GoodFeedback_{t-1}</i>	-0.88 (-3.08) ^{***}	-0.27 (-1.54)	0.56 (3.50) ^{***}
<i>BadFeedback_{t-1}</i>	0.62 (2.26) ^{**}	0.77 (5.17) ^{***}	-0.45 (-2.66) ^{**}
<i>NeutralFeedback_{t-1}</i>	-0.12 (-0.36)	0.23 (1.49)	-0.07 (-0.44)
<i>FeedbackLikely_t</i>	0.56 (1.52)	-0.55 (-2.81) ^{***}	-0.31 (-1.79) [*]
<i>Output_{t-1}</i>	0.75 (13.43) ^{***}		
<i>ExpectedRank_{t-1}</i>		0.79 (13.52) ^{***}	
<i>Rank_{t-1}</i>			0.66 (10.15) ^{***}
<i>Male</i>	1.35 (3.33) ^{***}	-0.39 (-2.19) ^{**}	-0.69 (-2.63) ^{**}
<i>Round_t</i>	0.05 (3.79) ^{***}	-0.01 (-1.49)	-0.00 (-0.20)
Adj. R^2	0.664	0.701	0.540
No. of obs	918	918	918

Table 3: 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 (0.19)	-0.73 (-4.19)***	0.25 (0.94)	-1.70 (-3.82)***	-0.22 (-1.12)	1.03 (4.26)***
$BadFeedback_{t-1}$	1.20 (2.71)***	0.89 (4.98)***	-0.48 (-1.84)*	0.40 (0.79)	0.24 (1.23)	-0.32 (-1.36)
$ExPostFeedback_{t-1}$	-0.39 (-1.08)	0.09 (0.70)	0.05 (0.24)	0.17 (0.34)	0.31 (1.39)	-0.18 (-0.76)
$FeedbackLikely_t$	0.41 (1.39)	-0.31 (-1.82)*	-0.26 (-1.52)	0.75 (1.43)	-0.75 (-3.10)***	-0.37 (-1.61)
$Output_{t-1}$	0.81 (16.60)***			0.70 (10.42)***		
$ExpectedRank_{t-1}$		0.82 (16.74)***			0.77 (10.64)***	
$Rank_{t-1}$			0.69 (10.35)***			0.63 (8.64)***
$Male$	0.85 (2.20)**	-0.35 (-2.63)**	-0.59 (-2.07)**	1.80 (3.73)***	-0.42 (-1.88)*	-0.79 (-2.91)***
$Round_t$	0.08 (1.66)	-0.01 (-0.73)	-0.02 (-0.94)	0.06 (1.78)*	0.00 (0.04)	0.00 (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 included

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4: Impact of heterogeneity in subjects' competitive abilities on their estimated rank, actual rank and effort. 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 $Rank = 1$).

	Panel A: Women Only			Panel B: Men Only		
	$Output_t$ Coef./t	$ExpectedRank_t$ Coef./t	$Rank_t$ Coef./t	$Output_t$ Coef./t	$ExpectedRank_t$ Coef./t	$Rank_t$ Coef./t
$MenInGroup_t$	-1.10 (-2.59)**	0.37 (1.69)*	0.60 (2.68)**	0.25 (0.32)	-0.20 (-0.50)	-0.10 (-0.25)
$GroupSize_t$	0.42 (0.97)	0.29 (1.25)	0.19 (0.86)	-1.33 (-1.13)	0.66 (1.20)	0.63 (1.18)
$Round_t$	0.11 (4.57)***	0.01 (0.72)	0.01 (0.88)	0.22 (5.53)***	-0.02 (-1.00)	0.00 (0.17)
$Constant$	7.85 (2.97)***	1.47 (1.06)	1.56 (1.10)	20.75 (2.46)**	-0.92 (-0.32)	-1.33 (-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

* $p < .10$, ** $p < .05$, *** $p < .01$