Rank expectations, feedback and social hierarchies

Kuhnen, Camelia M. and Tymula, Agnieszka

Northwestern University – Kellogg School of Management

2008

Rank Expectations, Feedback and Social Hierarchies*

Camelia M. Kuhnen†

Agnieszka Tymula‡

This version: January 2009. First version: May 2008

Abstract

We develop and test experimentally a theoretical model of the role of self-esteem, generated by private feedback regarding relative performance, on the behavior of agents working on an effort provision task for a flat wage. 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. 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. 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. Hence, in environments where monetary incentives are weak, moral hazard may be mitigated by providing feedback to agents regarding their relative performance, and by optimally choosing the reference peer group.

*We thank Pierpaolo Battigalli, Shimon Kogan, Jonathan Parker, Paola Sapienza and participants at the 2008 ESA European Meeting, 2008 La Pietra-Mondragone Theory Workshop, 2008 International ESA Conference and 2008 NBER Behavioral Finance Meeting for valuable comments and discussion.

†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 literature[1], 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 two gaps in the literature.

Specifically, our goal is to understand the role of ego utility on productivity in competitive settings where participants receive private feedback about their relative standing. 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 seem to care more about their relative, rather then objective level of wealth (Clark and 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), 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 information may impact actions, we employ a setting where

[1] See Prendergast (1999) for a review.

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.

The theoretical model we develop and our experimental results imply 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 workers.

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 and Harkins (1987), White, Kjelgaard, and Harkins (1995)). Thus, individuals are willing to exert more costly effort to avoid falling behind the average, and 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.

This paper contributes to the theoretical and experimental literature on ego utility, intrinsic motivation and peer effects. On the theory side, Benabou and Tirole (2002) focus on the effect of self-esteem on the behavior of people with time inconsistent preferences. 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.

These models, however, do not take into account the fact that in most real life situations people exert externalities on one another. Usually, one's self-esteem is not shaped in isolation but is also influenced by the actions of others. 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. For instance, after receiving bad feedback about relative performance, people with self-image concerns could employ deception strategies as suggested in Benabou and 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 minimal, or may engage in the task again because it is the only way to regain self-confidence (Koszegi (2006)). Extending these prior models, our theoretical framework applies to multi-agent settings and makes clear predictions about which of these effects should be observed in the data.

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)), they value public recognition independently of any monetary consequence and are willing to trade off material gains to obtain it (Huberman, Loch, and Onculer (2004)). The quest for status has labor market implications, for instance regarding wage and promotion schemes, or job search and sorting (Cowen and Glazer (2007)). Using survey and experimental data, Clark, Masclet, and Villeval (2006) find that status measured as one's rank in the income distribution has a more powerful effect on work effort than does the others' average income, suggesting that social comparisons are more ordinal than cardinal.

Peer monitoring has also been proposed as an effective incentive mechanism (Major, Testa, and Bylsma (1991), Kandel and Lazear (1992)). Performing well in front of peers seems to matter even when output does not have an impact on monetary payoffs (Knez and Simester (2001), Falk and Ichino (2006) and Mas and Moretti (2007)). The increase in output observed when people work along peers seems to come mainly from low-output individuals, who work harder in the presence of higher productivity workers (Seta (1982), Bandura (1986), Mas and Moretti (2007))

In contrast to these two 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 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.[2] There is extensive evidence in the literature that external intervention (for example output-based pay or monitoring) crowds out intrinsic motivation and undermines productivity (see Deci, Koestner, and Ryan (1999) or Frey and Jegen (2001) for reviews). For instance, Gneezy and 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 position in the group - that can potentially reinforce intrinsic motivation in ego-driven individuals.

It is possible, though, that in environments where monetary incentives are strong enough to actually motivate people to work hard, they may crowd out the effect of feed-

[2]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.

back that we demonstrate here in a flat-wage environment. 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.

Our results, however, 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, 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 rankings may have on performance is of key importance to the motivational politics of a modern firm.

2 Model

2.1 Setup

In our model two agents, i and j, work on similar tasks. Each individual output is observable and verifiable. The individual output depends on the worker's skill and on the amount of effort he has put into the task. We assume the following production function

$$y_i = a_i + e_i + \widetilde{\varepsilon}_i \tag{1}$$

where a_i represents the agent's innate ability level, e_i is the amount of effort that the agent has put into his task and the term $\widetilde{\varepsilon}_i$ is the realization of an exogenous transitory shock ($\widetilde{\varepsilon}_i \sim N(0, \sigma^2)$) independently and identically distributed across agents. The agent does not know his own ability, nor the ability of his opponent.

We assume that each agent's utility is increasing in his own output, since people enjoy knowing that they are productive (Deci (1975)). Moreover, we assume that the agent's utility depends also on how his output compares to that of the other agent, with utility decreasing in the output of the opponent. This assumption captures the empirical regularity that people enjoy performing better relative to others (e.g., Szymanski and Harkins (1987)). Importantly, agents work for a fixed wage, and do not receive pay linked to performance. The agent is also free to choose how much to care, or pay attention to the feedback about the opponent's output, and therefore about his own relative rank in this task. The intensity with which the agent chooses to care about the other person's output is captured by parameter $s_i \geq 0$. This assumption is similar to that of Benabou and Tirole (2002) that people may ignore information about output in order to preserve their self-esteem.

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. Expression $\ln\left(\frac{k}{k-s_i}\right)$ is increasing in s_i. This means that, all else equal, the higher 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 other agent.

At the end of the working period each agent always knows how much he produced and he may also learn how much the other agent produced. In the beginning of the working period each agent knows the probability (denoted by p for agent i and q for agent j) with which he will get information about the output of his opponent. Working on the task is costly. Agent i, who will receive information about the output of the other agent with probability p, experiences the following disutility (cost of effort) while working:

$$c_i(a_i, e_i) = (\beta - \gamma a_i) \ln (e_i - p s_i) \tag{2}$$

where $\beta > 0$ and $0 < \gamma < 1$ are parameters. For the cost function to be well-defined, we assume that $\beta - \gamma a_i > 0$ and $e_i - p s_i > 1$[3].

Since $\gamma > 0$, we assume that effort is less costly for a more able worker. That is, being better skilled to do a task makes the job more enjoyable, while being less able makes working on the task more frustrating, stressful or disappointing.

Moreover, we assume that effort is less costly if agents set a higher standard for themselves (in other words, being more motivated makes the task less unpleasant), and also, if the probability of learning their rank is higher. This last assumption is technical and it assures that when $p = 0$, the standard set s_i by the agent does not change the cost function, since in that situation the agent will not actually learn their relative

[3] Always holds in equilibrium.

performance (so the agent will not be able to compare himself with his competitor).

Therefore, ex-ante 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:

$$E_i(u_i) = (1-p)E_i(y_i) + p\left(E_i(y_i) - E_i(y_j)\ln\left(\frac{k}{k-s_i}\right)\right) - (\beta - \gamma E_i(a_i))\ln(e_i - ps_i) \tag{3}$$

which is equivalent to

$$E_i(u_i) = E_i(a_i) + e_i - pE_i(y_j)(\ln k - \ln(k-s_i)) - (\beta - \gamma E_i(a_i))\ln(e_i - ps_i) \tag{4}$$

The endogenous reference standard s_i has two effects. As in Falk and Knell (2004), the benefits of getting feedback about relative performance, as well as the cost of producing output decrease with s_i. The latter assumption captures the positive motivational effect of goal setting. The same level of effort appears to be less costly when one works on ambitious and demanding tasks. The former assumption illustrates that the chosen standard s_i can be interpreted as a measure of how much the individual would be hurt by an increase in the output of the other player, or of how frequently he decides to compare himself to the other. The higher s_i is, the more ambitious is the goal set by agent i.

Agent i, therefore, faces the following problem:

$$\max_{e_i,s_i} E_i(a_i) + e_i - pE_i(y_j)(\ln k - \ln(k-s_i)) - (\beta - \gamma E_i(a_i))\ln(e_i - ps_i) \tag{5}$$

which gives the following first order conditions:

$$e_i = \beta - \gamma E_i(a_i) + ps_i \tag{6}$$

$$s_i = \frac{k(\gamma E_i(a_i) - \beta) + e_i E_i(y_j)}{\gamma E_i(a_i) - \beta + pE_i(y_j)} \tag{7}$$

From equations (6) and (7) we get that:

$$e_i^* = \beta - \gamma E_i(a_i) + p\left(k - E_i\left(y_j^*\right)\right) \tag{8}$$

$$s_i^* = k - E_i\left(y_j^*\right) \tag{9}$$

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, $E_i(E_j(p)) = E_j(p) = E_j(E_i(q)) = E_i(q) = \frac{1}{2}$. Given these assumptions we get that agent i expects agent j to produce:

$$\begin{aligned} E_i\left(y_j^*\right) &= E_i\left(a_j + \beta - \gamma E_j(a_j) + q\left(k - E_j a_i - E_j(e_i^*)\right)\right) \\ &= E_i(a_j) + \beta - \gamma E_i(a_j) + \frac{k}{2} - \frac{1}{2}E_i a_i - \frac{1}{2}E_i\left(E_j(e_i^*)\right) \end{aligned} \tag{10}$$

Let $\overline{e}_i^* \equiv E_i(E_j(e_i^*)) = \int_0^1 e_i^*(p)f(p)dp$. Then combining equations (10) and (8) we get that:

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

After rearranging we get:

$$e_i^* = \beta + \frac{pk}{2} - p\beta + \left(\frac{p}{2} - \gamma\right)E_i(a_i) - p(1-\gamma)E_i(a_j) + p\frac{1}{2}\overline{e}_i^* \tag{11}$$

Taking expectations with respect to probability p in equation (11) we obtain:

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

which gives:

$$\overline{e}_i^* = \frac{1}{3}\left(2\beta + k + (1-4\gamma)E_i(a_i) - 2(1-\gamma)E_i(a_j)\right) \tag{12}$$

Combing equations (12) and (11) we obtain the formula for the equilibrium level of effort of agent i who has beliefs $E_i(a_i)$ and $E_i(a_j)$:

$$e_i^* = \frac{2p(1-\gamma) - 3\gamma}{3}E_i(a_i) - \frac{4p(1-\gamma)}{3}E_i(a_j) + \beta - \frac{2\beta p}{3} + \frac{2kp}{3} \tag{13}$$

Using equation (9) and (13) we obtain the equilibrium level of standard:

$$s_i^* = \frac{2(1-\gamma)}{3}E_i(a_i) - \frac{4(1-\gamma)}{3}E_i(a_j) + \frac{2}{3}(k-\beta) \tag{14}$$

In equilibrium agent i produces the following amount of output:

$$y_i^* = a_i + \frac{2p(1-\gamma) - 3\gamma}{3} E_i(a_i) - \frac{4p(1-\gamma)}{3} E_i(a_j) + \beta - \frac{2\beta p}{3} + \frac{2kp}{3} + \widetilde{\varepsilon}_i \tag{15}$$

and ex-ante expects to produce:

$$E_i(y_i^*) = \frac{(2p+3)(1-\gamma)}{3} E_i(a_i) - \frac{4p(1-\gamma)}{3} E_i(a_j) + \beta - \frac{2\beta p}{3} + \frac{2kp}{3} \tag{16}$$

We use equations (13) - (16) to derive the main propositions in the paper.

2.2 Implications

2.2.1 Ex-ante effects

Our model predicts that feedback policy can influence both productivity and beliefs even before any information is revealed to the agents. In particular, agents who expect to receive information about their opponent's output with different likelihoods, other things being equal, will 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 (15) 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 relative performance using the difference in agents' 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:
$\frac{d\left(E_i\left(y_i^*\right) - E_i\left(y_j^*\right)\right)}{dp} = \frac{dE_i\left(y_i^*\right)}{dp} = \frac{2}{3}\left((1-\gamma)(E_i(a_i) - 2E_i(a_j)) + k - \beta\right)$
Since $\gamma < 1$, $\frac{d\left(E_i\left(y_i^*\right) - E_i\left(y_j^*\right)\right)}{dp} > 0 \Leftrightarrow E_i(a_i) > 2E_i(a_j) - \frac{k-\beta}{1-\gamma}$ and $\frac{d\left(E_i\left(y_i^*\right) - E_i\left(y_j^*\right)\right)}{dp} \leq 0 \Leftrightarrow E_i(a_i) \leq 2E_i(a_j) - \frac{k-\beta}{1-\gamma}$. ■

This proposition implies that giving subjects opportunity to compare themselves to others makes the sufficiently confident ones more productive and more optimistic about their relative position in the group, which is highly desirable from the principal's point of view.

2.2.2 Ex-post effects

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.[4] 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 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* $p < \frac{3\gamma}{2(1-\gamma)}$ *(sufficient condition is that* $\gamma > \frac{2}{5}$*)*
- *increase (decrease) if* $p \geq \frac{3\gamma}{2(1-\gamma)}$.

Proof. From equation (15) 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 $\frac{dy_i^*}{dE_i(a_i)} \geq 0 \Leftrightarrow p \geq \frac{3\gamma}{2(1-\gamma)}$
Notice that since $p \leq 1$, if $\gamma > \frac{2}{5}$ then $\frac{dy_i^*}{dE_i(a_i)} < 0$. ■

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 (15) we get $\frac{dy_i^*}{dE_i(a_j)} = -\frac{4p(1-\gamma)}{3} < 0$. ■

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 productivity hierarchy (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.

[4] Note that we do not explicitly model belief updating in this setting, that is, how agents use new information about their own or their competitors' performance to update beliefs about own and others' abilities.

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, $E_i(y_i^*) - E_i(y_j^*)$, and say that agent expects better relative performance when this difference increases.

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

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

$\Rightarrow$

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

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

According to the model feedback also affects agents' motivation. An agent who got good feedback will become more ambitious in the future, in the sense that he will set more demanding goals for himself. He will have to rank better in the future (produce more relatively to his opponent) in order to maintain the same satisfaction level.

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

Proof. Using equation (14) we obtain

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

■

2.3 Total output - some implications

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. The following subsection addresses this question.

2.3.1 Common knowledge of abilities and feedback probability

Recall equation (8)

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

and assume that there is common knowledge of abilities and feedback probability. We then get that in equilibrium

$$e_i^* = \frac{\beta(1-p) + pk(1-q) + p\gamma a_j + pqa_i - pa_j - \gamma a_i}{(1-pq)} \tag{17}$$

$$y_i^* = \frac{\beta(1-p) + pk(1-q) + (1-\gamma)(a_i - pa_j)}{(1-pq)} + \widetilde{\varepsilon}_i \tag{18}$$

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

$$Y^*(i,j) = y_i^* + y_j^* = \frac{\beta(2-p-q) + qk(1-p) + pk(1-q)}{1-pq} + \frac{(1-\gamma)}{(1-pq)}(a_i(1-q) + a_j(1-p)) \tag{19}$$

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 intensity 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. While our experimental analysis precludes us from having common knowledge of abilities and feedback probabilities, we can not test this proposition directly. However, it suggests that feedback can be optimally provided to agents of different types, to maximize effort provision when monetary incentives are weak or non-existent.

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 and therefore we use a controlled experimental setting to test our theory.

In the experiment we ask subjects to solve simple multiplication problems (multiply one-digit numbers by two-digit numbers) during several identically structured rounds. Therefore, participants make real effort choices. 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 come to the lab, and their ability to solve these problems does not improve during the duration of the task. Moreover, 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. 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 data to be meaningful, it is necessary to control for the difficulty level of the multiplication problems. If randomly generated numbers were used to generate multiplication problems, and a participant solved more problems in round two than in round three this could mean two things. Either the person's effort remained the same across the rounds but the problems in period two were easier, or he 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)[5] in order to avoid this possible confound.

Problems were presented to each subject on a computer screen. Each time the subject solved the multiplication correctly one point was added to his score and the next problem was presented. If the subject provided a wrong answer, the score remained unchanged and he was 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 was solved, we avoid a situation where participants may strategically skip difficult problems looking for

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

easy ones.

The experiment consisted of 18 rounds. Each of them had the same structure and three feedback conditions were possible. The conditions differed with respect to the probability with which the subject received feedback about his relative rank at the end of the round. This probability was either 0, 0.5 or 1. We refer to these as the "No", "Maybe" or "Sure" treatments, respectively. The feedback condition was determined randomly and independently for every subject at the beginning of each round.[6] Therefore, in the same round different subjects faced different feedback conditions.

The sequence of events in each round is shown in Figure 1. First, 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.

Afterwards, subjects were asked to report their expected rank in that round.[7] Following that, 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). After the 90 seconds passed, subjects were asked to asses 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".

In the final stage of each round, that lasted for fifteen seconds, each subject either saw the performance ranking 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

[6]Using alternating messages within one session allows us to control for session effects.

[7]We 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 that others) changes behavior we are confined to a flat-wage environment.

that nobody else could associate his identity to his actual rank and score.

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. Importantly, 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 rank, as well as on their actual output, measured as the number of multiplication problems solved correctly. These effects are illustrated in Fig. 2.

Output is 7.28% higher (11.35 vs. 10.58 solved problems per round, $p < 0.07$ 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. 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 12.91 for men, and 8.69 for women ($p < 0.001$ in a one-sided mean comparison test), 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). This is consistent with prior experimental findings. For instance, Huberman, Loch, and Onculer (2004) observe that males seek status more than women, and Falk and Knell (2004) find that women have significantly lower aspiration levels than men regarding college education accomplishments.

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). 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 Tables 1 and 2 show 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 in Table 1, and relative to not getting any feedback at all in Table 2.

As predicted by Propositions 2, 3 and 4, 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. As above, these effects are measured relative to receiving neutral feedback in Table 1, and relative to not getting any feedback at all in Table 2. We 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 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** dispersion 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 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. We base this argument on the results in 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. Hence, we proxy heterogeneity 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 test more formally whether stable hierarchies do get formed, and if so, how soon it happens. Fig. 4 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. We revisit these two effects at the end of this section.

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 theses 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 5 and 6 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 4. 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 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 productivity that is shown in Figure 4 comes mainly from high productivity subjects who fight to maintain or improve their rank. A recent quote[8] 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."

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

[8]http://www.brainyquote.com/quotes/quotes/v/vijaysingh183223.html

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).[9] 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. 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), produce higher output (14.90 versus 6.65 multiplication problems per round) and have a lower cost of effort (0.30 versus 0.61). All of these differences are statistically significant ($p < 0.01$).

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 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 reputation,

[9]The average rate 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.

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 as much.

Therefore, our results suggest that in competitive settings productivity and beliefs are influenced by privately observed information about relative position in the group hierarchy. Importantly, the effects of private rank feedback on output are comparable to those of peer monitoring mechanisms documented in prior work. Mas and Moretti (2007) 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 and 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 compare themselves to others (by increasing the probability of feedback from 0 to 0.5) raises individual output on average by 7.28%, a sizeable effect compared to that of peer monitoring.

In light of our findings, it is natural to ask whether an optimal feedback policy exists. In other words, we would like 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 directly compare such complex feedback policies, our results have several implications for improving productivity. Those implications should be taken with caution, as their external validity remains to be examined in future work.

For instance, the principal could take advantage of the ex-ante effect of feedback likelihood on effort provision. Our model suggests that an organization could produce more if it used different feedback likelihood policies for agents of different skill. Feedback should be given more frequently to agents who either believe that they have, or actually posses, relatively high ability. To prolong the effectiveness of relative rank information, the principal could either provide noisy feedback to slow down the learning of one's rank in the hierarchy, or reshuffle work groups once the hierarchy is established and known.

Since in more homogeneous groups incentives are preserved for all members while in heterogeneous groups members split into top performers, who keep fighting for high ranks, and bottom performers, who compete much less, reshuffling may allow low-rank workers to climb the hierarchy in another group, and as a result, to generate more output. Finally, the principal could manipulate the beliefs of the agents. Both the model and the data suggest that if competitors appear to be too tough, an agent's performance deteriorates. Therefore, improving workers' beliefs about their relative ability may have a positive impact on productivity.

References

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

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

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*.

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 Economcs 10*(2), 171–178.

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.

Frey, B. and R. Jegen (2001). Motivation crowding theory: A survey of empirical evidence. *Journal of Economic Surveys*.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: *g*ender differences. *Quarterly Journal of Economics 118*(3), 1049–1074.

Gneezy, U. and A. Rustichini (2000, August). Pay enough or don't pay at all. *The Quarterly Journal of Economics 115*(3), 791–810.

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 19*, 743–772.

Koszegi, B. (2006). Ego utility, overconfidence and task choice. *Journal of the European Economic Association*.

Luttmer, E. (2005). Neighbors as negatives: *r*elative 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.

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: Sequence of events in a round.

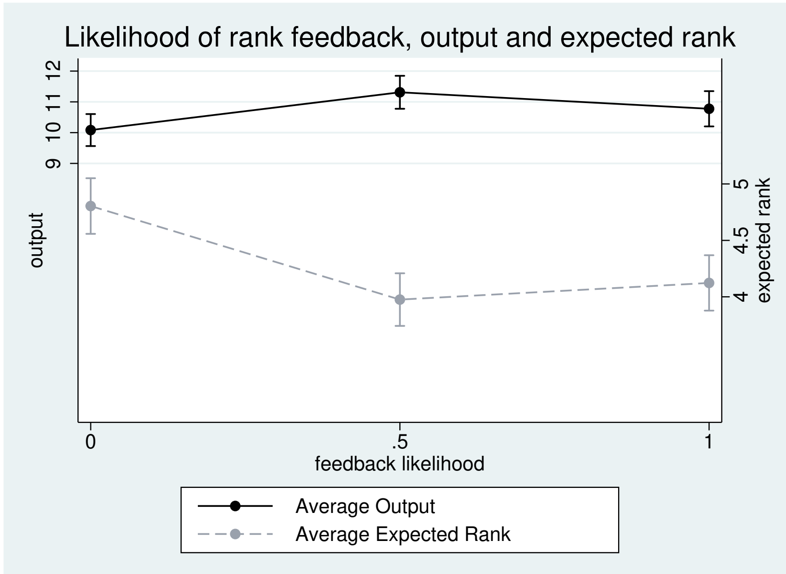

Figure 2: Feedback likelihood, output and expected rank

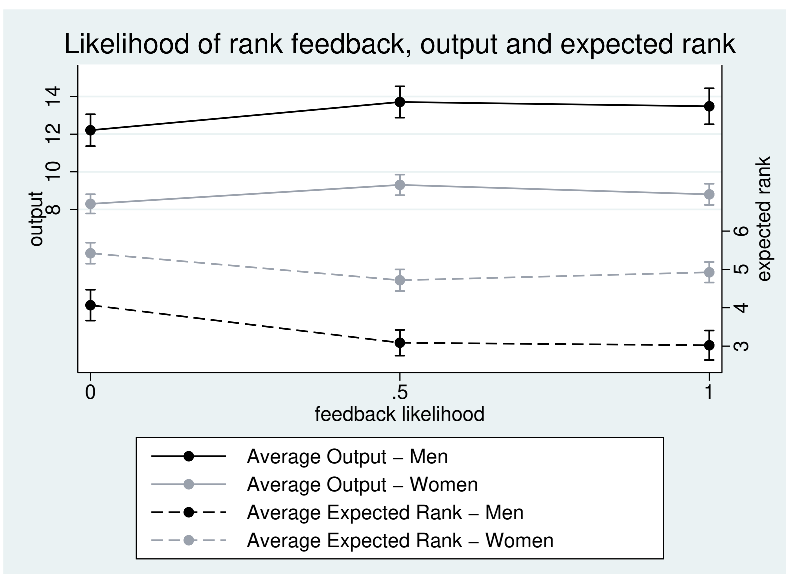

Figure 3: Feedback likelihood, output and expected rank, by gender

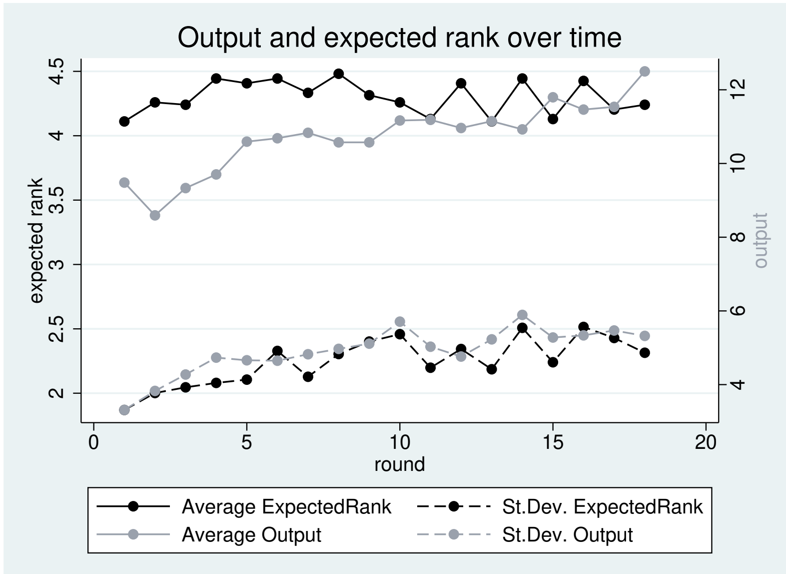

Figure 4: Output and expected rank, by round

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

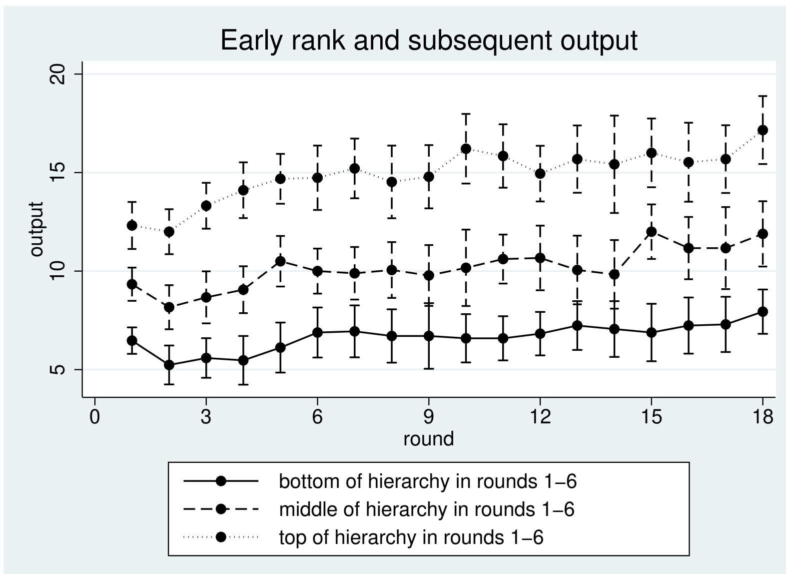

Figure 7: The average output produced each round by subjects who were at the top, in the middle or at the bottom of the rank hierarchy during the first six rounds.

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. 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.

	$Output_t$	$ExpectedRank_t$	$Rank_t$
	Coef./t	Coef./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 included

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

Table 2: Ex-post impact of feedback on estimated rank, actual rank and effort
Alternative specification for models in Table 1. The reference feedback category is given by observations where subjects did not receive relative rank information at the end of the prior round ($ExPostFeedback_{t-1} = 0$). T-statistics are in parentheses.

	$Output_t$ Coef./t	$ExpectedRank_t$ Coef./t	$Rank_t$ Coef./t
$GoodFeedback_{t-1}$	–0.88	–0.27	0.56
	(–3.08)***	(–1.54)	(3.50)***
$BadFeedback_{t-1}$	0.62	0.77	–0.45
	(2.26)**	(5.17)***	(–2.66)**
$NeutralFeedback_{t-1}$	–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 included
$^{*}p < .10, ^{**}p < .05, ^{***}p < .01$

Table 3: 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 $Rank = 1$). 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 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.

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