

# Rankings and Social Tournaments: Evidence from a Field Experiment<sup>1</sup>

Iwan Barankay<sup>2</sup>

*University of Pennsylvania, CEPR and IZA*

July, 2010

*Abstract:*

People often compare their performance to that of others. The question is whether these comparisons affect effort provision so that they can potentially be exploited to substitute for monetary incentives. In this paper we present evidence from a field experiment where employees were given feedback about how they rank in terms of performance compared to others doing the same task. The context is such that rank had no implication for current or future compensation. Compared to a control group with no rank feedback, those employees who received feedback about their rank reduced effort along both the intensive and the extensive margin: they were significantly less likely to return to work and when they did they were less productive when getting feedback about their rank. We find no heterogeneous effects by gender or by whether employees had a stated or revealed preference for receiving feedback about their rank. We find some evidence that giving people unexpectedly positive feedback did not significantly reduce performance.

Keywords: Rank Incentives, Social Tournaments, Feedback, Field Experiments.  
JEL Codes: L2, M5

---

<sup>1</sup> This paper benefited from financial support by the Economic and Social Research Council, U.K., and the Mack Center at the Wharton School. I greatly benefited from comments by Peter Cappelli, seminar participants at the Applied Economics Workshop at the Wharton School, the Behavioral Economics Conference in Jerusalem 2010, and the CEPR Public Policy Symposium 2010 in London. The usual disclaimer applies. Melvin Kanasseril provided outstanding programming support.

<sup>2</sup> The Wharton School, University of Pennsylvania, CEPR and IZA. Email: [barankay@wharton.upenn.edu](mailto:barankay@wharton.upenn.edu).

Even though monetary incentives can be very effective under certain conditions (Green and Stokey, 1983, Lazear and Rosen, 1981), it is well known that they exhibit limitations for instance under moral hazard or contractual incompleteness. Moreover people may not be driven by financial incentives alone but may also have social concerns of some kind. This paper focuses on one important type of social concern that receives vivid attention recently, which is a preference people hold over their rank order position in terms of performance compared to that of others. Rankings are a very common and prominent feature of workplace environments. Cardealers rankings are displayed on Walls of Honor in back-offices; Economists receive an email once a month by Repec inviting them to see a whole list of ranking statistics related to their publications and working papers; students are always keen to know how they are placed within their cohort; and there is a general curiosity in rankings as evidenced by a tremendous interest in sports rankings. More importantly perhaps, rankings are used more and more to determine the allocation of funding – both public but also private in the form of donations – to schools, hospitals, or universities but also by the vivid discussions in companies about whether to display rankings in the workplace and who gets to see them.

Despite the clear evidence for an interest and an opinion about the use of such rankings, we still know little about how providing information about rankings affects the behavior of individuals in organizations. To study such rank preferences is therefore important both theoretically and empirically. As is made clear below, how people respond to receiving updates about their rank is theoretically ambiguous. Even though it can encourage some to catch up or to excel even more,

others may be de-motivated or become complacent. Which behavior prevails is thus pre-dominantly an empirical question. Empirically, however, the challenge is to disentangle the effect of feedback about rankings from purely pecuniary interests: do people really care about their rank or do they simply care about the monetary benefits that come with it?

In light of these challenges, a field experiment has been designed to see how people adjust their effort, both along the intensive and the extensive margin, when they are given feedback about their place in rankings along dimensions that does explicitly not affect their current or future compensation. For the purpose of this experiment, workers are recruited via Amazon's crowd-sourcing website Mechanical Turk. Crowd-sourcing is an important employment phenomenon whereby employees log on to a website to select work offered to them online.<sup>3</sup>

For the experiment, job advertisements are placed online on the job listings of Amazon's Mechanical Turk and workers were recruited to analyze images. Workers were paid a piece-rate for the quantity of their work irrespectively of their work's quality. The key treatment was to randomize employees into two groups. Workers in the first group received feedback about how they rank in terms of accuracy of their work – note that accuracy had no bearing on their compensation – whilst workers in the other group received no feedback about their rank.

---

<sup>3</sup> The tasks and the work are also conducted online and submitted to the employer who pays for their output. Amazon's Mechanical Turk is a platform that allows the posting of such online work and facilitates the payment via their servers. Many jobs on this platform revolve around marketing and advertising campaigns where a human input is required to analyze and categorize images, to write and edit text, or to enter data thus capturing a large proportion of routine office work.

Comparing the behavior between the control and the treatment group, those who receive feedback are 30% less likely to return to work and when they do come back they are 22% less productive.

The theoretical framework highlights that there can be important heterogeneities in the effort response to feedback about rank depending on workers' prior belief about their rank, their preference over the difference between their actual rank, and their personal reference points with respect to which they compare themselves to. In light of such a rich set of underlying parameters the experiment attempted to make the first steps to address aspects of these heterogeneities. We explored heterogeneous treatment effects along a number of observable characteristics such as gender, rank preference and whether they received unexpectedly high rank feedback. We approximated the preference over rank by gender, the revealed, and the stated preference for rank feedback. We find no heterogeneous treatment effects using these proxies. Women responded no different than men to information about their rank both on the intensive and the extensive margin. Similarly, those with stated or with revealed preference for rank feedback also did not show a differential treatment effect. We did however find some evidence that telling people that they were in the Top 10 did not significantly reduce their performance.

There is now an active literature on status and rankings. This goes back to a rich literature in sociology (Durkheim, 1951) but also is in many ways akin to conspicuous consumption in economics (Veblen, 1949). In economics, status and rank received more and more attention on a conceptual level (Frank, 1985), theoretically (Auriol and Renault, 2008, Besley and Ghatak, 2008, Moldovanu, et al.,

2007) and also with empirical work in the laboratory (Charness and Kuhn, 2007, Eriksson et al. 2008), event studies and natural experiments (Azmat and Iriberry, 2010, Blanes-i-Vidal and Nossol, 2009, Ehrenberg and Bognanno, 1990) and field experiments on feedback in a team context (Bandiera et al, 2010b and Dur et al, 2010). Finally, recent work stressed the heterogeneity in competitive behavior by observable characteristics especially by gender (Dohmen and Falk, 2010, Gneezy et al. 2003, Gneezy et al 2009, Niederle and Vesterlund, 2007) but also by relative ability within social networks (Bandiera et al. 2010a).

This paper is the first field experiment that explores rank preferences as a social tournament and its effect both on the intensive and the extensive margin. Furthermore, it is also among the first papers to explore in detail the heterogeneous treatment effects associated with rank information.

The next section discusses the theoretical framework, followed by a description of the empirical context. Then we present the experimental design with the results before the paper closes with conclusions.

# Theoretical Framework

To clarify how employees respond to updates about their rank, it is important to show that the behavioral response is very heterogeneous with respect to treatment. In what follows, we set up a very simple framework whose sole purpose it to clarify some underlying causes of the heterogeneity, which informs the experimental design and the empirical analysis.

We start off with a modified version of the model proposed by Fehr and Schmidt (1999) to study rank preferences more generally. Assume that a worker  $i$  in addition to preference for piece-rate wages  $w_i(e) = e_i$  also has a preference over her rank order position  $r_i$ , with respect to a reference or focal point  $\bar{r}_i$ . The idea here is that people have a rank, for instance they are ranked 27<sup>th</sup>, but what they compare themselves to is how they are placed compared to a reference point, for instance the employee in 10<sup>th</sup> place. This reference point is specific to each person and it captures that people have different reference groups or focal points. Both wages and rank increases in effort  $e$  at cost  $c(e_i; \theta_i) = \frac{1}{2} \theta_i e_i^2$  where  $\theta_i > 0$  captures worker heterogeneity of the cost of effort. Finally assume that rank order position  $r_i$  increases in effort but at a decreasing rate.<sup>4</sup> Employees then have the following utility function

$$U(e; \alpha_i, \beta_i, \theta_i) = e_i - \frac{\theta_i e_i^2}{2} - \alpha_i \max\{\bar{r}_i - r_i, 0\} - \beta_i \max\{r_i - \bar{r}_i\}.$$

---

<sup>4</sup> Formally  $\frac{\partial r_i}{\partial e_i} > 0$  and  $\frac{\partial^2 r_i}{\partial e_i^2} \leq 0$ . These and other functional form assumptions are made for expositional purposes and could be relaxed or generalized.

The first element,  $e_i$ , is the piece-rate wage, the second is the cost of effort:  $\alpha_i$  then measures the preference over having a lower rank  $r_i$  than the reference point  $\bar{r}_i$ , whereas in the last element  $\beta_i$  captures the preference for having a higher rank than the reference point. This is a very flexible utility function that represents a broad combination of preferences over rank. First, for  $\alpha_i, \beta_i = 0$  workers don't care about rank; second, those with  $\alpha_i, \beta_i < 0$  are *inequity loving* as they enjoy both being better but also being worse than the reference point; third, *status seeking* workers have  $\alpha_i > 0$  but  $\beta_i < 0$  as they like being better but dislike being worse than their reference point; fourth, at least theoretically there can be *status avoiding* employees who have  $\alpha_i < 0$  but  $\beta_i > 0$  as they like being worse but dislike being better than their reference point.

The first order condition  $\frac{\partial u_i}{\partial e_i} = 0$  for people who are better ranked than their reference point,  $r_i - \bar{r}_i > 0$ , yields  $1 = \theta_i e_i + \alpha_i \frac{\partial r_i}{\partial e_i}$  and for those with a rank lower than their reference point  $1 = \theta_i e_i + \beta_i \frac{\partial r_i}{\partial e_i}$ .

In the field experiment below we will be comparing workers who receive information about their rank to those who don't. In the control group those who don't receive feedback nevertheless form *beliefs* about their rank  $\tilde{r}_i$ , which may be

higher or lower than their true rank. We take the existence of this prior belief as given.<sup>5</sup>

In the treatment group, workers receive feedback about their rank so that they update their beliefs about their rank. Denote  $\tilde{r}_i < \hat{r}_i$  an update where the true rank is below the initial belief and by  $\tilde{r}_i > \hat{r}_i$  an upward revision as the true rank is higher than the initial belief.

Exploring the first order condition above reveals how a worker adjusts effort in light of an update about the rank.

First, it is worthwhile to point out two reasons why workers do not change their effort. Either they don't care about their rank,  $\alpha_i, \beta_i = 0$ , or they do but the update is in line with their beliefs, i.e.  $\tilde{r}_i = \hat{r}_i$ . Empirically this implies that not seeing a change in behavior after feedback about rank does not necessarily rule out rank preferences. For instance, employees with a lot of experience or very good access to information about others' performance could be less responsive to feedback as they already know their rank order position.

Second, it is often argued that providing feedback generates a competitive environment that entices people to work harder as they are status seeking. This assumption is not necessarily the case. To see this, take the situation of a status seeking employee who loves being ahead but hates being worse than the reference

---

<sup>5</sup> We take them as given for ease of exposition. These beliefs could be modeled as the outcome of prior experience or experimentation. Often, people state that they have no prior over their rank. This captured three ways theoretically: First that they have no preference over rank, i.e.  $\alpha = \beta = 0$ , second that they do but they only realize it once they learn their rank, and finally that the preferences change with the environment that that they adjust their preferences in light of feedback about their rank.



point,  $\alpha_i > 0$  but  $\beta_i < 0$ . Whether the worker will work harder then depends on the worker's ability parameter  $\theta_i$  and whether the update is better or worse than the prior belief. For instance, a person may become complacent and reduce effort in light of an update, when (i) the person is of high ability (small  $\theta_i$ ), when the belief is higher than the reference point and the update is higher than the belief so that  $\bar{r}_i < \tilde{r}_i < \hat{r}_i$ . So it is not the feedback per se but how it affects the utility

Indeed, focusing more on high ability workers, we can characterize four cases in which people will reduce effort in light of feedback about the rank. First, an overly *optimistic person* who thinks and enjoys being well ahead of the reference point but learns that she is not as well placed as she thinks ( $\bar{r}_i < \check{r}_i < \tilde{r}_i$  and  $\beta_i > 0$ ) will reduce effort as she learns that the return to status from effort  $\partial r_i / \partial e_i$  is lower than she thought it would be; Second the *complacent worker* mentioned above ( $\bar{r}_i < \tilde{r}_i < \hat{r}_i$  and  $\beta_i < 0$ ); third there is the *slacker* who believes that he is lower ranked than the reference point but actually learns that the true rank is even lower than that yet dislikes being worse than the reference point ( $\check{r}_i < \tilde{r}_i < \bar{r}_i$  and  $\alpha_i > 0$ ) as the update reveals that the return to status from effort  $r_e$  is less than was the case under the belief; finally the *fatalist* who enjoys being below the reference point and withholds effort to instill a utility flow from low rank but then learns that she is doing better than she thought.

In general, the key insight is that rank generates utility which after the update is higher or lower than was the case under the prior belief meaning that the return to status from effort was higher or lower than under the prior. In response to that update, workers have to

trade-off the cost of effort, the effect of changing effort on their piece-rate wage, and the effect on rank of adjusting behavior. Whether the effort will increase or decrease depends on whether ability is high or low, the true rank is higher or lower than the belief, whether people are above or below their reference point, and finally whether they like or dislike being away from their reference point. In summary, the response to an update is theoretically complex. Empirically it is unlikely for a policy maker or someone in charge for designing workplace incentives to know all the aspects of a workers beliefs, and preferences over rank to correctly anticipate their response. What this theoretical framework tells is it to seek three types of results in the experimental framework below. First, to establish an empirical regularity about how, on average, employees respond to feedback about their rank order position. Second, whether observable characteristics suggested by the prior literature, like for instance gender or experience, can help to uncover the heterogeneities revealed in this theoretical section. Third, we use the model more directly to categorize instances when people might respond positively or negatively to feedback about rank. Ideally, we would like to know or back out from the data the preferences over rank, the  $\alpha_i$ 's and  $\beta_i$ 's, and the people's prior beliefs about their rank, the  $\bar{r}_i$ 's. This is of course very challenging and below, we will document the attempts made into that direction and suggestions for future empirical work.

## Context

To shed light on the effect of feedback about the rank order position we implemented a sequence of experimental treatments on Amazon's Mechanical Turk crowd-sourcing webpage (<http://www.mturk.com>). As this is a fairly new environment for field experiments it worthwhile to give some further details about it. The webpage is a platform for work conducted and submitted online. Employers, or *requesters* in the terminology of Mechanical Turk, post jobs on the website's job listings section. Each posting contains a title, a short description of the job, keywords, and the payment structure of the job. Typically jobs offer piece-rates, but bonuses conditional on a broad range of measures (e.g. accuracy, speed or completeness) are very common, which implies that virtually any monetary incentive scheme can be implemented. In the experiment below, however, we will only use piece-rates based on the quantity of work. Workers, also called *turkers*, peruse the list and pick a job on offer. The number of jobs to choose from varies by day and time and currently ranges between 100,000 and 250,000. When they select a job, they are led to a new webpage that presents the work to them. Each of those web-pages are so-called *HITs*, short for human intelligence task. The computer literacy required is really minimal and it is as easy as setting up an account on Amazon itself.

Most jobs on Mechanical Turk revolve around tasks that are amenable to be split into small increments, that can be digitized, and that require a human input. This may sound more abstract than it is. A brief investigation of the jobs posted by the

principal employers on Mechanical Turk reveals that many tasks are those that would otherwise be done by in house office personnel such as the analysis or documentation of the content of marketing material such as images of merchandise, but also the composition or editing of texts for catalogues or corporate blogs, the entry of data, and other related task. <sup>6</sup> To be able to earn money, workers currently need to have a bank account either in the US or in India.<sup>7</sup>

A worker can click on a job that leads them to a website, the *HIT*, where, after accepting the job, the worker can do and submit the work. After, a worker can abandon the job or can choose to be lead to the next webpage in the sequence. For instance, an employer may ask workers to answer questions relating to a number of images. There is but one image per webpage, so workers must proceed to another page to do more work.

The advantages using Mechanical Turk for field experimental work are manifold. First, a central appeal of field experiments is that subjects don't initially know that they are part of an experiment<sup>8</sup> and that experiments are conducted in a natural

---

<sup>6</sup> Amazon's business model apparently is to offer the software platform for the posting and hosting of the jobs and to facilitate the financial transaction between employers and employees. Their legal position is that it is the requester that enters the employer-employee relationship and not them. This has tax implications that other researchers need to be aware of. Enrolling people into experiments on Amazon's Mechanical Turk means that one is hiring them. If one is using University funds to pay workers, then the treatment of employees is determined by and varies according to University human resource departments. Typically there is a yearly income threshold below which a full set of paperwork, e.g. W6 and their social security number, is not required.

<sup>7</sup> This is due to the Patriot Act's money laundering provisions, the requirements of which Amazon is currently able to satisfy in the US and in India.

<sup>8</sup> Institutional review board requirements vary across universities in order to obtain a waiver of written consent by human research subjects that would inform them that they participate in an experiment. Usually these requirements are that the purpose of the job needs to be stated (e.g. University research), a contact person needs to be available, that the task poses no more than minimal risk, and that they will be debriefed at the end. All these conditions can be met on Mechanical Turk.

environment such as a workplace. Both these criteria are met. Furthermore, in contrast to field experiments conducted in collaboration with firms (Bandiera et al. 2010b), the experimenter can pose as the firm giving him substantial control about the protocol and thereby eliminating many project risks relating to field experiments. Also, the timeframe is much shorter to conduct experiments on Mechanical Turk: The whole experiment can be completed in a couple of hours, yet one retains the flexibility to implement long-term experiments as well.<sup>9</sup> Finally, the costs of running experiments can be low but one has to factor in the programming costs which depending on the experimental set-up can be substantial.

There are, however, two very appealing aspects about Mechanical Turk. First, as will be shown in the next section, it allows a very neat study of the extensive margin. Second, the characteristics of workers are much broader than can be found in most companies or laboratories. Table 1 displays data from a survey conducted among the workers enrolled into this paper's jobs after the completion of experiments. Roughly over half of the workers are female, and the age range is younger compared to the overall population, which explains also the somewhat lower educational attainment. The self-reported income range is lower than in the overall population but is also very broad. The intensity with which these people work on Mechanical Turk reflects that some workers do this alongside their main job, but some spend a good part of the week on Mechanical Turk: It takes workers between 30-60 seconds

---

<sup>9</sup> This contrasts very favorably with field experiments conducted in collaboration with firms where it takes at least one year but typically closer to two years from initial meetings to the completion of the experiments. There is no hard data on completion rates but my perception is that only a fifth of field experiments that actually get under way also reach the final experimental stage.

per HIT so that the 1000 HITs take them 8 – 17 hours per week. The risk attitudes and patience are not far from those found in larger samples. The key point is that while the people who work on Mechanical Turk are by no means representative of the overall population, they are not particularly remarkable in terms of characteristics.<sup>10</sup>

Any experiment or indeed any empirical study faces the challenge of external validity, and this setting is no exception. To have a meaningful discussion on external validity, one first has to define the population of interest and then, if the population of interest is different from that in the sample, how that may bias the results. Clearly people who work on Mechanical Turk are special by the simple fact that they choose work there. But based on the survey results, they don't strike us as unusual in other aspects.

---

<sup>10</sup> Moreover, it is noteworthy, although not done in this paper, that given this broad range of characteristics, one could aim to implement experiments with subjects that are otherwise hard to recruit in laboratories or typically underrepresented in companies, e.g. single mothers, or people over 65. This is important for replication studies or for studies requiring such detailed subject pools.

**Table 1: Characteristics of workers participating in the experiments (n = 227)**

Gender: 54.18% female

Age:

16-20	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65	65+
14.1%	16.7%	19.4%	12.8%	6.6%	8.4%	9.3%	6.2%	4%	2%	0.4%

Education:

Some High School	High School	Some College	Associates degree	Bachelor's Degree	Master's Degree	Doctorate
2%	12%	29%	15%	28%	12%	1%

Income in 2009

<\$10K	\$10K-15K	\$15K-25K	\$25K-40K	\$40k-60K	\$60K-75K	\$75K-100K	\$100K-150K	\$150K-200K	\$200k-250K
26%	11%	11%	22%	16%	4%	5%	4%	1%	1%

Experience

"Over the last month how many HITs did you submit on average per week?"

1-4	5-9	10-19	20-49	50-99	100-249	250-499	500-1000	1000-4999
9%	7%	10%	15%	15%	19%	9%	10%	6%

"For how long have you been active on Mechanical Turk?"

0-3 months	3-6 months	6-12 months	1-2 years	2years+
73%	9%	9%	6%	4%

Risk<sup>11</sup> - Zero means: "not prepared to take risks" and 10 means: "fully prepared to take risks."

0	1	2	3	4	5	6	7	8	9	10
2%	1%	4%	9%	11%	20%	15%	18%	13%	4%	5%

Patience<sup>12</sup>

Very impatient	Somewhat patient	Neither	Somewhat patient	Very patient
2%	25%	10%	42%	22%

<sup>11</sup> The wording of the question was "How do you see yourself? Are you generally a person who is fully prepared to take risks, or do you avoid taking risks? (Please choose a number from 0 to 10 where 0 means: "not prepared to take risks" and the value 10 means: "fully prepared to take risks.")"

<sup>12</sup> The wording of the question was "How do you see yourself? Are you generally a person who is patient or impatient?"

## Experimental Design

For the purpose of these experiments we posted jobs on Mechanical Turk that are very similar to other jobs encountered on Mechanical Turk in terms of difficulty, pay,<sup>13</sup> prior experience needed, and time. We also wanted to make sure that some elements of the tasks require real human effort.

We posted jobs in which people had to answer four questions relating to a picture. Figure 1 gives an example of a picture. The pictures are of walls and pavements on and around the campus of the University of Pennsylvania. The four questions are, (i) whether the picture is in focus, (ii) whether there are any clear cracks or lines in the picture, (iii) what the total sum of the length of lines in % of the width of the picture is, and (iv) to assess the overall quality of the picture. The precise wording of the questions is given in the Appendix. We offered a \$0.05 piece-rate when they answered the four questions relating to each picture. It is important to note that payment was independent of quality, which we communicated to workers at various stages in the experiment. In addition we paid them between the experimental stages so that they knew that pay only depended on the quantity worked and not the quality. When designing this job it was of key importance to keep its appearance in line with what these workers were used to. Even though this job may seem

---

<sup>13</sup> As experiments on Mechanical Turk become more common the payment practices become a point of contention. Browsing the Mechanical Turk webpage one encounters many jobs with very low pay. No comprehensive data exist on how popular those jobs are. It is true that one could do jobs very cheaply but that would limit the external validity of the results. An aim in this study was to offer pay where it was under plausible conditions a worker could earn the minimum wage.



abstract, uninteresting or even of questionable use,<sup>14</sup> it is very comparable to other jobs on Mechanical Turk and perhaps even of some office work.

The experiment proceeded in several stages.

*Stage 1: Eliciting revealed and stated preferences for feedback about rank*

In the first stage of the experiment we wanted to elicit preferences for feedback about the rank order position. This was important for three reasons. First, it is often stated that people want to know how they rank. This has been the principal stimulus to the status and the rank literature. So it is useful to establish if this is empirically the case. Second, we wanted to know if giving feedback about rank affected workers differently depending on whether they had a stated or revealed preference for rank feedback. Third, we will use the revealed preference for rank as indicative of the fact the workers care about their rank, i.e. in terms of the model that their  $\alpha_i$ 's and  $\beta_i$ 's are different from zero.<sup>15</sup>

We posted two jobs simultaneously on Mechanical Turk. These two jobs were entirely identical except only for the title of the job. The first job had the title "Answer four questions relating to a picture. No typing required." The second job had the title "Answer four questions relating to a picture and get feedback on your rank in terms of accuracy. No typing required."

---

<sup>14</sup> In the survey to workers we asked them why they chose to work for us. 24% said the pay was good, 43% said that the work looked interesting, 9% stated the work looked fun, and 23% chose that the work looked easy.

<sup>15</sup> It has also been attempted to ask the workers directly what their reference point is and whether and in what way they cared about being better or worse than their reference point. Validating pre-testing those questions revealed a number of problems of framing and misinterpretation of the question. We therefore chose a more reliable yet limited approach of asking them whether they like to get feedback about their rank order position.

### *The effect of offering feedback on job selection*

Faced with that choice, the surprising result is that the first job *without* the feedback was substantially more popular. During the time we posted the jobs,<sup>16</sup> the no feedback job attracted 254 workers compared to 76 workers who chose the job with feedback. So when workers are offered a choice between two identical jobs but one with rank feedback and the other without, we cannot confirm the prior that there is a strong demand for rank feedback. It is important to note for what follows that at this stage of the experiment none of the workers received feedback – who receives feedback will actually be randomized in the next step.

As noted above, we administered a survey to workers *after* all experimental treatments. In other studies it has been shown how simply asking people about their preference has strong predictive power for their revealed preferences via actions. As there was little guidance in the literature as to whether it is necessary to get data on revealed preference for rank or whether simply asking about their preferences is enough, we wanted to get data on both to see what the correlation between stated and revealed preferences is. In the worker, survey we thus asked several questions about rank feedback. When asked how much they agree or disagree with the statement “[y]ou want to receive feedback about your relative performance, i.e. how your performance ranked compared to others”, 44% answered strongly agree, 31% agree somewhat, 17% neither agree or disagree, 6% disagree somewhat, and 2% selected strongly disagree. So there is already strong

---

<sup>16</sup> The jobs were posted twice for 24 hours each on different days to deal with potential day of week effects.

discrepancy between stated and revealed preferences. Whereas a 74% majority stated they want rank feedback, only a 23% minority actually chose the job where feedback was offered.

Next we can investigate the correlation between stated and revealed preferences. The correlation between the survey response of 227 workers<sup>17</sup> and which job the worker chose is only 0.0308.

Clearly, the stated and revealed preferences are orthogonal to each other. One could say that this low correlation may be specific to the job at hand or the wording of the survey question. Yet it would be hard to imagine that noise and framing alone could bias a true relationship by that much.<sup>18</sup> One lesson here is that data on revealed preference adds information over and above stated preference for feedback about rank. In the appendix we explore with logistic regressions which workers chooses the feedback over the no feedback job and how that selection affected subsequent performance.

In summary, the job without feedback was much more popular among the workers. We fail to predict, based on observable characteristics, which type of worker chooses the feedback rather than the no rank feedback job. Finally there is no discernable selection effect across workers that lead to a difference in the productivity of workers. One final note is that at the end of this stage we paid workers for any and all the work they did irrespective of the accuracy of the work.

---

<sup>17</sup> We checked whether the survey respondents were a selected sample but, based on observable we find no evidence for a bias. Details are available upon request from the author.

<sup>18</sup> This result however is perhaps in line with a casual observation among students for: they have a keen interest to get feedback about how their performance ranks compared to others but have a rather strong emotional response when they are given feedback about their rank.

They knew, therefore, that compensation did not depend on the quality – the accuracy – of their work.

*Stage 2: Giving workers feedback about their rank in terms of accuracy*

In the next stage of the experiment we block-randomized workers into two groups to determine who was given feedback about their rank order position. More specifically we took half the workers who chose the no feedback job in the first stage and half the workers who chose the rank feedback job and allocated them to a control group leaving the other half in the treatment group.

We then sent an email to all those workers via the Mechanical Turk messaging system. The 142 workers in the control group (*without rank feedback in email*) received an email inviting them to come back again to do more work for us. In the email we informed them that all workers were sent this email, as we wanted to make sure they don't feel selected in any way beyond the fact that they worked for us before.<sup>19</sup> The appendix gives the exact wording of the emails. The 142 workers in the treatment group (*with rank feedback in email*) received the same invitation, but in that email we informed them about their rank in terms of accuracy of their answers compared to others. No other element was different across the two groups, e.g. the pay and any aspect of the work they were asked to do was identical.

---

<sup>19</sup> Mechanical Turk allows employers to give out qualifications to individual workers. In this stage of the experiments we qualified the workers to do more work for us and only these qualified workers could work in this stage of the experiment. Just to be clear all the workers who participated in the first stage of the experiment were sent emails and invited to participate in the second stage.

*The average treatment effect of rank feedback on the intensive and extensive margin*

We can now study whether giving people feedback about rank affects on average their extensive margin – in our context whether people come back to do more work when invited to do so – and the intensive margin, the quantity of their work conditional on coming back to work.

Of the 142 workers in the control group (*without rank feedback in email*), 88 (62%) returned to do more work. In the treatment group (*with rank feedback in email*) only 59 returned to work (42%). Thus providing feedback substantially reduces the extensive margin. This is also confirmed in logistic regressions in column 1 of Table 2.

We can also study the intensive margin here by measuring the number of pictures done by those people. This is interesting for a specific reason: it may be the case that sending people feedback about their rank might dissuade some workers to come back to work but might help to entice the high performing workers to return. This is, however, not that case. On average, those who received feedback about their rank order position were 22% less productive conditional on returning to work. Those who received no information about their rank in the invitation email completed on average 19.6 pictures compared to 15.3 pictures by those who did receive feedback about their rank in the invitation email. Column 1 of Table 3

confirms this result in a regression model where the coefficient on the treatment dummy is significant at the 1% level.<sup>20</sup>

*Heterogeneous treatment effects of feedback on the intensive and the extensive margin*

We now turn to explore heterogeneous treatment effects. As highlighted in the theoretical section, whether people increase or reduce their effort depends on a number of underlying conditions and parameters: whether the update is positive or negative, their individual reference points  $\bar{r}_i$ , and what preferences they have over their own rank with respect to their reference point, i.e. the value of their  $\alpha_i$ 's and  $\beta_i$ 's. First we follow the prior literature and heterogeneous treatment effects in competitive environments by gender. In other studies (Gneezy et al. 2003, Niederle et al. 2007) it has been found that women shy away from competition and perform differently when in them.<sup>21</sup> Next we can study if the treatment effect on effort of telling employees their rank order position differs by gender. As can be seen in column 2 of Table 2 we find no heterogeneous treatment effect by gender. In Table 3 column 2 we also explore if there is a heterogeneous treatment effect by gender on the intensive margin and again we find no difference.

Finally we take the model more seriously and see if (i) those with a revealed preference for rank and (ii) those with an unexpectedly positive ( $\hat{r}_i > \bar{r}_i$ ) update compared to their prior belief about their rank raise or lower their effort after the

---

<sup>20</sup> One could also use hurdle models to jointly estimate the extensive and the intensive margin. Using simple regression models conveys the result in a more accessible way. Hurdle models generate qualitatively the same results.

<sup>21</sup> As was shown earlier in the paper there is no difference between men and women in their preference to get feedback. See the Appendix and in particular Table A1 for further details.

update. As highlighted in the model it is among these workers that we should expect to see an effort response since those who don't care about rank should not change their behavior in light of and those who get feedback that is in line with their prior also won't update their behavior as the feedback contains no new information.

It is very difficult to have an independent measure of whether people care about their rank, what their reference point is, and what their prior belief about their rank is. Yet, we made some first attempts into this direction that should inspire further exploration to better elicit these important underlying parameters.

In this paper we attempted to approach this challenge in the following way. We measure the preference over rank by whether workers chose the feedback job in stage 1 or by whether they have a stated preference for feedback.<sup>22</sup> As this could proxy for a general propensity to work more it is important to control for the main along with the interaction effects. In column 3 of Table 2 we added these variable and also interacted them with the treatment dummy. We find no evidence that either having a stated or a revealed preference for feedback about rank has a significant differential treatment effect.<sup>23</sup> This result is also confirmed when studying the intensive margin. In Table 3 column 3 we test if these preference terms had a differential treatment effect on the quantity worked but find no evidence for it.

---

<sup>22</sup> As explained above we tried in pre-tests other ways to get a more direct measure of their reference points and whether they (dis-)like to be ahead or behind their reference point. Subjects were however confused by what was being asked and also revealed an incentive to misreport information in the questionnaire.

<sup>23</sup> We also ran regressions adding only the main and interaction effects of the stated preference or only for the revealed preference with qualitatively the same result.

We finally attempt to estimate the effect of getting an unexpectedly positive feedback about rank performance. In terms of the model this refers to the case when the update about the rank order position is higher than the prior belief. As we have no good data about prior beliefs we make the following proxy. We create a dummy variable equal to one when a worker appeared in the Top 10 in any of the four accuracy rankings.<sup>24</sup>

In column 4 of Table 2 we see that the interaction term between *rank feedback* and *Top 10 performance* is not significant. One can do one final test here to see if giving unexpectedly positive feedback is less destructive to performance than giving feedback in general. For this we do a joint test the null hypotheses if the sum of the coefficients on *Rank Feedback* plus  $(Top\ 10\ Performance) * (Rank\ Feedback)$  equal to zero but we fail to reject the null ( $\chi^2(1) = 0.63$ , p-value 0.4281).

We can confirm the same results as we study the intensive margin in column 4 of Table 3. Again the interaction term on  $(Top\ 10\ performance) * (Rank\ Feedback)$  is not significantly different from zero but the joint test of whether the sum of the coefficients on  $(Rank\ Feedback) + (Top\ 10\ Performance) * (Rank\ Feedback) = 0$  can not be rejected ( $F(1, 143) = 0.77$ , p-value 0.3805).

Taken together, although we find that giving feedback about rank reduces performance both on the intensive and the extensive margin we at least find that

---

<sup>24</sup> As this could measure ability it is important to control for the main along with the interaction effect. This variable is equal to one when a person appeared in any of the four accuracy rankings within the top 10 the mean on this dummy variable is 0.2993 with a standard deviation of 0.4588. We experimented with a number of different definitions for this variable all yielding the same qualitative results: We defined alternative variables as to whether people are in the top and/or the bottom 25 in terms of ranking; if they were in the top or the bottom 10% or 25% in the accuracy performance distribution; and broke down the results by the ranking on each of the four categories.



giving positive feedback (i.e. telling workers that they are in the Top 10) does not significantly reduce performance. One should be cautious when interpreting this last result but it suggests that feedback does not always reduce performance.

As mentioned before these metrics are not perfect but struck us as the right balance at this stage in terms of precision and statistical power. Future research should further explore the correct measurement of reference points, of preferences parameters for rank, and workers' prior beliefs more precisely – perhaps with incentivized revelation mechanisms.

Table 2: Effect of feedback about rank order position on the extensive margin

Dependent variable is equal to one when person returns to do more work after being sent an invitation email	(I) Baseline	(II) Gender	(III) Rank Preference	(IV) Top 10 news
Rank Feedback	-0.8297*** (.2449)	-0.9892** (.4831)	-.2649 (.7788)	-1.132*** (.3021)
Gender		0.0488 (.4691)		
Gender*(Rank Feedback)		-0.1864 (.6519)		
Revealed Rank Preference			0.7214 (1.151)	
(Revealed Rank Preference)* (Rank Feedback)			-1.282 (1.269)	
Stated Rank Preference			-.0941 (.2447)	
(Stated Rank Preference)* (Rank Feedback)			-.2571 (.3581)	
Top 10 performance				-2.069 (.4190)
(Top 10 performance)* (Rank Feedback)				.7196 (.6293)
Number of observations	284	191	191	284
Pseudo R <sup>2</sup>	0.034	0.052	0.074	0.131

Notes: These are logistic regressions in which the estimated coefficients are reported (rather than odd ratios) with bootstrapped standard errors with 1000 replications. The data is on all workers who worked in the first stage of the experiment who were then sent one of the two email. See the appendix for the precise wording of the emails. The dependent variable is a dummy variable equal to unity when the worker chose to return to work after the email. *Rank Feedback* is the treatment index variable equal to unity when the invitation email contained feedback about rank order position in terms of the accuracy of their work; the control group received an email without rank information. The sample in column 2 and 3 is somewhat smaller as it uses data from the worker survey. *Revealed Rank Preference* is equal to one when a worker chose the job with offered feedback in the first stage of the experiment. *Stated Rank Preference* is equal to one when a worker confirmed in the survey to like getting feedback about their rank *Top 10 performance* is equal to one when a worker has been ranked among the Top 10 along any of the four accuracy measures. See the Appendix and the text for further details and definition of variables.

Table 3: Effect of feedback about rank order position on the intensive margin

Dependent variable measures the quantity worked conditional on returning to work in the second stage.	(I) Baseline	(II) Gender	(III) Rank Preference	(IV) Top 10 news
Rank Feedback	-4.275** (1.688)	-4.319 (2.958)	-7.920** (3.927)	-4.238*** (1.645)
Gender		-1.204 (1.890)		
Gender*(Rank Feedback)		-.9013 (3.915)		
Revealed Rank Preference			0.6739 (1.976)	
(Revealed Rank Preference)* (Rank Feedback)			3.736 (4.124)	
Stated Rank Preference			-1.887** (.8709)	
(Stated Rank Preference)* (Rank Feedback)			0.9217 (1.974)	
Top 10 performance				-14.70*** (2.635)
(Top 10 performance)* (Rank Feedback)				1.321 (3.700)
Number of observations	147	121	121	147
R <sup>2</sup>	0.0457	0.0692	0.1038	0.3008

Notes: The data used in these regressions is based on those workers who return to work after being send an invitation email at the start of the second stage. All workers who worked in the first stage of the experiment were then sent one of the two emails. See the appendix for the precise wording of the emails. The dependent variable measures the number images completed by a worker. These are OLS regressions with robust standard errors. Using count data regressions generates qualitatively the same results. See Table 2 for further information about the definitions of the variables.

## Conclusions

There has been a lot of interest both theoretically and empirically in whether people derive a utility from their rank. This behavioral pattern is very interesting for contract and policy design as it could potentially elicit effort when monetary incentives cause unintended consequences or are ineffective. Following a prior literature on social psychology and economics we aimed to provide evidence from a field experiment on how effort changes both along the intensive and the extensive margin after workers are informed about their rank. We find that telling people their rank significantly reduces both the intensive and the extensive margin in a context where rank had no bearing on their current or future compensation.

The results have been established in one specific context in this paper but the same results have been found in two other settings. Using data from a large office furniture company in the US and from medical authors writing for a clinical online encyclopedia (Barankay, 2010) studied a very similar field experiment and qualitatively the same results have been found that telling people about their rank reduces their effort. Thus there is evidence from several settings that feedback about rank reduces performance and that these results are not driven by the specificities of Amazon's Mechanical Turk.

Studying social preferences in general and rank preferences in particular is appealing for a number of reasons. First it is a more realistic representation of human concerns and behavior in the workplace. Second, it received a lot of attention in education and health policy without much prior empirical investigation.

Third, social preferences are potential candidates to help eliciting effort when monetary incentives fail or show unintended consequences.

The clean structure and the strong control of this paper's setting helped to highlight both the extensive and the intensive margin at play here. The field experimental design also helped to show the aggregate detrimental treatment effect. The underlying theory explained how the direction of the effort response is very sensitive to the underlying parameters. In many settings however these underlying parameters (the worker's reference points, prior beliefs and their preference over rank) are available neither to the econometrician nor the policy maker so the hope to get implement rank feedback targeted only towards those who will respond positively to is not realistic.

Future work needs to provide more evidence from other settings and also explore whether the underlying parameters can be recovered to pinpoint more detailed mechanisms in the data. Only then can we establish if targeted feedback that takes into account the underlying heterogeneities can be established to generate a positive causal effect on performance. At this stage the aggregate result is that feedback about rank is detrimental to performance.

## References

- Auriol, Emmanuelle, Régis Renault (2008), "Status and Incentives," *RAND Journal of Economics*, vol. 39(1), 305-326.
- Azmat, Ghazala, and Nagore Iriberry (2010), "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students," *Journal of Public Economics*, forthcoming.
- Bandiera, Oriana, Iwan Barankay and Imran Rasul (2010a), "Social Incentives in the Workplace." *Review of Economic Studies*, 77 (2), 417 - 458.
- Bandiera, Oriana, Iwan Barankay and Imran Rasul (2010b), "Team Incentives: Evidence from a Field Experiment," Mimeo, University of Pennsylvania.
- Besley, Timothy, and Maitreesh Ghatak (2008), "Status Incentives," *American Economic Review Papers and Proceedings*, 98(2), 206—211.
- Blanes i Vidal, Jordi, and Mareike Nossol (2009), "Tournaments without Prizes: Evidence from Personnel Records," Mimeo, London School of Economics.
- Dohmen, Thomas and Armin Falk (2010), "Performance Pay and Multi-Dimensional Sorting: Productivity, Preferences, and Gender," *American Economic Review*, forthcoming.
- Durkheim, Emile (1951), *Suicide: A Study In Sociology*. The Free Press, NY: 1951.
- Ehrenberg, Ronald G and Bognanno, Michael L (1990), "Do Tournaments Have Incentive Effects?," *Journal of Political Economy*, 98(6), 1307-1324
- Green, Jerry R., and Nancy L. Stokey (1983), "A comparison of Tournaments and Contracts," *Journal of Political Economy*, 91(3), 349-364.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini (2003), "Performance in Competitive Environments: Gender Differences," *Quarterly Journal of Economics*, 118(3), 1049-1074.

Gneezy, Uri, Kenneth L. Leonard, and John A. List (2009), "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society," *Econometrica*, 77(5), 1637-1664.

Lazear, Edward P. and Sherwin Rosen (1981), "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89(5), 841-864.

Moldovanu, Benny, Aner Sela, Xianwen Shi (2007), "Contests for Status," *Journal of Political Economy*, 115(2), 338-363.

Niederle, Muriel, and Lise Vesterlund (2007), "Do women shy away from competition? Do men compete too much?" *Quarterly Journal of Economics*, 122(3), 1067-1101.

Veblen, Thorstein (1949), *The Theory of the Leisure Class – An Economic Study of Institutions*, George Allen & Unwin.

Figure 1: Example of a picture





# Appendix

## Job Descriptions

Title as it appears on the Mechanical Turk job listing:

Job 1: "Answer four questions relating to a picture. No typing required."

Job 2: "Answer four questions relating to a picture and get feedback on your rank in terms of accuracy. No typing required."

Further description on the webpage with the work (HIT) itself:

"To improve our database as part of an environmental university research project, please answer the following questions relating to the picture below. There is no need to type. Simply select your answers by clicking on a radio button. The quality of your answers is very important to us. We greatly appreciate the diligence you devote to this task."

## Wording of the questions in the jobs

Questions workers were asked to answer the following four questions:

Question 1. Is the picture sharp and in focus?

A1 ALL parts of the picture are sharp and in focus.

A2 SOME parts of the picture are sharp and in focus.

A3 NO part of the picture is sharp and in focus.

A4 Don't Know/Can't tell.

Question 2. Can you see any clear cracks or lines in the picture?

A1 Yes

A2 No

A3 Don't Know/Can't tell.

Question 3. We want to know how long, in total, the lines or cracks are. What is the TOTAL sum of all lines or cracks on the picture as compared to the WIDTH of the picture? When there are several cracks or lines add them up and then compare that total length to the width of the picture. Note: The width measures the extent of the pictures from left to right. Example: When the sum of all lines is less than a quarter of the pictures width it is less than 25% of the picture's width.

The sum of all crack or lines are...

A1 ...less than 25% of the picture's width.

A2 ...25% to 50% of the picture's width.

A3 ...50% to 75% of the picture's width.

A4 ...50% to 75% of the picture's width.

A5 ...75% to 100% of the picture's width.

A6 ...100% to 150% of the picture's width.

A7 ...more than 150

A8 Don't know/Can't tell

A9 There are no lines or cracks

Question 4. The overall quality of the pictures in our data-base is very important to us. How would you rate the overall quality of the picture? Note: Please try to ignore the sharpness of the picture when judging the overall quality.

A1 Well Below Average

A2 Below Average

A3 Average

A4 Above Average

A5 Well Above Average

A6 Don't know/Can't tell

#### Further evidence on selection into the feedback and the no feedback job

Table A1 explores with logistic regressions with reported coefficients whether survey based characteristics help predict the attributes of worker who choose the feedback job. Thus basic demographic characteristics do not differ across the two groups thus not inducing a selection bias along those dimensions. Thus we cannot, for instance, confirm in this field experiment recent laboratory evidence (Niederle et al 2007) that women prefer to shy away from competitive environments.<sup>25</sup> We can also confirm in this logistic estimation that the stated preference for feedback holds no predictive power for the revealed preference for feedback.

---

<sup>25</sup> This statement is made under the common interpretation in the social psychology literature that one is willing to interpret the presence of rank feedback to be indicative of a more competitive environment.

### Table A1: Who chooses the rank feedback job?

Logistic regressions with the dependent variable equal to one when a worker chose the feedback job in the first stage of the experiment. See notes below for further details.

Female	-.0516	(.3534)
Age	.0154	(.0134)
High education	-.5534	(.3407)
Experience	-.0001	(.0003)
Risk	.0089	(.0824)
Patience	-.1869	(.3487)
Likes rank feedback	.0345	(.1623)
Number of observations	227	
Pseudo R <sup>2</sup>	.0196	

Note: Logistic regressions with estimated coefficients. Bootstrapped standard errors with 1000 replications. *High education* is a dummy equal to unity when a worker has at least a completed high school; coding this variable differently does not alter the result. *Experience* is the level of experience the worker has on Mechanical Turk. *Risk* is a dummy variable equal to one when the person stated to be risk loving in the survey. *Patience* is a dummy variable equal to one when the person stated to be patient in the survey. *Likes rank feedback* is a dummy variable equal to one when the person agreed to the statement in the survey “[y]ou want to receive feedback about your relative performance, i.e. how your performance ranked compared to others”.

Next, one can explore a possible selection effect by looking at whether the quantity or the quality of work was any different between those who chose the feedback job and those who chose the no feedback job. Quantity is simply measured by the number of pictures done by a worker. In the no feedback job workers did on average 21 pictures and in the feedback job they did 23 pictures, which is not significantly different at the 5% level. Quality is measured by the fraction of question that the worker answers accurately. Note that the first three questions (whether the picture is in focus; whether it contains a crack or line; and how long the crack is) the correct answer can be objectively determined. We find no evidence for a difference in terms of accuracy in the work done for the feedback and the no feedback job.

Table A2: Accuracy of work in the feedback and in the no feedback job

	No rank feedback	With rank feedback
Is the picture in focus?	.75 (.15)	.76 (.15)
Is there are a line/crack?	.78 (.21)	.79 (.22)
How long is the line/crack?	.43 (.20)	.45 (.19)

Note: Each cell gives the fraction of all questions that was answered correctly. See the appendix for the complete wording of the questions and possible answers. Based on data from 254 workers who chose the no rank feedback job and the 76 workers who chose the job where we offered rank feedback.

Wording of the email sent to the workers at the beginning of the second stage of the experiment

*Email sent to workers in the control group*

“Hi,

We greatly appreciate the work you did for us recently.

Thank you for completing the assignment on Mechanical Turk for us. We would like you to answer a set of HITs similar to the ones you have already answered. Please go to the link provide below, after logging into Mturk.com, where we assigned a HIT for you.

Note: All workers who worked on these HITs were given this qualification.

[URL to the job]”

*Email sent to workers in the control group*

“Hi,

We greatly appreciate the work you did for us recently. Here is a feedback of your work.

You were ranked ww out of WWW workers in terms of accuracy on the first question.

You were ranked xx out of XXX workers in terms of accuracy on the second question.

You were ranked yy out of YYY workers in terms of accuracy on the third question.

You were ranked ZZ out of ZZZ workers in terms of accuracy on the fourth question.

Thank you for completing the assignment on Mechanical Turk for us. We would like you to answer a set of HITs similar to the ones you have already answered. Please go to the link provide below, after logging into Mturk.com, where we assigned a HIT for you.

Note: All workers who worked on these HITs were given this qualification.

[URL to the job]”