



University of Pennsylvania
ScholarlyCommons

Management Papers

Wharton Faculty Research

2011

Rankings and Social Tournaments: Evidence From a Crowd-Sourcing Experiment

Iwan Barankay
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/mgmt_papers



Part of the [Management Sciences and Quantitative Methods Commons](#)

Recommended Citation

Barankay, I. (2011). Rankings and Social Tournaments: Evidence From a Crowd-Sourcing Experiment. Retrieved from https://repository.upenn.edu/mgmt_papers/240

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/mgmt_papers/240
For more information, please contact repository@pobox.upenn.edu.

Rankings and Social Tournaments: Evidence From a Crowd-Sourcing Experiment

Abstract

There is a growing interest in behavioral incentive schemes exploiting people preference about how they rank compared to others as a non-monetary mechanism to shape effort. In this paper we present evidence from a crowd-sourcing 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, employees who received feedback about their rank were less likely to return to work and also less productive on the job.

Keywords

Rank Incentives, Social Tournaments, Feedback, Field Experiment

Disciplines

Management Sciences and Quantitative Methods

Rankings and Social Tournaments: Evidence from a Crowd-Sourcing Experiment^{*}

Iwan Barankay[†]

University of Pennsylvania, CEPR and IZA

September 12, 2011

Abstract:

There is a growing interest in behavioral incentive schemes exploiting people preference about how they rank compared to others as a non-monetary mechanism to shape effort. In this paper we present evidence from a crowd-sourcing 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, employees who received feedback about their rank were less likely to return to work and also less productive on the job.

Keywords: Rank Incentives, Social Tournaments, Feedback, Field Experiment.

JEL Codes: **D23, J33, M52**

^{*} This paper benefited from financial support by the Economic and Social Research Council, U.K., and the Mack Center at the Wharton School. I thank Peter Cappelli, and 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 for valuable suggestions. The usual disclaimer applies. Melvin Kanasseril provided outstanding programming support.

[†] The Wharton School, University of Pennsylvania, 3620 Locust Walk, 3620 Locust Walk, SHDH Suite 2000, Philadelphia, PA-19104, Tel: +1 215 898 6372. Email: barankay@wharton.upenn.edu.

Introduction

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 others.

Rankings are a very common and prominent feature of workplace environments. In car-dealerships 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. More importantly perhaps, rankings are becoming more prevalent to evaluate teachers or medical professionals and consulted to determine the allocation of funding – both public but also private in the form of donations – to schools, hospitals, or universities.

The literature about the effect of providing information about rankings on performance is still developing. To study such rank preferences is therefore important both theoretically and empirically.

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?

We designed a field experiment on a crowd-sourcing webpage 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. In this context, rankings explicitly did not affect their current or future compensation. For the purpose of this experiment, workers are recruited via Amazon's crowd-sourcing website Mechanical Turk (www.mturk.com). Crowd-sourcing is an important emerging employment phenomenon whereby employees log on to a website to select work offered to them online. 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.

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 its quality. The key treatment was to randomize employees into two groups. One group received feedback about how they rank in terms of performance of their work. The other group received no feedback about their rank.

Three main results emerge. First, comparing the behavior between the control and the treatment group, those who receive feedback are 30% less likely to return to work. Second, the productivity of those who were told their rank after

they return is 22% lower compared to those who returned to work but were not told their rank. Third, we explored heterogeneous treatment effects along a number of observable characteristics such as, age, prior ranking, self-reported rank preference and whether they received exceptionally high or low ranking feedback but find no heterogeneous treatment effects using these proxies. Those with self-reported 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.

Building on insights in sociology and social psychology there is now a rich theory in economics on the role of self-image (Benabou and Tirole, 2003, Koszegi, 2006), social status (e.g. Robson, 1992, Becker et al, 2005 Ellingsen and Johannesson, 2007, Frey, 2007, Moldovanu, et al, 2007, Auriol and Renault, 2008, Besley and Ghatak, 2008, Dur, 2009, Dubey and Geanakoplos, 2010, Ederer and Patacconi, 2010), equity theory (Adams, 1965) and identity (Akerlof and Kranton, 2005). This paper also builds on theoretical work on interim performance feedback (Lizzeri et al., 2002, Aoyagi, 2010) that highlights how its effect can be heterogeneous across agents but also points out that the optimality of feedback hinges on whether agents know their ability and the shape of their cost of effort functions (Aoyagi, 2007, Ederer, 2010).

A meta-analysis of psychology studies about feedback interventions¹ covering some 131 studies with over 13,000 subjects by Kluger and Denisi (1996) revealed that the effect of feedback on performance is very heterogeneous. Even though feedback improved performance on average, it reduced performance in one third of the surveyed studies.

¹ See also Smither et al., 2005.

Given this heterogeneity in response to rankings and in attitudes towards self-image there is scope to clarify empirically the patterns behind these differences. There are a few papers that studied the effect of status and performance feedback with field experiments and quasi-experiments. An early paper is Greenberg (1988) who has data from a field experiment of 198 employees in an insurance under-writing department. The firm renovated offices of that company during which employees had to be relocated to other offices. The clever aspect of this paper is that this relocation has been random so that employees have been moved to offices that correspond to the same, lower, or higher pay-grades. Compared to those employees who were relocated to office in line with their current pay-grade, those reassigned to higher status offices increased their performance whereas those assigned to lower-status offices reduced it. Under the perhaps strong assumption that the status of the office has no direct productive impact on performance, the effect of the relocations on performance can be attributed to the behavioral response to variations in status corresponding to each office type.

The effect of information about relative performance has been studied experimentally in the context of electricity consumption (Costa et al, 2010), job satisfaction (Card et al, 2010)², and it speaks to work on interpersonal wage comparisons (Cohn et al, 2011).

Another closely related study from the education context is Azmat and Iriberry (2010) who show how relative performance feedback raises high school student's educational attainment using data from a naturally occurring change in some Spanish school districts. The interpretation of their results needs, however,

² For evidence from event studies without randomized contemporaneous control and treatment groups see Bandiera et al (2011b) and Blanes i Vidal and Nossol (2010)

to be cast within the education context. Feedback in an education setting conveys information about ability, and thus resolves an uncertainty about the return to effort, and not just about relative performance. Furthermore their results may be driven by changes in behavior of the parents rather than the students themselves. In contrast to that paper the field experiment described here only gives feedback about rank in a setting where employees have full information about absolute performance bringing the interpretation more in line with self-image and status.

Rankings are often used to hand out symbolic awards which has been studied by Kosfeld and Neckermann (2011) who hired students to enter data for three weeks as part of a non-governmental organization project. The treatment was to honor the best performance publically with a symbolic award. They find that the award treatment raises performance by 12%. In their experiment the authors left vague vis-à-vis the student what the exact criteria were to determine the award winners. The authors argue that the results are due to the status conveyed by the award but it may be possible that the award component of the task made the job more interesting to the students thus raising their intrinsic motivation.

In a notable and complementary laboratory experiment³, Charness et al (2010) study whether people are willing to learn about and to also improve their rank when rank conveys no monetary benefits. They find that simply informing subjects about their rank increases their performance. Their result is in contrast to the findings of this paper but despite the similarity of the set-up there are

³ Other laboratory experiments involving feedback about rank are Hannan et al (2008), Freeman and Gelber (2010), and Kuhnen and Tymula (2011).

many differences in the population, the nature of the task, the scope of sabotaging each other's work, and the duration of the experiment.

In a companion paper to this field experiment, Barankay (2011) replicated the main effect of a reduction in performance due to rank feedback. The setting of that paper were furniture salespeople who are paid according to a commission rate scheme based on their absolute performance. For a given level of sales a change in rankings had no impact on sales. He found that those salespeople who were informed about their rank sold over 20% less furniture.

Another closely related field experiment is by Delfgaauw et al. (2010). In their study they introduce a tournament, with or without prizes, in a team setting among stores of a clothing chain. They find weak overall effects of the tournament but show that when the gender of the store-keeper is aligned with that of the employees, the tournaments raise performance more. In contrast to that work, employees who work in my experiment and are compensated as individuals and only the information about rank is experimentally varied.

A conclusion of my study is that more information is not always better for the performance of agents which speaks to results from earlier laboratory experiments (Camerer et al, 1989). In a sequence of experiments, Loewenstein et al (2006), for instance, document how more information can reduce the performance and earnings even when experimental subjects are willing to pay for that information. In one of the experiments, a significant number of participants pay for information – the solution to a puzzle – that then hurts their ability to predict how many others will solve that puzzle. The adverse effect of more information also speaks to the concept of curiosity in psychology whereby people are in a state of deprivation due to the lack of information. One salient

characteristic of curiosity is, however, its tendency to disappoint when satisfied (Loewenstein, 1994).

The result that more information lowers performance may also motivate policies based on asymmetric paternalism (Thaler and Sunstein, 2003, Loewenstein et al, 2007) where an attempt is made to help individuals achieve their own goals, which requires protecting them from themselves and from information they would seek.

The rest of the paper is structured as follows. In the next section we explain the theoretical intuition behind the results. We then describe the context and experimental design. This is followed by an in-depth analysis of the main treatment effect and the mechanism behind. Finally we conclude with a discussion of future work.

Theoretical Framework

To help organize the results of this paper, we set up a stylized framework that is inspired by the theoretical literature reviewed above. Suppose salesperson i 's payoff contains three components. First, she derives utility from the monetary benefit of her effort. This benefit $b_i(e_i)$ approximates the commission based compensation scheme salespeople face and reflects how their effort maps into commission payouts. Second, each worker has a convex cost of effort, $c_i(e_i, \vartheta_i)$ which depends on the worker's ability ϑ_i . In the third element, we assume that workers derive utility from the performance rank they have $r(b_i, \mathbf{b}_{-i})$, which depends on the monetary payout she receives, b_i and that of all the other

salespeople captured by the vector \mathbf{b}_{-i} . This third component of the payoff function generates *rank incentives*.

Salespeople choose their effort simultaneously to maximize their payoff

$$e_i^* = \operatorname{argmax}_{e_i} \{b_i(e_i, m_i) - c_i(e_i, \vartheta_i) + r(b_i, \mathbf{b}_{-i})\}.$$

We assume that the monetary benefit increases in effort, hence $\partial b_i(e_i, m_i)/\partial e_i \geq 0$, but the marginal benefit from a change in rank $\partial r(b_i, \mathbf{b}_{-i})/\partial e_i$ can be more flexible. It can be positive when people want to be better than others but it can be negative when they want to conform to a low performance norm by lowering their rank. For completeness, it is also possible that they may only care about their monetary payoff for their work $b_i(e_i, m_i) - c_i(e_i, \vartheta_i)$ but not about their rank. This set-up can now be adjusted to capture the design of the experiment in this paper. In our study, we randomize salespeople into two groups. In one group people are privately told their rank and in the other group they are not. Those who are *not* told their rank form a belief, based on their information and prior experience, about how their effort maps into the ranking and what they estimate the other salespeople's effort to be. Those in the other group, who *are* told their rank, actually go through three stages. First they choose their effort, based on a prior belief r_i^b of how their effort will map into a rank, second, they learn the realization of their rank, i.e. how their effort actually did map into a rank, and third they adjust the effort based on the update.

We can thus classify the reaction to this update into two categories depending on how they enter into the salesperson's payoff. First, the update may be *good news* in the sense that the actual rank was better than their prior belief. Denote $\alpha(r_i, r_i^b; r_i > r_i^b)$ to be the *marginal change in effort*, where $\alpha(\cdot, \cdot)$ is a parameter that can be positive or negative. Second, the update may be *bad news* in the sense that the realized rank was less than the prior belief and the marginal

change in effort resulting from this is $\beta(r_i, r_i^b; r_i > r_i^b)$, where $\beta(\cdot; \cdot)$ is a parameter that again can be positive or negative. There is also a third possibility that agents' update is exactly in line with their prior belief and we normalize their effort response then to be zero.

More generally $\alpha(\cdot; \cdot)$ and $\beta(\cdot; \cdot)$ could be non-degenerate functions but given the sample sizes in our experiment we only have statistical power to estimate a parameter and test for its statistical significance.

We can then distinguish four empirical predictions in the presence of *rank incentives*.

First, when $\alpha(\cdot; \cdot) > 0$ there is an *encouragement or motivating effect* whereby agents will try to catch up even though they are told that they fell short of their expected rank.

Second, when $\alpha(\cdot; \cdot) < 0$ there is a *discouragement or demoralization effect* whereby agents reduce their effort when they learn that they fell short of their expected rank. This can have two interpretations. First, learning about a low rank may reduce the enjoyment derived from the job and thus reduces the intrinsic motivation of the agent. Similarly there is the possibility that agents learn something constructive from negative feedback which encourages them to exit and to pursue other opportunities.

Third, $\beta(\cdot; \cdot) > 0$ is an *acceleration effect*, whereby people who are positively surprised by the rank they achieved are driven to excel even more.

Finally, when $\beta(\cdot; \cdot) < 0$ there is a *complacency effect* whereby salespeople who learn that they are better ranked than expected slack off.

Here we presented a very reduced-form representation of these predictions but each of these four cases has been derived in the theoretical literatures on self-image (Benabou and Tirole, 2002, Koszegi, 2006), and interim performance

evaluation (Lizzeri et al, 2002, Aoyagi, 2010). They have also been described in observational studies about feedback (Kluger and DeNisi, 1996, Smither et al, 2005). The experiment below was designed to test for the significance and the sign of these parameters and the mechanism behind them.

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.

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.⁴ To be able to earn money, workers currently need to have a bank account either in the US or in India.⁵

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.

⁴ 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.

⁵ 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.

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 experiment⁶ and that experiments are conducted in a natural 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.⁷ 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

⁶ 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.

⁷ 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.

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 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.⁸

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.

⁸ 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+
Frequ.	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
Frequ.	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
Frequ.	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%

Risk⁹ - 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%

Patience¹⁰

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

⁹ 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.")"

¹⁰ 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,¹¹ 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 abstract,

¹¹ 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.

uninteresting or even of questionable use,¹² 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.¹³ 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.”

¹² 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.

¹³ 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.

Treatment Effects

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,¹⁴ 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 discrepancy between stated and revealed preferences.

¹⁴ The jobs were posted twice for 24 hours each on different days to deal with potential day of week effects.

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 workers¹⁵ and which job the worker chose is only 0.0308.

Clearly, the stated and revealed preferences are almost 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.¹⁶ 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. They knew, therefore, that compensation did not depend on the quality – the accuracy – of their work.

¹⁵ 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.

¹⁶ 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.

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 and that ranking was not relevant for compensation, as we wanted to make sure they don't feel selected in any way beyond the fact that they worked for us before.¹⁷ 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.

Main 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

¹⁷ 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.

more work when invited to do so – and the intensive margin, the quantity and the quality 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 might dissuade some workers to come back to work but might help to attract back high performing workers. 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.¹⁸

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 especially on whether the

¹⁸ 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.

update is positive or negative. As was shown above observable characteristics do not predict whether workers have a preference to get feedback.¹⁹

Finally we take the model more seriously and see if those with a revealed preference for rank and those with a positive or a negative update compared to their prior belief about their rank raise or lower their effort after the update. As we highlighted in the model it is among these workers that we should expect an effort response: those who don't care about rank should not change their behavior in light of feedback about their rank if it does not affect their compensation; 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.²⁰ As this could proxy for a general propensity to work more it is important to control for the main along with the interaction effect. In column II 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

¹⁹ See the Appendix and in particular Tabel A1 for further details.

²⁰ 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.

rank has a significant differential treatment effect.²¹ 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.

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.²²

In column III 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 to see if giving positive feedback is less destructive to performance than giving feedback irrespective of performance. For this we do a joint test if the sum of the coefficients on *Rank Feedback* plus $(\text{Top 10 Performance}) * (\text{Rank Feedback})$ is significantly different from zero. We fail to reject the null that the sum of these coefficients is equal to zero ($\chi^2(1) = 0.63$, p-value 0.4281).

We can confirm the same results as we study the intensive margin in column III of Table 3. Again the interaction term on *Top 10 performance* and *Rank Feedback* is not significantly different from zero but the joint test of whether the sum of the

²¹ 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.

²² 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. We experimented with a number of different definitions for this variable all yielding the same qualitative results notable. We tried 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.

coefficients on *Rank Feedback* plus *(Top 10 Performance)*(Rank Feedback)* is equal to zero 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 giving unexpectedly positive feedback (i.e. telling workers that they are in the Top 10) does not significantly reduce performance. One should be cautious in 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. Future research should further explore the correct measurement of reference points and preferences parameters 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) Rank Preference	(III) Top 10 news
Rank Feedback	-0.8297*** (.2449)	-.2649 (.7788)	-1.132*** (.3021)
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	284
Pseudo R ²	0.034	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) Rank Preference	(III) Top 10 news
Rank Feedback	-4.275** (1.688)	-7.920** (3.927)	-4.238*** (1.645)
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	147
R ²	0.0457	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 is a growing literature both theoretically and experimental that studies whether people derive a utility from their rank. These behavioral patterns are very interesting for contract and policy design as they could potentially elicit effort when monetary incentives cause unintended consequences or are ineffective. Following a prior literature on social psychology we aimed to provide evidence from a field experiment using a crowd-sourcing webpage on how effort changes of workers both along the intensive and the extensive margin when they are told their performance rank. We find that telling people their rank reduces both the intensive and the extensive margin in a context where rank had no bearing on their current or future compensation.

This paper complements other laboratory and field experimental studies. One particular contribution is to show the impact of rank feedback on the extensive margin, i.e. differential attrition. In many laboratory studies, subjects can only respond along the intensive margin, i.e. productivity, and thus the results may differ if attrition would be permitted.

Studying social preferences in general and rank preferences in particular is important for several 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, rank incentives could be potential candidates for behavioral incentive schemes to shape habit formation.

One can derive two general policy prescriptions from this study.

The first lesson is that if we are to envisage behavioral incentive schemes based on rank feedback it is imperative to elicit whether people are prone to complacency and demoralization effects, as was the case in our context, or whether they will respond positively to rank feedback.

The second lesson, and a suggestion for future work, is to see if there exists an alternative framing of rank feedback that will elicit a more positive productivity

response from the one we used in this study. Possible candidates here could be to attach rank information to goals and benchmarks. Also, public instead of private feedback about rank may be able to elicit a different response.

Given the large behavioral response generated by such a trivial and cost-effective intervention it seems a fruitful area for more field experiments.

References

- Adams, J.S. (1965): Inequity in Social Exchange. in L. Berkowitz (ed.) *Advances in Social Psychology*, 2, 267-299. New-York: Academic Press.
- Akerlof, G.A., Kranton, R., 2005. Identity and the economics of organizations. *Journal of Economic Perspectives* 19, 9–32.
- Aoyagi, Masaki. 2010. Information feedback in a dynamic tournament. *Games and Economic Behavior*, 70(2), 242-260
- 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*, 94(7-8), 435-452.
- 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, Imran Rasul. 2011. Team Incentives: Evidence from a Field Experiment. *Mimeo*, University of Pennsylvania.
- Barankay, Iwan. 2011 “Gender differences in productivity responses to performance rankings: Evidence from a randomized workplace experiment. *Mimeo*, University of Pennsylvania.
- Becker, Gary S., Kevin M. Murphy, and Ivan Werning. 2005. The equilibrium distribution of income and the market for status, *Journal of Political Economy* 113, 231–248.
- Benabou, R., Tirole, J., 2003. Intrinsic and extrinsic motivation. *Review of Economic Studies* 70, 489–520.
- Besley, Timothy, and Maitreesh Ghatak. 2008. Status Incentives. *American Economic Review*, 98(2), 206–11.
- Blanes i Vidal, Jordi, and Mareike Nossol. 2010. Tournaments without Prizes: Evidence from Personnel Records, with Mareike Nossol. *Management Science*, forthcoming.
- Camerer, Colin, George Loewenstein, and M. Weber. 1989. The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, 97, 1232-1254.
- Card, D., A. Mas and E. Moretti. 2010. Inequality at Work: The Effect of Peer Salaries on Job Satisfaction. *NBER Working Paper* 16396.
- Charness, Gary, David Masclet and Marie Claire Villeval. 2010. Competitive Preferences and Status as an Incentive: Experimental Evidence. *IZA Discussion Paper* 5034.

- Costa, D., M. E. Kahn. 2010. Energy Conservation "Nudges" and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment. *NBER Working Paper No. 15939*.
- Delfgaauw, Josse, Robert Dur, Joeri Sol, and Willem Verbeke. 2010. Tournament Incentives in the Field: Gender Differences in the Workplace. *Working paper* Erasmus U Rotterdam
- Dubey, Pradeep, and John Geanakoplos. 2010. Grading Exams: 100, 99, 98, ... or A, B, C? *Games and Economic Behavior*, 69(1): 72–94.
- Dur, Robert. 2009. Gift Exchange in the Workplace: Money or Attention? *Journal of the European Economic Association*, 7(2–3): 550–60.
- Ederer, Florian. 2010. Feedback and Motivation in Dynamic Tournaments. *Journal of Economics & Management Strategy*, 19(3), 733-769.
- Ederer, Florian, and Andrea Pataconi. 2010. Interpersonal Comparison, Status and Ambition in Organizations. *Journal of Economic Behavior and Organization*, 75(2): 348–63.
- Ehrenberg, Ronald G and Bognanno, Michael L. 1990. Do Tournaments Have Incentive Effects? *Journal of Political Economy*, 98(6), 1307-1324
- Ellingsen, Tore, and Magnus Johannesson. 2007. Paying Respect. *Journal of Economic Perspectives*, 21(4): 135–49.
- Freeman, Richard B. and Alexander Gelber. 2010. Prize Structure and Information in Tournaments: Experimental Evidence. *American Economic Journal: Applied Economics*, 2(1), 149–164.
- Frey, Bruno S. 2007. Awards as Compensation. *European Management Review*, 4(1): 6–14.
- Green, Jerry R., and Nancy L. Stokey. 1983. A comparison of Tournaments and Contracts. *Journal of Political Economy*, 91(3), 349-364.
- Greenberg, J., 1988. Equity and workplace status: a field experiment. *Journal of Applied Psychology* 73, 606–613.
- Hannan, R. Lynn, Ranjani Krishnan, and Andrew H. Newman. 2008. The Effects of Disseminating Relative Performance Feedback in Tournament and Individual Performance Compensation Plans. *Accounting Review*, 83(4), 893-913.
- Kluger, Avraham and Angelo DeNisi. 1996. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*. 119(2), 254-284.

- Kosfeld, Michael and Susanne Neckermann. 2011. Getting More Work for Nothing? Symbolic Awards and Worker Performance. *American Economic Journal: Microeconomics*, 3, 86-99.
- Koszegi, Botond, 2006, Ego utility, overconfidence and task choice, *Journal of the European Economic Association* 4, 673–707.
- Kuhnen, Camelia M., Agnieszka Tymula. 2011. Feedback, Self-Esteem and Performance in Organizations. *Management Science*, forthcoming.
- Lazear, Edward P. and Sherwin Rosen (1981), "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89(5), 841-864.
- Lizzeri, A., M. Meyer, and N. Persico. 2002. The incentive effects of interim performance evaluations. *CARESS Working Paper* 02-09.
- Loewenstein, G. 1994. The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, 116(1), 75-98.
- Loewenstein, G. Moore, D. and R.W. Weber. 2006. Misperceiving the value of information in predicting the performance of others. *Experimental Economics*. 9(3), 281-295.
- Loewenstein, G., T. Brennan, and K.G. Volpp. 2007. Asymmetric paternalism to improve health behaviors. *Journal of the American Medical Association*. 298(20), 2415-2417.
- Moldovanu, Benny, Aner Sela, and Xianwen Shi. 2007. Contests for Status. *Journal of Political Economy*, 115(2): 338–63.
- Moldovanu, Benny, Aner Sela, Xianwen Shi (2007), "Contests for Status," *Journal of Political Economy*, 115(2), 338-363.
- Robson, Arthur J. 1992. Status, the distribution of wealth, private and social attitudes to risk. *Econometrica* 60, 837–857.
- Smither, James W., Manuel London, and Richard R. Reilly. 2005. Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology*, 58, 33–66.
- Thaler, R. H., C. R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth and Happiness*, Yale Univ. Press.
- 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. We can also confirm in this logistic estimation that the stated preference for feedback holds no predictive power for the revealed preference for feedback.

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 ²	.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]”