

Tournaments and Office Politics:  
Evidence From a Real Effort Experiment

by

Jeffrey Carpenter, Peter Hans Matthews and John Schirm

September 2007

MIDDLEBURY COLLEGE ECONOMICS DISCUSSION PAPER NO. 07-09



DEPARTMENT OF ECONOMICS  
MIDDLEBURY COLLEGE  
MIDDLEBURY, VERMONT 05753

<http://www.middlebury.edu/~econ>

# TOURNAMENTS AND OFFICE POLITICS: Evidence from a real effort experiment\*

Jeffrey Carpenter<sup>†</sup>    Peter Hans Matthews<sup>‡</sup>    John Schirm<sup>§</sup>

August 7, 2007

## Abstract

In many environments, tournaments can elicit more effort from workers, except perhaps when workers can sabotage each other. Because it is hard to separate effort, ability and output in many real workplace settings, the empirical evidence on the incentive effect of tournaments is thin. There is even less evidence on the impact of sabotage because real world acts of sabotage are often subtle manifestations of subjective peer evaluation or “office politics.” We discuss a real effort experiment in which effort, quality adjusted output and office politics are compared under piece rates and tournaments. Our results suggest that tournaments increase effort only in the absence of office politics. Competitors are more likely to sabotage each other in tournaments and, as a result, workers actually provide less effort simply because they expect to be the victims of sabotage. Adjusting output for quality with the rating of an independent auditor shrinks the incentive effect of the tournament even further since output tends to become more slipshod.

"The person who says 'I'm not political' is in great danger ... Only the fittest will survive, and the fittest will be the ones who understand their office's politics."

Jean Hollands, quoted in *Playing Office Politics*, *Newsweek*, 16 September 1985

## 1 Introduction

There is still much economists do not know about the incentive effects of tournaments, despite the widespread use of compensation schemes based on relative performance measures. The causes and consequences of sabotage, for example, are much better understood in principle

---

\*We thank the NSF (Career 0092953) and Middlebury College for funding. In addition, we thank the participants at the 2007 IZA Workshop on Behavioral and Organizational Economics for thoughtful comments.

<sup>†</sup>Department of Economics, Middlebury College & IZA; jpc@middlebury.edu

<sup>‡</sup>Department of Economics, Middlebury College; pmatthew@middlebury.edu.

<sup>§</sup>Compensation Department, Google; john.schirm@gmail.com.

(Lazear 1989, Chen 2003, Grund and Sliwka 2005) than in practice. To our knowledge, there are just two empirical studies of sabotage based on non-experimental data: Garicano and Palacios-Huerta (2006) find that when soccer teams have more incentive to win, they devote more effort both to scoring goals and to dirty play, with no net change in scoring and, in a more traditional vein, Drago and Garvey (1998) conclude on the basis of a survey of Australian manufacturing that when the incentives for promotion are sharp, workers expend less "helping effort."

Given Falk and Fehr's (2003) observation that experimental methods are especially well-suited to the study of tournaments, it comes as a surprise how few controlled studies of sabotage have been published. The principal contributors to this small literature have been Harbring and Irlenbusch (2003a, 2003b, 2005) and Falk and Fehr (2002), who have considered the effects of variations in the number of competitors, the number of prizes, the prize spread, the contract choices available to principals and communication on destructive activities.

Our paper extends this literature in at least two important directions. First, we explore the effects of sabotage within the context of a real effort tournament. While there is some debate about the differences between real and chosen effort designs (Bruggen and Strobel 2007, for example), we were concerned about the representativeness of chosen effort designs and share the concerns of van Dijk, Sonnemans and van Winden (2001, p.189), who remind us that real work "involves effort, fatigue, boredom, excitement and other affectations not present" in chosen effort.

Second, and no less important, our design reflects a different and, in some work environments, more plausible notion of sabotage. Sabotage in the lab is almost always *diffuse* and *blunt*. It is diffuse in the sense that, with the notable exception of Harbring, Irlenbusch, Krackel and Selten (2004), it is not directed at individuals: destructive activities are assumed to reduce the output of all other subjects. It is our impression, however, that in practice, the saboteur's aim is often much narrower, in part because diffuse punishment is more difficult for the target to interpret and, therefore, costlier to impose. It is blunt because what is represented in most, if not all, experiments is the physical destruction of output, one of the most extreme forms of sabotage. In reality a worker doesn't always need to produce more output than her rivals to win a promotion tournament, but rather create an impression, well-founded or otherwise, that she has. When individual effort levels are difficult to rank order, let alone measure, this is often a simpler, but more subtle, task, and one with *indirect* effects

on output: a worker who fears that her contributions to output will be misrepresented could well decide to expend less effort. Our experimental design allows for two forms of directed sabotage: subjects both counted, and evaluated the quality of, the output of each of their rivals and, in some treatments, compensation depended, in part, on these evaluations.

These peer evaluations are perhaps best understood as a metaphor for office politics, one purpose of which is to influence, at some cost, decision makers' beliefs about relative performance. To achieve this end, the "politician" in this experiment has two sorts of misinformation at her disposal, one more subtle, and perhaps easier to rationalize, than the other. As a theoretical matter, we know that "influence activities" (Milgrom and Roberts 1988) of this sort can even lead firms to abandon internal promotion in favor of external recruitment (Chan 1996).

A more literal, but we believe complementary, interpretation is that the protocol embodies what industrial psychologists and others would call "360° review" or "multi-source feedback" (MSF), since compensation depends on the evaluations of both supervisor and peers. Until recently, the consensus (McEvoy and Buller 1987, for example) was that most workers disliked MSF, and that this dislike was more pronounced when the results were used to determine compensation or otherwise evaluate performance. Furthermore, consistent with our results, resistance to peer evaluation is not limited to one's own assessment, but the effects of "friendship bias" (Kane and Lawler 1978) and other norms on the evaluations of others.

We find that sabotage or office politics more than reverse the incentive effects of a tournament: adjusted output per worker is less than that achieved under piece rates. Furthermore, the principal manifestation of this decrease is quality, not quantity. Expectations are then identified as the proximate cause of this decline: if workers expect that their contributions to the firm will be misrepresented, effort is scaled back. In addition, these expectations are warranted: as the differences between workers increase, so does sabotage. We also discern evidence of an affective response to tournaments, even in the absence of an incentive to sabotage.

The experimental design is described in the next section and in the section after that, we discuss our results in broad terms. The fourth section then reports estimates for output and sabotage functions.

## 2 Experimental Design

Instead of following in the rich tradition of “effort choice” experiments (e.g., Fehr et al., 1993 or Harbring and Irlenbusch, 2005) which are particularly adept at identifying the factors that influence the decisions to provide effort and sabotage, we decided to design a real effort experiment. In terms of the underlying effort task, our experiment is similar to Konow (2000) or Falk and Ichino (2006) in that our 160 participants were asked to spend 30 minutes stuffing envelopes.

In each of the 20 sessions (5 per treatment), 8 student participants were provided with their own computer, work table, “output box,” list of names and addresses and access to a shared printer. The task was to complete a form letter with names and addresses from the list, hand address an envelope, print the letter, stuff it into an envelope and then add it to the output box. The substance of the letter was not contrived: it concerned official department business that based on debriefings appeared to be salient to the students. The task was not as simple as first seems. From start to finish, each letter required between 60 and 90 seconds to complete.

After the production period, all of the participants went around the room and examined the output boxes of the other workers. The “supervisor,” one of the experimenters, also examined all of the output boxes. Each person counted and recorded the number of completed envelopes in each of the output boxes and then, on the basis of one envelope chosen at random from each box, estimated the quality of production, on a scale from 0 to 1. Because an objective measure of quality, or at least one in which neither the experimenters nor the participants had a vested interest, was later needed to determine levels of sabotage, we hired a letter carrier from the US Postal Service to count and evaluate the “deliver-ability” of all the envelopes.

Finally, at the end of each session, subjects completed a short survey that allowed us to collect much of the usual demographic data, some information about expectations, in particular whether each subject expected his or her own output to be reported accurately, and a measure of risk preferences.

The students participated in one of four treatments that differ with respect to the method of compensation and the opportunities for sabotage. In the baseline *Piece Rate* treatment, participants were paid \$1 for each quality adjusted envelope produced. In this case quality adjusted output depended only on the count and assessment of the supervisor. Where the

supervisor's count of participant  $i$  is  $N_{S \rightarrow i}$  and his quality assessment is  $Q_{S \rightarrow i}$ , the dollar payoff of worker  $i$  in the Piece Rate treatment was:

$$\pi_i^{PR} = N_{S \rightarrow i} \times Q_{S \rightarrow i}$$

In an effort to maintain the internal validity of the experiment and link the change in the design to the incentive effects of a tournament, the compensation scheme in the *Tournament* treatment is nearly identical to the one in the Piece Rate treatment:

$$\pi_i^T = \left\{ \begin{array}{ll} 25 + (N_{S \rightarrow i} \times Q_{S \rightarrow i}) & \text{if } N_{S \rightarrow i} \times Q_{S \rightarrow i} > N_{S \rightarrow j} \times Q_{S \rightarrow j} \quad \forall j \neq i \\ N_{S \rightarrow i} \times Q_{S \rightarrow i} & \text{otherwise} \end{array} \right\}$$

As one can see the only difference was that the highest producer earned a bonus of \$25 for winning the tournament which means that any changes in behavior can be attributed to the bonus.<sup>1</sup>

We take advantage of the peer assessments in the *Tournament with Sabotage* treatment. In this condition, workers are compensated based on the average quality adjusted output assessed by all 8 auditors. Now, one's pay and chances of winning the tournament depend on the evaluation of the supervisor as in the previous two treatments but one's peers can also affect your final output assessment. The average assessed output of worker  $i$  is:

$$\overline{N_{j \rightarrow i} \times Q_{j \rightarrow i}} = \left( \sum_{j \neq i} N_{j \rightarrow i} \times Q_{j \rightarrow i} \right) / 8$$

and, as a result, her payoff in the Tournament with Sabotage treatment is:

$$\pi_i^{T\&S} = \left\{ \begin{array}{ll} 25 + \overline{(N_{j \rightarrow i} \times Q_{j \rightarrow i})} & \text{if } \overline{N_{j \rightarrow i} \times Q_{j \rightarrow i}} > \overline{N_{j \rightarrow k} \times Q_{j \rightarrow k}} \quad \forall k \neq i \\ \overline{N_{j \rightarrow i} \times Q_{j \rightarrow i}} & \text{otherwise} \end{array} \right\}$$

Lastly, to be complete we also ran a *Piece Rate with Sabotage* treatment that was identical to the Tournament with Sabotage treatment except no bonus was paid to the producer of the highest quality adjusted output. While it is clear that to assess the pure incentive effect of the tournament in the presence of the possibility for sabotage, one needs to compare

---

<sup>1</sup>Clearly the expected payoffs will be different between the Piece Rate and Tournament treatments unless all participants place zero weight on winning the contest. An alternative would have been to try to maintain the same expected payoff in the two treatments, but given ability is unobserved in this real effort environment (although it should be randomly distributed across treatments), it would have been impossible to calibrate such an alternative design.

the Tournament with Sabotage treatment to the Piece Rate with Sabotage treatment, it would be odd if there were differences between the two piece rate treatments because there is absolutely no incentive to sabotage each other when everyone is paid a piece rate. Indeed, we find no statistically significant differences between the two piece rate treatments and therefore we pool these data for the purposes of our analysis.

Based on our description of how the four treatments were run, it should be obvious that there are two avenues through which office politics and sabotage might affect outcomes in the experiment. Because winning a promotion tournament often has as much to do with the perception of being the most productive as it does with actually being the most productive, overt forms of sabotage, like the destruction of one another's output, sometimes give way to more subtle forms. These acts frequently take the form of rumors or comments aimed at diminishing the perceived ability or accomplishments of one's competitors. To some degree, our experiment is designed to capture both forms of office politics.

In our design, *brazen* sabotage occurs when people undercount each other's output. From a strategic point of view, it should be obvious that each worker has the incentive to report zero units produced for each of the 7 other competitors in a tournament. This is how one maximizes the chance of winning the tournament. That said we suspected that only people with extremely little integrity would engage in such obvious acts of sabotage.

To provide an environment more conducive to office politics, we purposely had the participants hand write the addresses on all of the envelopes so that there would be both objective, and possibly subjective, differences in the assessed quality of the output. This possibility of subjective peer assessment is at the very heart of office politics. Just as one has the incentive to undercount the output produced by one's competitors, one also has the incentive to underrate the quality of their output. In fact, if counting zero for everyone else dominates, so does assigning them quality equal to zero. We posited that because quality is so much more subjective than count, saboteurs might feel more comfortable using it to lower the perceived productivity of their peers.

### **3 Descriptive Statistics**

Before conducting controlled tests for differences in our treatments, we begin by describing our participants and the broad patterns we see in output and sabotage. Table 1 summarizes the characteristics of our participants based on data from the exit survey. In terms of the

standard demographics, our participants were roughly evenly split between men and women, 11% were international students, the mean Grade Point Average (GPA) of the participants was high, and as is typical in the student population, there was not much variation in age. The two demographics that we were particularly interested in are sex and whether or not one is an international student because it is easy to formulate hypotheses about these traits being predictors of ability in this task. Women tend to have better dexterity and hand writing and international students will be less familiar with English and the U.S. postal system. To control to some extent for the competitiveness of our participants we collected data on birth order and the number of siblings one has (Falbo, 1981).

Based on past experience we have little confidence in the self reports of students about personal or family income and wealth. We therefore asked indirect questions that may be imperfect proxies for income and wealth but are probably measured with less error and bias. We asked them for the number of bathrooms in their parent’s house, whether they had a car on campus and whether they were employed during the school year. Precise measurements of the marginal effects of income and wealth are well beyond the purview of the current experiment – we simply want to control for differences.

We collected two other control variables in the survey. To account for the risk preferences of our participants we had the participants complete the Weber et al (2002) risk preference scale and created an indicator variable for those persons who ranked among the top 10% in risk taking. We also asked each participant to tell us the number of other participants in the session that they knew. We considered this to be a control for social distance, and the possibility of collusion, within a session. On average, people knew one or two of the other participants.

Lastly, we were interested in the extent to which participants would anticipate sabotage and how this expectation would affect their performance in the production task. We asked, “Did you expect your teammates to correctly report your production output?” Overall, most (88%) of people expected the other participants in their session to correctly assess their output. As we will see however, this expectation is less “rosy” for participants in the Tournament with Sabotage treatment. Here only 60% thought their output would be counted correctly). We will also see that these expectations matter a lot.

Three important aspects of our data are depicted in Figure 1 which reports the mean postal worker count of raw output (black bars), the mean quality adjusted count of the postal



worker (dark grey bars) and the mean peer adjusted count (light grey bars) for the three treatments.<sup>2</sup> Perhaps the most important thing to notice is the disincentive effect of the threat of sabotage. As reported elsewhere (e.g., Nalbantian and Schotter, 1997 or van Dijk et al., 2001), adding the \$25 bonus for the highest producer increases raw effort (proxied here by the simple postal worker envelope count – black bars). The number of envelopes produced increases by 1.125 on average ( $z = 1.84, p = 0.06$ ) when we move from piece rates to the tournament.<sup>3</sup> The difference of 1.125 envelopes does not seem large, but remember that our participants only produced for 30 minutes. Over the course of an 8 hour day, the 8 workers would produce 144 more envelopes in the Tournament condition. However, what is important is that raw output actually falls compared to both the Tournament ( $z = 3.36, p < 0.01$ ) and the Piece Rate treatment ( $z = 2.03, p = 0.04$ ) when competitors are able to sabotage each other. It is one thing to discover the presence of sabotage and the lowering of output as the direct result of sabotage but to find that just the potential for sabotage acts as a large disincentive to providing effort is unique to this experiment. On average, workers produce 2.475 fewer envelopes when sabotage can alter the course of a tournament.

If one compares the black raw output bar to the dark grey quality adjusted output bar within each treatment one gets a sense of the extent to which quality varied by treatment. One might expect quality to be lower in the tournaments because people feel more pressure to rush but because they will be paid based on the number of quality adjusted envelopes, it is risky to produce slipshod output. The 2.53 quality adjusted envelope difference in the Piece Rate treatment is similar to the 2.48 difference in the Tournament but the difference is slightly larger in the Tournament with Sabotage treatment, 2.67 envelopes. If one regresses quality on treatment indicators only, the small differences appear to be significant. The Tournament with Sabotage treatment yields lower quality of 0.06 compared to the Piece Rate treatment ( $p < 0.10$ ) and 0.07 less than the Tournament ( $p < 0.05$ ).<sup>4</sup>

Figure 1 also suggests that the experiment was successful in eliciting differences in sabotage that make sense. Recall that there is no monetary incentive to sabotage one’s peers in either the Piece Rate or Tournament treatments, but nonetheless, we allowed participants to count and rate each other to provide important benchmarks. In the Piece Rate treatment it

---

<sup>2</sup>Note: the peer adjusted count is the mean assessment of the 7 other participants only and recall that there are no statistically significant differences between the two piece rate treatments so they have been pooled to simplify and sharpen the analysis.

<sup>3</sup>We report  $z$  statistics from the nonparametric Wilcoxon test.

<sup>4</sup>Recall that quality is measured on  $[0,1]$  scale.

appears as if people actually made small gifts to each other, on average. One can see this by comparing the dark grey quality adjusted count of the postal worker to the light grey peer adjusted count.<sup>5</sup> When nothing is at stake, there is no harm in being nice to one's peers.

The same is not true in the Tournament. Here we find that people sabotage each other even when there is no material incentive to do so. The difference of almost one quality adjusted envelope is just significant ( $z = 1.66$ ,  $p = 0.09$ ) and can be symptomatic of a simple, maybe even unconscious or affective/emotional response to competition (Schultheiss et al., 1999). When given the material incentive in the Tournament with Sabotage treatment, one can see much lower peer adjusted output. Compared to the evaluation of the letter carrier, peers credit each other with producing an average of about two adjusted envelopes less than the letter carrier and the difference is highly significant ( $z = 2.81$ ,  $p < 0.01$ ). As a rough comparison of magnitudes, if the effects of the bonus and the ability to reduce the output of others are additive, then it appears that affect accounts for 47% of the overall sabotage that occurs in the Tournament with Sabotage treatment.<sup>6</sup> The remaining 53% of sabotage is likely to be cognitive and strategy-driven.

## 4 Econometric Results

Our first and most important results concern the relationship between compensation and adjusted output, and these are contained in Table 2. The first column reports least squares estimates, with robust standard errors, for the regression of adjusted individual output on the two tournament treatment indicators. Both coefficients are significant at the 10 percent level. Relative to the combined Piece Rate treatment, adjusted output is estimated to rise 1.171 per person, or  $9.368 = 8(1.171)$  per team, in the Tournament. When sabotage is possible, on the other hand, adjusted output per worker falls not just relative to the tournament ( $t = 2.87$ ,  $p = 0.01$ ), but to the combined Piece Rate treatments.

The addition of indicators for sex, international student status and risk preferences has little effect on these treatment effects: the estimated coefficients retain both their size and significance. Also, the estimated coefficients for sex and international student status are negative and significant in both the statistical and economic senses, and are consistent, we believe, with the nature of the task, one in which keyboard dexterity, neat handwriting and

---

<sup>5</sup>However, the difference in the two means is not significant ( $z = 1.03$ ,  $p = 0.30$ ).

<sup>6</sup>0.95 of the 2.03 difference between postal worker assessment and peer assessment in the Tournament with Sabotage treatment.

fluency in English were favored.

The observation that risk takers are estimated to produce 1.781 fewer envelopes, and that the difference is significant at the 10 percent level, calls for a more subtle explanation. It is our hypothesis that risk takers are more likely than other subjects to skimp on quality in an effort to finish more envelopes, a conjecture with important implications for personnel management. We find support for this view in Tables A1 and A2 in the Appendix, which report the results for separate (raw) quantity and quality regressions. Combined, the second and third columns of Tables A1 and A2 suggest that risk takers produce about the same number of envelopes but spend less effort on each. If the strong quality effect comes as no surprise, the insignificant quantity effect is unexpected: our prior was that a risk taker would attempt to produce more, but inferior, envelopes.

The last columns in Table 2 reveal that the addition of the expectations variable, whether or not subjects anticipate that their output will be reported correctly, has two marked effects, with or without other demographic controls. First, the coefficient for risk takers shrinks in both size and significance.

Second and most importantly, the treatment effect for tournaments with sabotage vanishes. It seems reasonable to conclude, then, that quality adjusted output falls in this environment *because* fewer individuals expect their co-workers to be truthful about their contributions to production and, as a result, effort is reduced. In more evocative terms, firms in which promotion is the result of office politics will become less profitable, not because output is ever lost or ruined, but because it is never produced.<sup>7</sup>

But were such expectations reasonable? That is, did our subjects sabotage one another more often in the tournament? And what other influences can be identified? Figure 2 illustrates the incidence and amount of output sabotage in the three treatments. As one can see almost 80% of the evaluations were correct in the Piece Rate treatment but this number fall to only less than 60% in the Tournament and further still to less than 50% in the Tournament with Sabotage. Furthermore, there are a lot of observations considerably to the right of zero in the Tournament with Sabotage indicating significant sabotage. It is one thing to mis-count by one or two envelopes as occurs in the Piece Rate treatment but there are 22 instances of the target's output being evaluated as less than five units and 9 instances in which the evaluator reduced the target's output to zero. All of these instances

---

<sup>7</sup>The results reported in tables 2, A1 and A2 are all robust to clustering standard errors at the session level or using the Tobit regressor (because there are so few censored observations).

occurred in the Tournament with Sabotage.

For a more systematic analysis of output sabotage, Table 3 reports GLS estimates with individual random effects ( $v_{ij}$ ) for three models of output sabotage, all based on the spline function:

$$\begin{aligned}
 OS_{ij} = & \beta_0 + \beta_1 T + \beta_2 TS + \beta_3(Q_j - Q_i) + \beta_4(Q_j - Q_i)T + \beta_5(Q_j - Q_i)TS \\
 & + \beta_6 \max\{Q_j - Q_i, 0\} + \beta_7 \max\{Q_j - Q_i, 0\}T + \beta_8 \max\{Q_j - Q_i, 0\}TS \\
 & + \vec{X}\theta + u_i + v_{ij}
 \end{aligned}$$

where  $OS_{ij}$  is  $i$ 's undercount (positive) or overcount (negative) of  $j$ 's output,  $T$  and  $TS$  are the tournament treatment indicators,  $Q_j - Q_i$  is the objective difference in output and  $\vec{X}$  is a vector of controls. In the first specification, for example, no controls are included, and the only estimated coefficient that is significant at the 10 percent level or better is that on  $\max\{Q_j - Q_i, 0\}TS$ . That is, it is only in Tournaments with Sabotage, and then only when the target has produced more output, that the model predicts systematic miscounting of any kind. As the implied sabotage functions in Figure 3 reveal, however, those who produce more output should expect substantial undercounts: in particular, a four envelope difference will be reported as three, a surprise inasmuch as there are other, more subtle, forms of sabotage available.

The estimates in the second and third columns hint that the result is a robust one. No matter what other controls are included, neither the size nor the significance of the relevant coefficients are affected. Furthermore, results not reported here reveal that the results are also robust with respect to the specification of the error term: the same patterns prevail when errors are instead clustered on the individual. The data also demonstrate that, once more, expectations matter. Those who expect co-workers to report their output correctly will sabotage them less, an example of complex or expectational reciprocity: if A expects B to report A's output correctly, then A will correctly report B's, and so on. Inasmuch as one of the controls included in the third specification is the number of other subjects known, this is more than a "friendship effect."

Given the differences between the two forms of office politics, we expected false reports of quality, in either direction, to be more common and more responsive to output differences. Figure 4 shows that there is much more variation in the quality sabotage data reflecting the partially subjective nature of this sort of assessment. As in Figure 2, however, we see that the number of cases in which the peer evaluator agreed with the letter carrier decreases

dramatically from the Piece Rate to the Tournament and even further to the Tournament with Sabotage. While the distribution of evaluation differences is fairly symmetric around zero in the Piece rate treatment, it is skewed slightly to the right in the Tournament and skewed dramatically to the right in the Tournament with Sabotage. Indeed, as expected, there is a lot of quality sabotage in the Tournament with Sabotage and there is even a significant amount in the Tournament despite there being no incentive to sabotage one's coworkers.

Table 4 reports estimates for three models that, apart from the dependent variable, are identical to those used to characterize quantity sabotage:

$$\begin{aligned}
QS_{ij} = & \gamma_0 + \gamma_1 T + \gamma_2 TS + \gamma_3(Q_j - Q_i) + \gamma_4(Q_j - Q_i)T + \gamma_5(Q_j - Q_i)TS \\
& + \gamma_6 \max\{Q_j - Q_i, 0\} + \gamma_7 \max\{Q_j - Q_i, 0\}T + \gamma_8 \max\{Q_j - Q_i, 0\}TS \\
& + \vec{X}\Lambda + e_i + \epsilon_{ij}
\end{aligned}$$

where  $QS_{ij}$  is now  $i$ 's underestimate (positive) or overestimate (negative) of the quality of  $j$ 's output. The most immediate difference between Tables 3 and 4 is that almost all of the coefficients  $\gamma_0, \dots, \gamma_8$  are significant, consistent with our prior.

To understand better their implications, consider the implied sabotage functions for the benchmark specification in which no other terms are included, as depicted in Figure 5. Even when there is no difference in output, subjects engage in significant positive sabotage in *both* tournament treatments. Underestimation is more pronounced when it matters most, in tournaments with sabotage, but even in the standard tournament, it exists as what we believe to be an affective response. The further observation that the quality sabotage function more or less levels off for positive output differences in the standard tournament lends credence to this view. The response to co-workers who produce more is not "calibrated" in any obvious sense.

In tournaments with sabotage, on the other hand, underestimation always becomes more pronounced with differences in output, even if the small, but statistically significant, decrease in the slope of the function to the right of the "knot" is difficult to rationalize.

Figure 5 also illustrates that the "sabotage threshold" - that is, the difference in output sufficient to induce underestimation - is not just negative, but almost equal, in the two sorts of tournaments, and that below this, the amount of "negative sabotage" (i.e., gifts) increases more slowly in the standard tournament.

All of this stands in marked contrast to behavior in the combined piece rate treatment, in which the extent of sabotage is predicted to *decrease*, albeit slowly, as output differences increase, and the level of sabotage associated with no difference in output is negative. For that matter and consistent with Figure 1, at most values of the difference in output, participants in the Piece Rate condition actually make small quality gifts to their coworkers. Furthermore, because the estimated change in the slope of the sabotage function at the knot is statistically insignificant, the rate at which sabotage declines is independent of whether the difference in output is positive or negative.

Once more, we are confident that our results are robust. On the one hand, the results are not much different when the errors are clustered at the individual level. More important, as the second and third columns in Table 4 attest, the addition of the now familiar controls for sex, international student status, risk preferences, expectations and so on has little effect, whether or not other controls are also used. There is perhaps less evidence of reciprocity, however: the coefficient on expectations has the right (that is, positive) sign, but is no longer significant.

## 5 Concluding Remarks

In lieu of the usual rehash of motives and results, let us instead focus on some of their implications. On a practical level, the obvious question is whether managers should foster competition among coworkers or explicitly set up promotion tournaments? One clear answer is that if office politics can have an effect on output, either directly or indirectly in terms of lost productivity due to political maneuvering, then between worker competition should be avoided. Workers have good reason to be wary of the sorts of peer review mentioned in the introduction. Our results suggest that when there is any ambiguity in the performance of a competitor, workers are likely to engage in sabotage. Not only will time be spent on unproductive tasks, the atmosphere itself created by a tournament can be a disincentive to work hard. The second effect is the main empirical contribution of our study.

It is also not clear whether one should use tournaments even when office politics can be kept to a minimum. Recall the estimates of the effect of the tournament on adjusted output in Table 2. If the bonus is substantial (i.e., the difference in pay grades is large from one level to the next in the context of Lazear, 1989 or Drago and Garvey, 1998), it is not clear that the increased productivity provided by the competition outweighs the added

labor costs. The per worker increase when comparing piece rates to the tournament is 1.171 units. This means an increase of  $1.171 \times 8 = 9.368$  units for the team. Since labor costs in the tournament are therefore  $34.368 = 25 + 9.368$  more, the tournament is only worth it if  $9.368\rho \geq 34.368$  or  $\rho \geq 3.57$ , where  $\rho$  is the sales price per finished envelope.

## 6 References

Bruggen, Alexander and Strobel, Martin, 2007. Real effort versus chosen effort in experiments. *Economics Letters*, forthcoming.

Chan, William, 1996. External recruitment versus internal promotion. *Journal of Labor Economics*, 14, 555-570.

Chen, Kong-Pin, 2003. Sabotage in Promotion Tournaments. *Journal of Law, Economics, and Organization*, 19(1), 119-140.

Drago, Robert and Garvey, Gerald T., 1998. Incentives for Helping on the Job: Theory and Evidence. *Journal of Labor Economics*, 16(1), 1-25.

Falbo, Toni U. Texas Austin, 1981. Relationships between birth category, achievement, and interpersonal orientation. *Journal of Personality and Social Psychology* Vol 41(1), 121-131 Additional Info US American Psychological Association; Other Source Info *Journal of Abnormal & Social Psychology*.

Falk, Armin and Fehr, Ernst, 2002, The power and limits of tournament incentives, University of Zurich working paper.

Falk, Armin and Fehr, Ernst, 2003. Why labor market experiments? *Labour Economics*, 10(4), 399-406.

Falk, Armin and Ichino, Andrea, 2006. Clean evidence on peer pressure. *Journal of Labor Economics*, 24(1), 39-57.

Fehr, Ernst, Kirchsteiger, Georg and Riedl, Arno, 1993. Does Fairness Prevent Market Clearing? An Experimental Investigation. *Quarterly Journal of Economics*, 108, 437-459.

Garicano, Luis and Palacios-Huerta, Ignacio, 2006, Sabotage in tournaments: making the beautiful game a little less beautiful, mimeo.

Grund, Christian and Sliwka, Dirk, 2005. Envy and compassion in tournaments. *Journal of Economics and Management Strategy*, 14, 187-207.

Harbring, Christine and Irlenbusch, Bernd, 2003a. How many winners are good to have? On tournaments with sabotage. *Journal of Economic Behavior & Organization*, forthcoming.

Harbring, Christine and Irlenbusch, Bernd, 2003b, Sabotage and the endogenous design of tournaments, University of Bonn Working Paper.

Harbring, Christine and Irlenbusch, Bernd, 2005. Incentives in Tournaments with Endogenous Prize Selection. *Journal of Institutional and Theoretical Economics*, 161(4), 636-663.

Harbring, Christine, Irlenbusch, Bernd, Krackel, Matthias and Selten, Reinhard, 2004. Sabotage in asymmetric contests - an experimental analysis. *International Journal of the Economics of Business*, forthcoming.

Kane, J.S. and Lawler, E.E., 1978. Methods of peer assessment. *Psychological Bulletin*, 85, 555-586.

Konow, James, 2000. Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *American Economic Review*, 90(4), 1072-1091.

Lazear, Edward P., 1989. Pay Equality and Industrial Politics. *Journal of Political Economy*, 97(3), 561-580.

McEvoy, Glenn and Buller, Paul, 1987. User acceptance of peer appraisals in an industrial setting. *Personnel Psychology*, 40, 785-797.

Milgrom, Paul and Roberts, John, 1988. An economic approach to influence activities in organizations. *The American Journal of Sociology*, 94, S154-S179.

Nalbantian, Haig and Schotter, Andrew, 1997. Productivity Under Group Incentives: an experimental study. *American Economic Review*, 87(3), 314-341.

Schultheiss, Oliver, Campbell, Kenneth and McClelland, David, 1999. Implicit power motivation moderates men's testosterone response to imagined and real dominance success. *Hormones and Behavior*, 36, 234-241.

van Dijk, Frans, Sonnemans, Joep and van Winden, Frans, 2001. Incentive Systems in a Real Effort Experiment. *European Economic Review*, 45, 187-214.

Weber, Elke, Blais, Ann-Renee and Betz, Nancy, 2002. A domain-specific risk attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15, 263-290.



## 7 Tables and Figures

TABLE 1: Participant Characteristics

	N	Mean	Std. Dev.	Min	Max
Male	160	0.463	0.500	0	1
International Student	160	0.113	0.317	0	1
Risk Scale	160	128.600	25.176	0	203
Risk Taker (90th percentile Risk Scale)	160	0.100	0.301	0	1
E(Teammates to correctly report my output)	160	0.888	0.317	0	1
Age	160	20.019	1.385	18	23
GPA	158	3.486	0.281	2.55	4
First Born	160	0.581	0.495	0	1
Number of Siblings	160	1.569	1.164	0	7
Number of Bathrooms in Parent's House	160	3.066	1.528	0	9
Have a Car on Campus	160	0.406	0.493	0	1
Employed	160	2.225	0.997	0	4
Number of Participants Known	159	1.277	1.272	0	8

TABLE 2: Analysis of Objectively Adjusted Output

	(1)	(2)	(3)	(4)
Tournament	1.171*	1.138**	1.165**	1.370**
	(0.59)	(0.54)	(0.55)	(0.55)
Tournament with Sabotage	-1.494*	-1.680**	0.188	0.347
	(0.77)	(0.72)	(0.79)	(0.76)
Male		-1.265**	-1.220**	-1.300**
		(0.52)	(0.49)	(0.51)
International Student		-3.125***	-3.239***	-3.561***
		(0.72)	(0.65)	(0.75)
Risk Taker		-1.781*	-0.624	-0.447
		(0.93)	(0.97)	(1.01)
E(Teammates to correctly report my output)			4.746***	4.515***
			(1.08)	(1.16)
Intercept	11.121***	12.291***	7.481***	15.989***
	(0.36)	(0.45)	(1.23)	(5.42)
Includes demographic controls	No	No	No	Yes
R-squared	0.07	0.20	0.32	0.35
N	160	160	160	157

Note: OLS with robust standard errors; the omitted treatment is piece rate; column (4) includes controls for age, GPA, birth order, number of siblings, employment status, the number of other participants known and proxies for family wealth.

TABLE 3: Analysis of Output Sabotage

	(1)	(2)	(3)
Tournament	0.130 (0.34)	0.095 (0.34)	0.084 (0.35)
Tournament with Sabotage	-0.025 (0.33)	-0.373 (0.36)	-0.306 (0.38)
(Target's Output - Saboteur's Output)	0.046 (0.05)	0.052 (0.05)	0.042 (0.05)
(Target's Output - Saboteur's Output) × Tournament	0.131 (0.09)	0.119 (0.09)	0.127 (0.09)
(Target's Output - Saboteur's Output) × Tournament with Sabotage	0.037 (0.08)	0.016 (0.08)	0.035 (0.08)
max {(Target's Output - Saboteur's Output), 0}	0.061 (0.07)	0.058 (0.07)	0.078 (0.08)
max {(Target's Output - Saboteur's Output), 0} × Tournament	-0.131 (0.14)	-0.123 (0.14)	-0.124 (0.14)
max {(Target's Output - Saboteur's Output), 0} × Tournament with Sabotage	0.260** (0.12)	0.263** (0.12)	0.229* (0.12)
Male		0.185 (0.22)	0.219 (0.23)
International Student		-0.372 (0.34)	-0.405 (0.38)
Risk Taker		0.194 (0.38)	0.228 (0.40)
E(Teammates to correctly report my output)		-0.905** (0.42)	-0.829* (0.44)
Intercept	0.120 (0.20)	0.957** (0.48)	-2.450 (2.27)
Includes demographic controls	No	No	Yes
rho	0.28	0.26	0.27
Wald Chi-squared	109	119	123
N	1093	1093	1072

Note: GLS with individual random effects; the omitted treatment is piece rate; column (3) includes controls for age, GPA, birth order, number of siblings, employment status, the number of other participants known and proxies for family wealth.

TABLE 4: Analysis of Quality Sabotage

	(1)	(2)	(3)
Tournament	0.147*** (0.04)	0.146*** (0.04)	0.123*** (0.04)
Tournament with Sabotage	0.232*** (0.04)	0.203*** (0.05)	0.222*** (0.05)
(Target's Output - Saboteur's Output)	-0.011** (0.01)	-0.011** (0.01)	-0.012** (0.01)
(Target's Output - Saboteur's Output)×Tournament	0.033*** (0.01)	0.032*** (0.01)	0.033*** (0.01)
(Target's Output - Saboteur's Output)×Tournament with Sabotage	0.044*** (0.01)	0.044*** (0.01)	0.045*** (0.01)
max {(Target's Output - Saboteur's Output), 0}	0.005 (0.01)	0.005 (0.01)	0.004 (0.01)
max {(Target's Output - Saboteur's Output), 0}×Tournament	-0.027* (0.01)	-0.027* (0.01)	-0.025* (0.02)
max {(Target's Output - Saboteur's Output), 0}×Tournament with Sabotage	-0.025** (0.01)	-0.025** (0.01)	-0.025** (0.01)
Male		-0.003 (0.03)	-0.009 (0.03)
International Student		0.008 (0.05)	0.015 (0.05)
Risk Taker		0.005 (0.05)	0.002 (0.06)
E(Teammates to correctly report my output)		-0.075 (0.06)	-0.049 (0.06)
Intercept	-0.044* (0.02)	0.031 (0.07)	-0.135 (0.31)
Includes demographic controls	No	No	Yes
rho	0.44	0.44	0.45
Wald Chi-squared	93	95	100
N	1093	1093	1072

Note: GLS with individual random effects; the omitted treatment is piece rate; column (3) includes controls for age, GPA, birth order, number of siblings, employment status, the number of other participants known and proxies for family wealth.

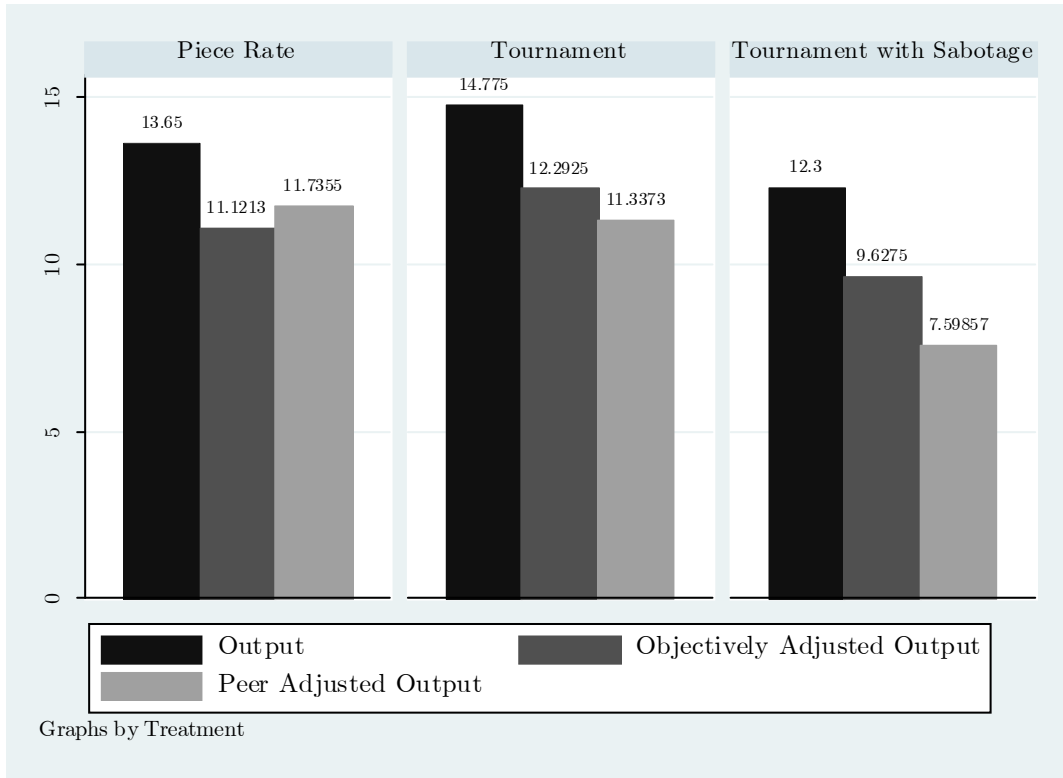


FIGURE 1: Mean Production Levels by Treatment

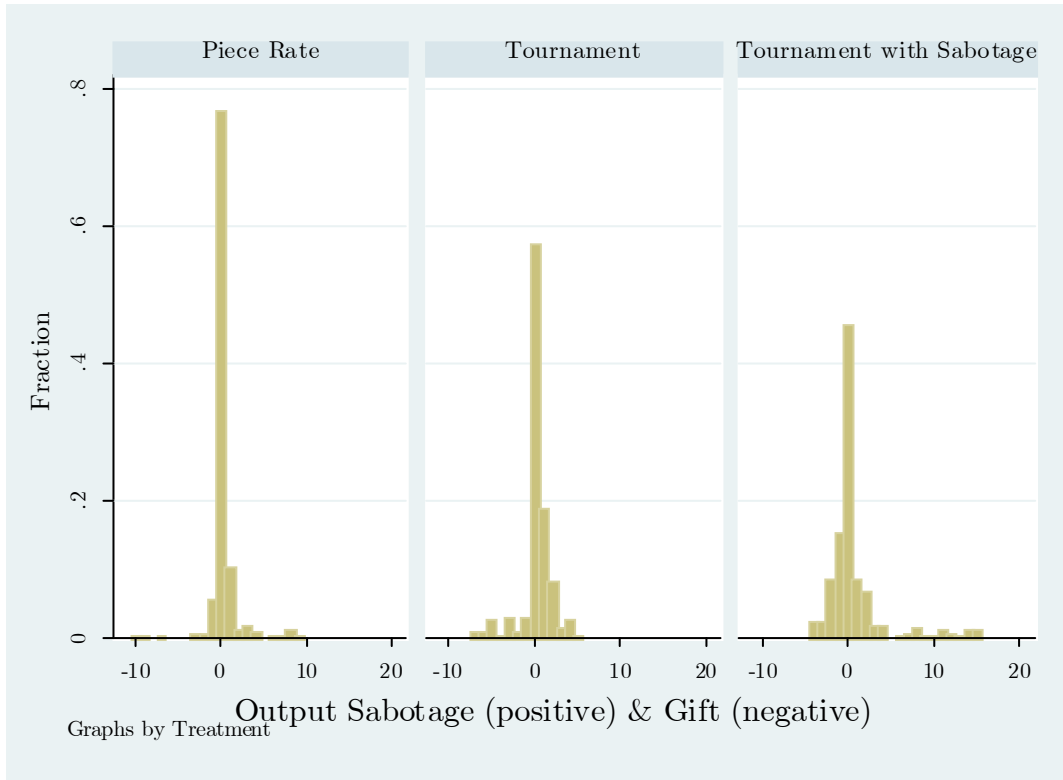


FIGURE 2: Output Sabotage by Treatment (Note: two observation in the Tournament with Sabotage treatment have been deleted  $\{-20, 24\}$  to make the graph easier to read).

### Output Sabotage

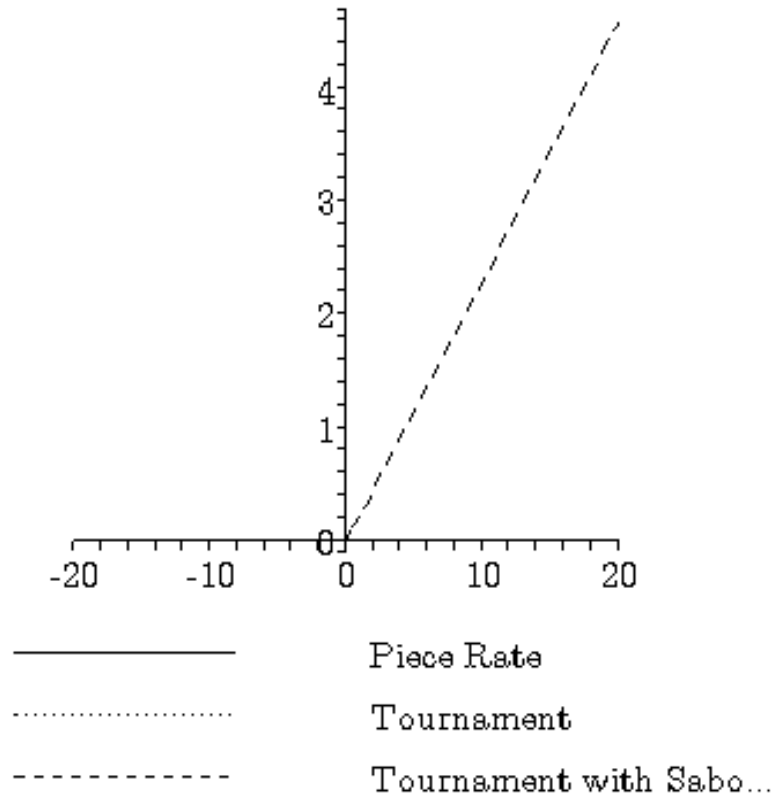


FIGURE 3: Estimated Output Sabotage Functions by Treatment (Note: the horizontal axis is the difference between the target's and the saboteur's raw output).

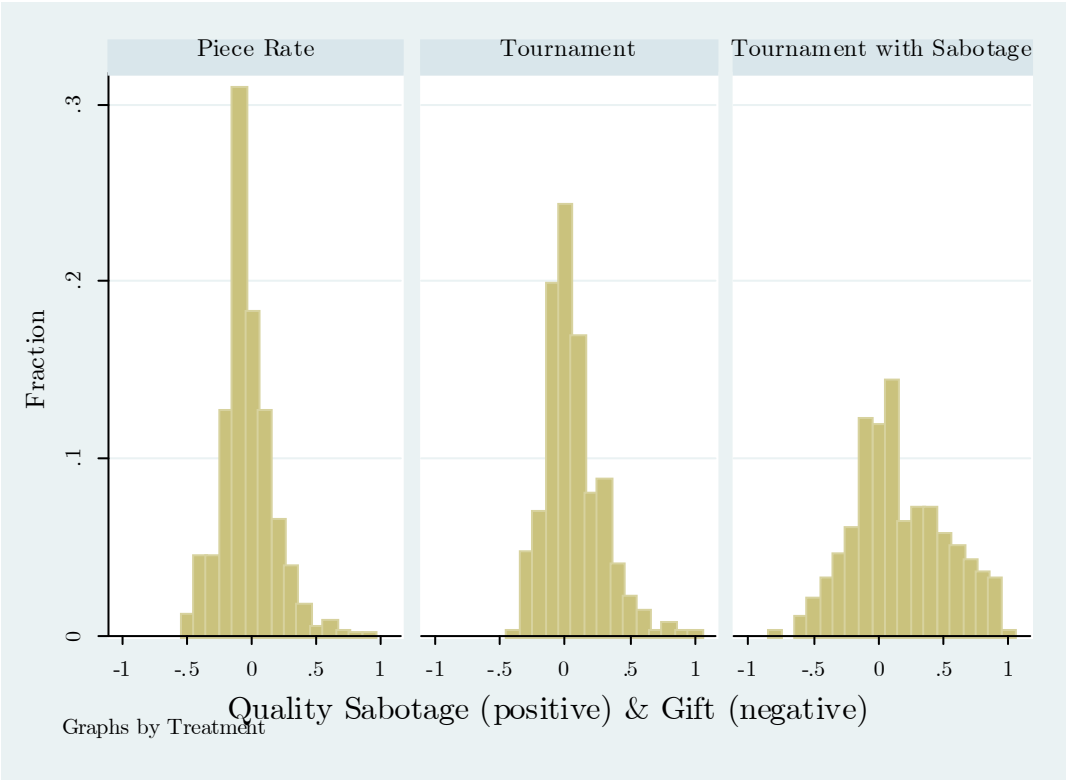


FIGURE 4: Quality Sabotage by Treatment



### Quality Sabotage

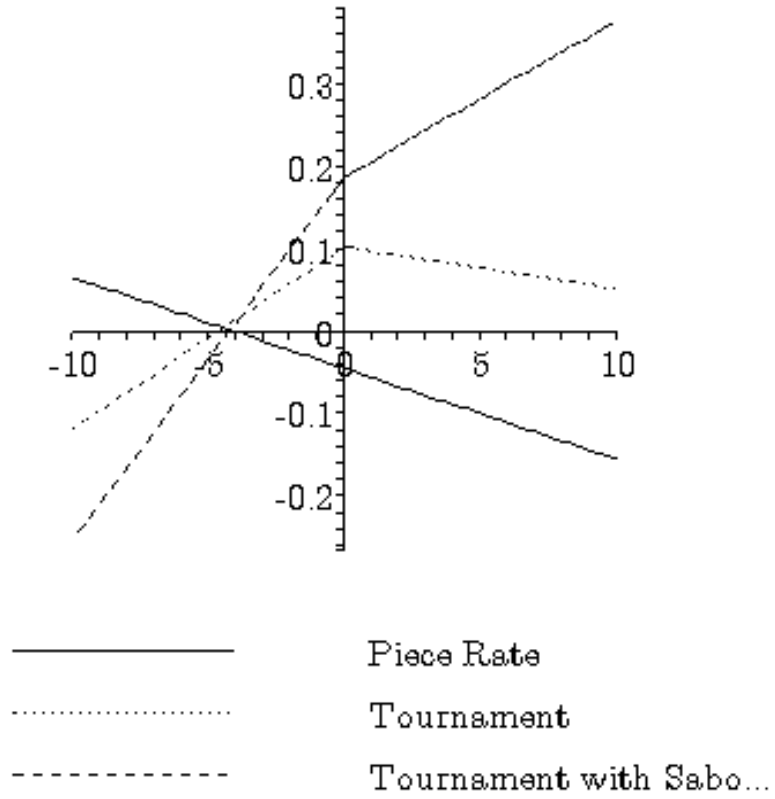


FIGURE 5: Estimated Quality Sabotage Functions by Treatment (Note: the horizontal axis is the difference between the target's and the saboteur's raw output).

## 8 Appendix: Supplemental Regression Tables and Experimental Instructions

TABLE A1: Analysis of Objectively Assessed Output

	(1)	(2)	(3)	(4)
Tournament	1.125*	1.062*	1.075*	1.316**
	(0.58)	(0.56)	(0.57)	(0.59)
Tournament with Sabotage	-1.350*	-1.445**	-0.573	-0.437
	(0.72)	(0.70)	(0.86)	(0.85)
Male		-0.722	-0.701	-0.805
		(0.53)	(0.53)	(0.55)
International Student		-2.439***	-2.492***	-2.670***
		(0.65)	(0.65)	(0.77)
Risk Taker		0.084	0.624	0.716
		(0.94)	(0.97)	(1.06)
E(Teammates to correctly report my output)			2.217*	2.124
			(1.26)	(1.35)
Intercept	13.650***	14.290***	12.043***	16.998***
	(0.39)	(0.50)	(1.44)	(5.84)
Includes demographic controls	No	No	No	Yes
R-squared	0.06	0.12	0.15	0.17
N	160	160	160	157

Note: OLS with robust standard errors; the omitted treatment is piece rate; column (4) includes controls for age, GPA, birth order, number of siblings, employment status, the number of other participants known and proxies for family wealth.

TABLE A2: Analysis of Objectively Assessed Quality

	(1)	(2)	(3)	(4)
Tournament	0.01 (0.02)	0.011 (0.02)	0.013 (0.02)	0.011 (0.02)
Tournament with Sabotage	-0.060* (0.04)	-0.068** (0.03)	0.042* (0.02)	0.048** (0.02)
Male		-0.048** (0.02)	-0.046*** (0.02)	-0.048** (0.02)
International Student		-0.099** (0.05)	-0.106*** (0.04)	-0.116*** (0.04)
Risk Taker		-0.136*** (0.04)	-0.068 (0.04)	-0.055 (0.04)
E(Teammates to correctly report my output)			0.281*** (0.04)	0.273*** (0.04)
Intercept	0.817*** (0.01)	0.866*** (0.02)	0.581*** (0.05)	0.951*** (0.16)
Includes demographic controls	No	No	No	Yes
R-squared	0.037	0.189	0.44	0.476
N	160	160	160	157

Note: OLS with robust standard errors; the omitted treatment is piece rate; column (4) includes controls for age, GPA, birth order, number of siblings, employment status, the number of other participants known and proxies for family wealth.

## Instructions (all treatments)

You have been asked to participate in a production experiment. For your participation today you will earn an amount of money that depends on your performance in the experiment. Any earned money will be paid to you in cash at the end of the experiment.

In this experiment you will be completing a production task that consists of:

- Addressing and printing letters
- Putting the letters in envelopes

- Hand-addressing the envelopes

Each of you will be assigned to one of the work stations in the room. After the experimenter demonstrates how the tasks should be completed, you will work for 30 minutes.

It is important that all these tasks are completed correctly because these letters are addressed to candidates for a new position in the Economics Department.

At the end of 30 minutes you will each be asked to go around to each of the other participant's stations and count their output. You will write down your count of their output on a record sheet that you will be provided. To assure quality, you will then randomly choose one of their letters to open and check, both the letter and the envelope, for mistakes. You will then write down your assessment of the quality of the person's work on the record sheet. Your quality assessment for each individual will be a number between 0 and 1 where 1 indicates that the work is acceptable and 0 indicates that it is not acceptable.

The experimenter will also go around to each work station to count the output and assess the quality of each participant's work.

Do not write your name on any of the sheets you are given. Only put your participant number in the space provided at the top of the record sheet. We do this to maintain anonymity in the experiment. By only using your participant number, the other participants will not know your assessment of their output and the experimenter will never be able to link your actions to you by name.

After all the counts and quality assessments are turned in, you will fill out a brief survey and the experimenter will enter all the count and quality information into a spreadsheet. When every one is done with the survey you will be paid according to the following compensation scheme.

#### *Compensation (Piece Rate)*

The group will be working under a piece rate compensation scheme. Your piece rate compensation will be:

$$Pay = N \times Q \times \$X$$

where  $N$  is the count of your production by the experimenter,  $Q$  is your quality rating assigned by the experimenter and  $\$X$  is the piece rate.

ARE THERE ANY QUESTIONS?

#### *Compensation (Tournament)*

The group will be working under a piece rate compensation scheme. However, there is

also the potential for you to win a bonus. Your piece rate compensation will be:

$$Pay = N \times Q \times \$X$$

where  $N$  is the average count of your production by the experimenter,  $Q$  is your quality rating assigned by the experimenter and  $\$X$  is the piece rate. On top of this, the participant who has the highest quality adjusted output ( $N \times Q$ ) will be awarded a bonus of  $\$Z$ .

ARE THERE ANY QUESTIONS?

*Compensation (Piece Rate with Sabotage)*

The group will be working under a piece rate compensation scheme. Your piece rate compensation will be:

$$Pay = N_{AVG} \times Q_{AVG} \times \$X$$

where  $N_{AVG}$  is the average count of your production by the other participants and the experimenter,  $Q_{AVG}$  is your average quality rating assigned by the other participants and the experimenter and  $\$X$  is the piece rate.

ARE THERE ANY QUESTIONS?

*Compensation (Tournament with Sabotage)*

The group will be working under a piece rate compensation scheme. However, there is also the potential for you to win a bonus. Your piece rate compensation will be:

$$Pay = N_{AVG} \times Q_{AVG} \times \$X$$

where  $N_{AVG}$  is the average count of your production by the other participants and the experimenter,  $Q_{AVG}$  is your average quality rating assigned by the other participants and the experimenter and  $\$X$  is the piece rate. On top of this, the participant who has the highest quality adjusted output ( $N_{AVG} \times Q_{AVG}$ ) will be awarded a bonus of  $\$Z$ .

ARE THERE ANY QUESTIONS?