American Economic Review 2010, 100:1, 1–18
http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.1.1

Tournaments and Office Politics: Evidence from a Real Effort Experiment

By Jeffrey Carpenter, Peter Hans Matthews, and John Schirm*

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

There is still much that 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 (Edward P. Lazear 1989; Kong-Pin Chen 2003; Christian Grund and Dirk Sliwka 2005) than in practice. To our knowledge, there are just two empirical studies of sabotage based on nonexperimental data: Luis Garicano and Ignacio Palacios-Huerta (2005) 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, Robert Drago and Gerald 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 Armin Falk and Ernst 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 Christine Harbring and Bernd Irlenbusch (2004, 2005, 2008), who have considered the effects of variations in the number of competitors, the number of prizes, the prize spread, 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 (Alexander Bruggen and Martin Strobel 2007, for example), we were concerned about the representativeness of chosen effort designs, and share the concerns of Frans van Dijk, Joep Sonnemans, and Frans van Winden (2001, 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 et al. (forthcoming), 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,

* Carpenter: Department of Economics, Middlebury College, 601 Warner Hall, Middlebury, VT 05753, and IZA (e-mail: jpc@middlebury.edu); Matthews: Department of Economics, Middlebury College, 305F Warner Hall, Middlebury, VT 05753, and IZA (e-mail: pmatthew@middlebury.edu); Schirm: Compensation Department, Google, 345 Spear Street, San Francisco, CA 94105 (e-mail: john.schirm@gmail.com). We thank Michael Bougor, Carolyn Craven, Corinna Noelke, and the participants at the 2007 IZA Workshop on Behavioral and Organizational Economics for thoughtful comments. We also acknowledge the financial assistance of the National Science Foundation (Career 0092953) and Middlebury College.

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 performance is 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" (Paul Milgrom and John Roberts 1988) of this sort can even lead firms to abandon internal promotion in favor of external recruitment (William 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 "multisource feedback" (MSF), since compensation depends on the evaluations of both supervisor and peers. Until recently, the consensus (Glenn McEvoy and Paul 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" (Jeffrey S. Kane and Edward E. Lawler 1978) and other norms on the evaluations of others, one manifestation of what Dennis Organ (1988) calls "organizational citizenship behavior."

We find that sabotage or office politics more than reverses 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 a reduction in 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.

The experimental design is described in Section I. In Section II we discuss our results in broad terms. Section III then reports estimates for the output and sabotage functions.

I. Experimental Design

Instead of following in the rich tradition of "effort choice" experiments (e.g., Fehr, Georg Kirchsteiger, and Arno Riedl 1993; 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 James Konow (2000) or Falk and Andrea Ichino (2006) in that our 224 participants were asked to spend 30 minutes preparing letters and envelopes.[1]

In each of the 28 sessions (7 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

EQ 1

[1] Our experimental instructions appear in the Web Data Appendix (available at http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.1.XX).

address an envelope, print the letter, stuff it into the 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 "deliverability" 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 attitudes.

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 $N_{S\to i}$ and his quality assessment is $Q_{S\to i}$, the dollar payoff of worker i in the Piece Rate treatment was

$$\pi_i^{PR} = N_{S\to i}\, Q_{S\to i}.$$

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:

$$\pi_i^{T} = \begin{cases} 25 + (N_{S\to i}\, Q_{S\to i}) & \text{if } N_{S\to i}\, Q_{S\to i} > N_{S\to j}\, Q_{S\to j} \;\; \forall j \neq i \\ N_{S\to i}\, Q_{S\to i} & \text{otherwise} \end{cases}.$$

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.[2]

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 eight 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 one's final output assessment. The average assessed output of worker i is

$$\overline{N_{j\to i}\, Q_{j\to i}} = \left(\sum_{j\neq i} N_{j\to i}\, Q_{j\to i} \right) \Big/ 8,$$

[2] 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 that ability is unobserved in this real effort environment (although it should be randomly distributed across treatments), it would have been difficult to calibrate such an alternative design.

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

$$\pi_i^{T\&S} = \begin{cases} 25 + \overline{(N_{j\to i}\, Q_{j\to i})} & \text{if } \overline{N_{j\to i}\, Q_{j\to i}} > \overline{N_{j\to k}\, Q_{j\to k}} \quad \forall\, k \neq i \\ \overline{N_{j\to i}\, Q_{j\to i}} & \text{otherwise.} \end{cases}$$

Finally, to be complete we also ran a *Piece Rate with Sabotage* treatment that was identical to the Tournament with Sabotage treatment except that no bonus was paid to the producer of the highest quality adjusted output. While it is clear that, in order to assess the pure incentive effect of the tournament in the presence of the possibility for sabotage, one needs to compare 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 material 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.[3]

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 a self-interested worker has the incentive to report zero units produced for each of the seven 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 is such overt acts of sabotage.

To provide an environment more conducive to office politics, we purposely had the participants write the addresses on all of the envelopes by hand 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.

While it is materially costless to sabotage another worker, it does not appear to be costless from a psychological point of view. Lying, and doing so in the context of work, have been demonstrated to create considerable cognitive and moral dissonance (Elliot Aronson 1969; Steven Grover 1993; Rob Holland, Ree Meertens, and Mark van Vugt 2002; Madan M. Palsane, 2005).

[3] Our experiment results in six important behavioral variables: objective output, objective quality, objectively adjusted output, peer adjusted output, output sabotage, and quality sabotage. Using parametric means statistics (student's *t*), nonparametric median statistics (rank sum), and nonparametric CDF statistics (Kolmogornov-Smirnov), we tested whether these six variables differed significantly between the Piece Rate and Piece Rate with Sabotage treatments. In only 1 of 18 instances (the rank sum test of output sabotage differences) did the statistic suggest a significant difference between the two treatments might exist. As one reviewer reminded us, however, the conclusion that there are no differences is nevertheless subject to type II error, a particular concern with nonparametric statistics.

II. 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 percent were international students and the mean grade point average (GPA) of the participants was relatively high. The two demographics that we were particularly interested in are sex and whether 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 handwriting, and international students will be less familiar with English and the US 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 (Toni U. Falbo 1981).

Based on 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 parents' 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 attitudes of our participants we had the participants complete the Elke Weber, Ann-Renee Blais, and Nancy Betz (2002) risk preference scale and created an indicator variable for those persons who ranked among the top 10 percent in risk taking. We chose to use the Weber et al. measure because we were interested in a broader measure of risk attitudes that included responses in the social and moral domains. We also asked each participant to tell us the number of other participants in the session whom 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 other participants.

Finally, 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 (86 percent) 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 59 percent 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.[4] Perhaps the most important thing to notice is the disincentive effect of the threat of sabotage. As reported elsewhere (e.g., Haig Nalbantian and Andrew Schotter 1997 or Van Dijk et al. 2001), adding the $25 bonus for the highest producer increases raw effort (proxied here by the black bars representing the postal worker envelope count). The number of envelopes produced increases by 1.40 on average ($z = 2.64$, $p < 0.01$) when we move from piece rates to the tournament.[5] The difference of 1.40 envelopes does not seem large, but remember that our participants produced

[4] Note that the peer adjusted count is the mean assessment of the seven other participants, 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.

[5] We report z statistics from the nonparametric Wilcoxon test.

Table 1—Participant Characteristics

	N	Mean	SD	Min	Max
Male	224	0.473	0.500	0	1
International student	224	0.152	0.360	0	1
Risk scale	224	127.670	23.693	78	203
Risk taker (90th percentile risk scale)	224	0.100	0.292	0	1
E (teammates to correctly report my output)	224	0.861	0.346	0	1
GPA	222	3.491	0.278	2.55	4
First born	224	0.585	0.494	0	1
Number of siblings	224	1.509	1.108	0	7
Number of bathrooms in parents' house	224	3.018	1.448	0	9
Have a car on campus	224	0.393	0.489	0	1
Employed	224	0.634	0.483	0	1
Number of participants known	223	1.264	1.214	0	8

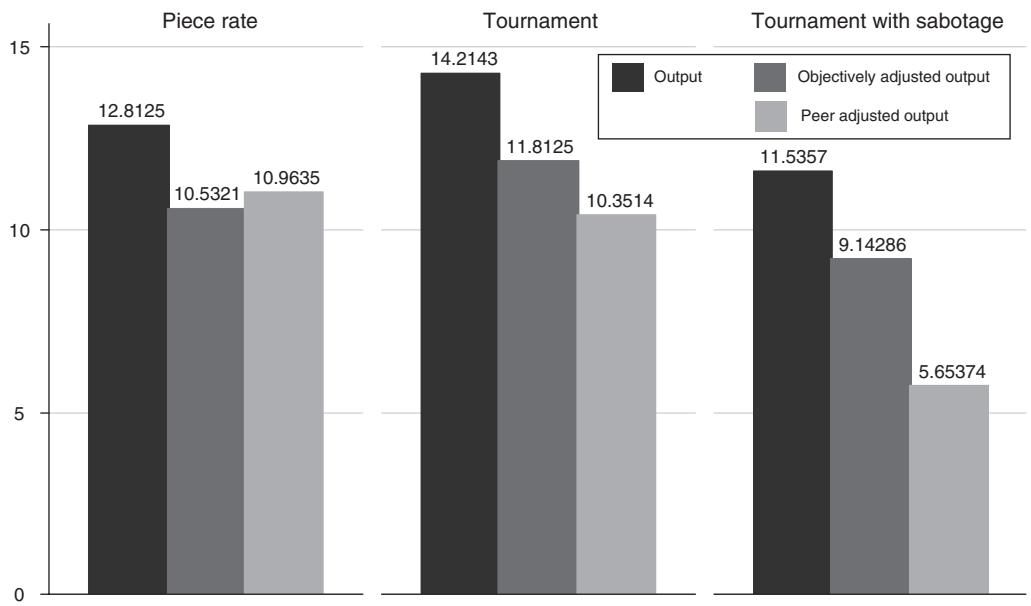

Figure 1. Mean Production Levels by Treatment

for only 30 minutes. Over the course of an eight hour day (and in the absence of diminishing returns) the eight workers would produce 179 more envelopes in the Tournament condition. What is important, however, is that raw output actually falls when competitors are able to sabotage each other relative to both the Tournament ($z = 4.15$, $p < 0.01$) *and* the Piece Rate ($z = 1.98$, $p = 0.05$) treatments. It is one thing to discover the presence of sabotage and the reduction of output as the direct result of that sabotage, but to find that the very potential for sabotage acts as a large disincentive to providing effort is unique to this experiment. On average, workers produce 2.68 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 2.28 quality adjusted envelope difference in the Piece Rate treatment is similar to the 2.40 difference in the Tournament and the 2.39 average envelope difference in the Tournament with Sabotage treatment. However, if one regresses quality on treatment indicators only, the small differences appear to be significant. The Tournament with Sabotage treatment yields lower quality of 0.04 compared to both the Piece Rate and the Tournament treatments ($p < 0.10$ in both cases).[6]

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 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.[7] 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 more than one quality adjusted envelope is significant ($z = 2.73$, $p < 0.01$) and could be symptomatic of a simple, maybe even unconscious or affective/emotional, response to competition (Oliver Schultheiss, Kenneth Campbell, and David McClelland 1999). Of course, without direct measures of emotions, this plausible interpretation is merely speculation.

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 more than two adjusted envelopes less than the letter carrier, and the difference is highly significant ($z = 6.24$, $p < 0.01$). As a rough comparison of magnitudes, if the effects of the bonus and the ability to reduce the output of others are additive, it appears that the affective or emotional response isolated in the Tournament accounts for 42 percent of the overall sabotage that occurs in the Tournament with Sabotage treatment.[8] The remaining 58 percent of sabotage is likely to be cognitive and strategy-driven.

III. 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 1 percent level. Relative to the combined Piece Rate treatment, adjusted output is estimated to rise 1.28 per person, or $10.24 = 8(1.28)$ per team, in the Tournament. When sabotage is possible, on the other hand, adjusted output per worker falls relative not only to the tournament ($t = 2.67$, $p < 0.01$), but also 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. They are consistent, we believe, with the nature of the task, one in which keyboard dexterity, neat handwriting, and fluency in English were favored.

[6] Recall that quality is measured on $[0, 1]$ scale.

[7] However, the difference is not significant ($z = 0.81$, $p = 0.41$).

[8] In the Tournament with Sabotage treatment, 1.46 of the 3.50 difference between postal worker assessment and peer assessment.

Table 2—Analysis of Objectively Adjusted Output

AQ2

	(1)	(2)	(3)	(4)	(5)
Tournament	1.280***	1.384***	1.299***	1.408***	1.397**
	(0.47)	(0.44)	(0.44)	(0.46)	(0.63)
Tournament with sabotage	−1.389***	−1.535***	−0.454	−0.345	−0.407
	(0.50)	(0.47)	(0.46)	(0.44)	(0.67)
Male		−1.260***	−1.184***	−1.156***	−1.050***
		(0.37)	(0.36)	(0.38)	(0.34)
International student		−2.118***	−2.069***	−2.141***	−1.932***
		(0.51)	(0.48)	(0.54)	(0.53)
Risk taker		−1.467**	−0.799	−0.754	−0.525
		(0.62)	(0.63)	(0.67)	(0.61)
E (teammates to correctly report my output)			3.028***	2.914***	2.911***
			(0.54)	(0.57)	(0.56)
Intercept	10.532***	11.598***	8.634***	9.641***	7.895***
	(0.29)	(0.35)	(0.64)	(2.78)	(2.50)
Includes demographic controls	No	No	No	Yes	Yes
Includes session random effects	No	No	No	No	Yes
R^2	0.09	0.21	0.30	0.31	-
Observations	224	224	224	221	221

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

AQ3

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

The observation that risk takers are estimated to produce 1.467 fewer envelopes, and that the difference is significant at the 5 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 some support for this view in two unreported regressions in which we analyze raw quantity and objective quality separately. The quantity regression suggests that risk takers do produce about one more envelope but the effect is only significant at the 20 percent level, while the second regression indicates that risk takers generate 11 percent less quality on each ($p < 0.01$).

Columns 3 and 4 of 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 more important, the treatment effect for tournaments with sabotage vanishes. These two key results also hold in the last column of Table 2 when we test the robustness of our results by adding session-level random effects to account for the fact that all the participants within a session shared a common printer. As the reader can see, neither the point estimates nor the significance of our results change appreciably. It seems reasonable to conclude, then, that quality adjusted output falls in this environment *because* fewer individuals expect their coworkers 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.

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 2A illustrates the incidence and amount of output sabotage in the three treatments. As one can see, almost 80

percent of the evaluations were correct in the Piece Rate treatment, but this number fall to less than 70 percent in the two tournament treatments. What differentiates the two tournament treatments, however, is that there are a lot of observations well to the right of zero in the Tournament with Sabotage, indicating significant sabotage. It is one thing to miscount by one or two envelopes as occurs in the Piece Rate and Tournament treatments, but there are 44 instances of the target's output being evaluated as less than five units and 17 instances in which the evaluator reduced the target's output to zero. All of these instances occurred in the Tournament with Sabotage.

For a more systematic analysis of output sabotage, column 1 of Table 3 reports GLS estimates with individual random effects (v_{ij}) for output sabotage based on the spline function:

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

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 $\vec{\mathbf{X}}$ is a vector of controls. As one can immediately see by the coefficient on the TS indicator, there was significantly more output sabotage in the Tournament with Sabotage than in the Piece Rate ($p < 0.01$) or the Tournament ($p = 0.07$) treatments. The saboteurs were also sensitive, in general, to the difference between their output and the target's output: for each additional envelope produced by the target over the saboteur, the saboteur reduces the target's count by approximately one-tenth of an envelope ($p < 0.01$), and this gradient is approximately twice as steep for saboteurs in the Tournament ($p = 0.02$). While the slope of the output sabotage function is not overall steeper for the Tournament with Sabotage saboteurs, it does become significantly steeper when the target produces strictly more than the saboteur ($p < 0.01$).

The results reported in the first column of Table 3 are also robust to the inclusion of session-level random effects. The mixed model reported in column 2 is almost identical, in fact, to the one reported in column 1.

All the interaction terms make it difficult to visualize the individual behavior of workers, so we graph output sabotage by treatment as a function of the output difference in Figure 3A. Recalling Figure 1, participants in the Piece Rate treatments have nothing to fear from sabotage, on average. The output sabotage function stays below the horizontal axis until very high levels of disadvantageous inequality. Likewise, Tournament producers begin to see positive amounts of sabotage once they produce four more units than the other participants and, as one might expect, even participants who produce up to three units *less* than the evaluator need to worry in the Tournament with Sabotage.

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 2B 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 2A, however, we see that the number of cases in which the peer evaluator agreed with the letter carrier or favored her peer on quality 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 the mode (which is to be slightly nicer to one's peers than the postal carrier) 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

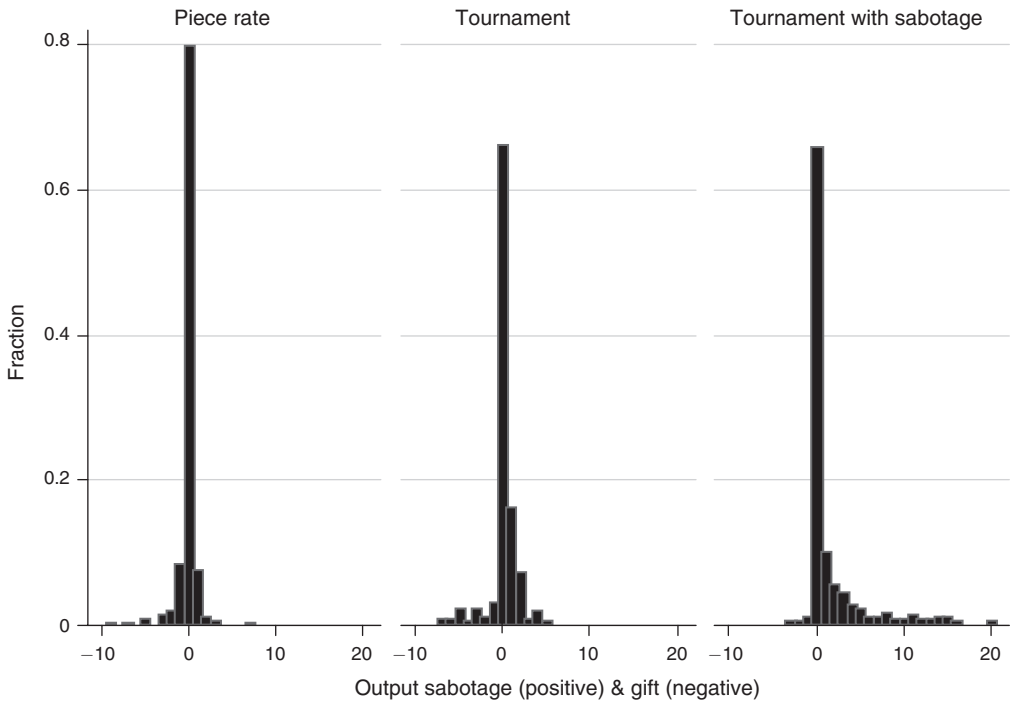

FIGURE 2A. OUTPUT SABOTAGE BY TREATMENT

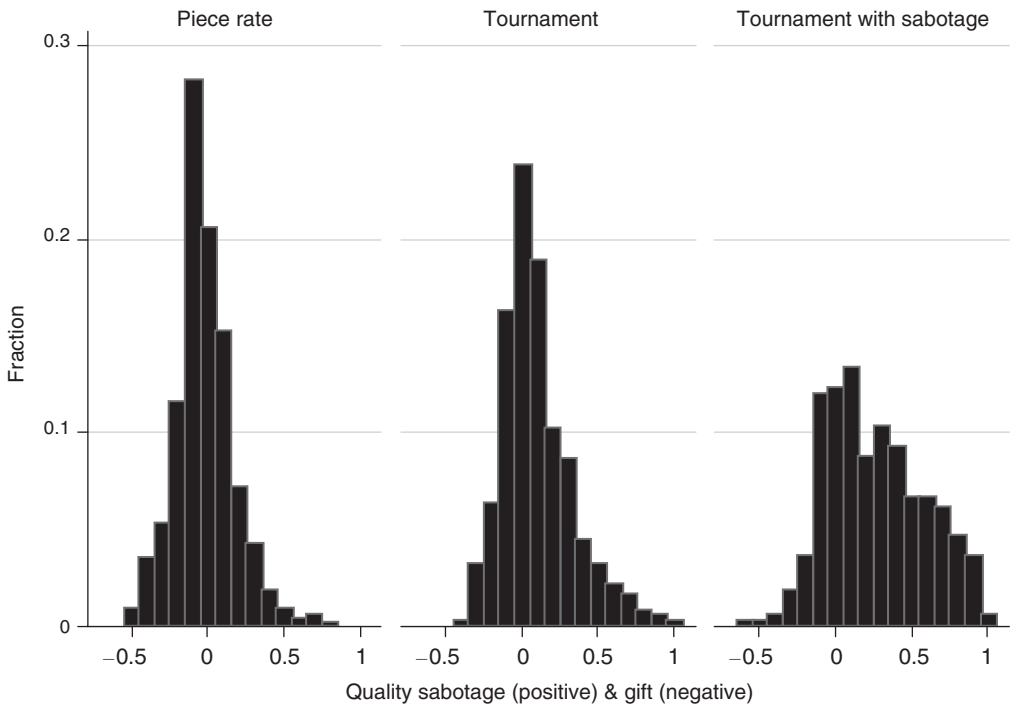

FIGURE 2B. QUALITY SABOTAGE BY TREATMENT

Table 3—Analysis of Sabotage

	Output sabotage (1)	Output sabotage (2)	Quality sabotage (3)	Quality sabotage (4)
Tournament	0.312	0.316	0.124***	0.121***
	(0.28)	(0.30)	(0.03)	(0.04)
Tournament with sabotage	0.947***	0.960***	0.241***	0.248***
	(0.30)	(0.32)	(0.03)	(0.04)
(Target's output − saboteur's output)	0.085***	0.085***	0.003	0.003
	(0.03)	(0.03)	(0.01)	(0.01)
(Target's output − saboteur's output) × tournament	0.135**	0.137**	0.015**	0.015*
	(0.06)	(0.06)	(0.01)	(0.01)
(Target's output − saboteur's output) × tournament with sabotage	0.031	0.035	0.008	0.008
	(0.04)	(0.04)	(0.01)	(0.01)
max {(target's output − saboteur's output), 0}	−0.073	−0.073	−0.008	−0.009
	(0.05)	(0.05)	(0.01)	(0.01)
max {(target's output − saboteur's output), 0} × tournament	−0.073	−0.075	−0.008	−0.006
	(0.09)	(0.09)	(0.01)	(0.01)
max {(target's output − saboteur's output), 0} × tournament with sabotage	0.381***	0.376***	0.024**	0.023**
	(0.07)	(0.07)	(0.01)	(0.01)
Intercept	−0.622	−0.605	0.100	0.118
	(1.38)	(1.45)	(0.15)	(0.16)
Includes demographic controls	Yes	Yes	Yes	Yes
Includes session random effects	No	Yes	No	Yes
Wald χ^2	268	285	157	133
Observations	1,538	1,538	1,538	1,538

Notes: GLS with individual random effects; the omitted treatment is piece rate; both columns include controls for sex, international student status, risk taking, expectations, GPA, birth order, number of siblings, employment status, the number of other participants known, and proxies for family wealth.

AQ 4

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

in the Tournament with Sabotage. There is even a significant amount in the Tournament, despite there being no material incentive to sabotage one's coworkers.

Column 3 of Table 3 reports estimates for a model that, apart from the dependent variable, is identical to that used to characterize quantity sabotage:

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

where QS_{ij} is now i's underestimate (positive) or overestimate (negative) of the quality of j's output. Quality sabotage seems to take a form that is very similar to output sabotage. This time, however, both of the tournament treatments elicit significantly more sabotage than the piece rate treatments, and the Tournament with Sabotage doubles the increment over the Tournament ($p < 0.01$). 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

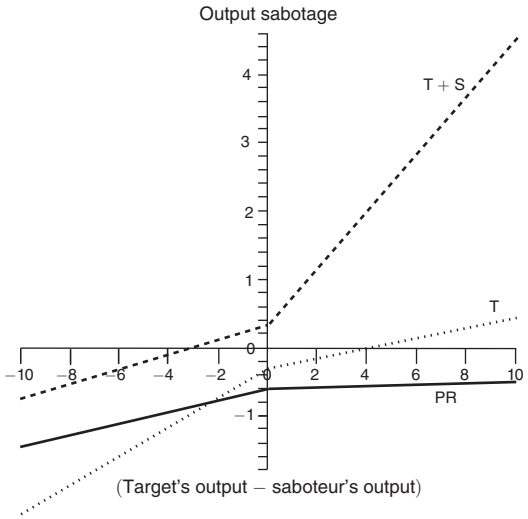

FIGURE 3A.
OUTPUT SABOTAGE FUNCTIONS BY TREATMENT

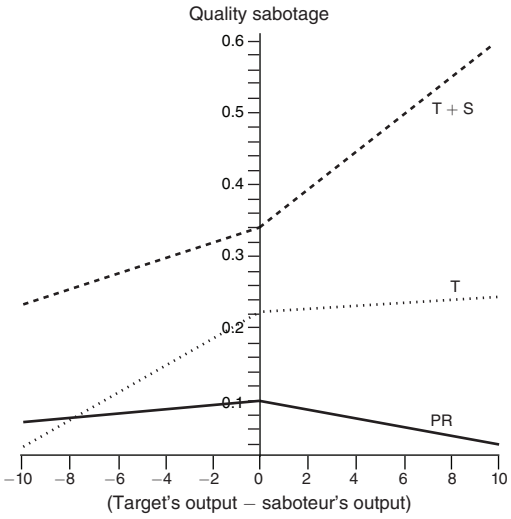

FIGURE 3B.
QUALITY SABOTAGE FUNCTIONS BY TREATMENT

tournaments with sabotage, but even in the standard tournament, it exists as what we conjecture to be an affective response.

While the general trend with respect to the output difference is weaker than with output sabotage, we see again that the gradient is steeper overall in the Tournament ($p = 0.05$) and considerably steeper in the Tournament with Sabotage when the saboteur is at a productive disadvantage ($p = 0.02$). Again, these results are robust to the inclusion of session-level effects (column 4). As shown in Figure 3B, one's chances of incurring quality sabotage in the Piece Rate treatment are the greatest when one produces as much as one's evaluator, but the amount of sabotage received in the other two treatments is monotonically increasing in the output difference. Figure 3B also illustrates that the "sabotage threshold"—that is, the difference in output sufficient to induce underestimation—is not just negative, but very different, in the two sorts of tournaments, and that below this, the amount of "negative sabotage" (i.e., gifts) increases more quickly in the standard tournament.

IV. 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, but the atmosphere itself created by a tournament can be a disincentive to work hard. Both of these results have implications for the firm's profits above the obvious problem of destroyed output studied in previous experiments. This second, disincentive to provide effort, effect is the main empirical contribution of our study.

Considering that the motivational disincentive effect of expected sabotage is a novel result, some speculation may be in order. It seems reasonable, for example, to conjecture that the expectation of sabotage causes the marginal benefit of effort schedule to shift down in the tournament, causing effort and therefore output to fall. It is also possible that, consistent with the recent neurophysiological work of Klaus Fliessbach et al. (2007), workers concerned that the "wrong" peer will win the tournament reduce effort even further.

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 (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 1.28 units. This means an increase of $1.28 \times 8 = 10.24$ units for the team. Since labor costs in the tournament are therefore $35.24 = 25 + 10.24$ more, the tournament is worth it only if $10.24\rho \geq 35.24$ or $\rho \geq 3.44$, where ρ is the sales price per finished envelope.

REFERENCES

Aronson, Elliot. 1969. "The Theory of Cognitive Dissonance: A Current Perspective." In *Advances in Experimental Social Psychology,* ed. L. Berkowitz, 1–34. New York: Academic Press.

Bruggen, Alexander, and Martin Strobel. 2007. "Real Effort versus Chosen Effort in Experiments." *Economics Letters*, 96(2): 232–36.

Chan, William. 1996. "External Recruitment versus Internal Promotion." *Journal of Labor Economics*, 14(4): 555–70.

Chen, Kong-Pin. 2003. "Sabotage in Promotion Tournaments." *Journal of Law, Economics, and Organization*, 19(1): 119–40.

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

Falbo, Toni. 1981. "Relationships between Birth Category, Achievement, and Interpersonal Orientation." *Journal of Personality and Social Psychology,* 41(1): 121–31.

Falk, Armin, and Ernst Fehr. 2003. "Why Labour Market Experiments?" *Labour Economics*, 10(4): 399–406.

Falk, Armin, and Andrea Ichino. 2006. "Clean Evidence on Peer Effects." *Journal of Labor Economics*, 24(1): 39–57.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics*, 108(2): 437–59.

Fliessbach, Klaus, Bernd Weber, Peter Trautner, Thomas Dohmen, Uwe Sunde, Christian Elger, and Armin Falk. 2007. "Social Comparison Affects Reward-Related Brain Activity in the Human Ventral Striatum." *Science*, 318(5854): 1305–08.

Garicano, Luis, and Ignacio Palacios-Huerta. 2005. "Sabotage in Tournaments: Making the Beautiful Game a Bit Less Beautiful." Centre for Economic Policy Research Discussion Paper 5231.

Grover, Steven. 1993. "Lying, Deceit, and Subterfuge: A Model of Dishonesty in the Workplace." *Organization Science*, 4(3): 478–95.

Grund, Christian, and Dirk Sliwka. 2005. "Envy and Compassion in Tournaments." *Journal of Economics and Management Strategy*, 14(1): 187–207.

Harbring, Christine, and Bernd Irlenbusch. 2004. "Sabotage and the Endogenous Design of Tournaments." http://www.iza.org/conference_files/wepe2004/irlenbusch_b1073.pdf.

Harbring, Christine, and Bernd Irlenbusch. 2005. "Incentives in Tournaments with Endogenous Prize Selection." *Journal of Institutional and Theoretical Economics*, 161(4): 636–63.

Harbring, Christine, and Bernd Irlenbusch. 2008. "How Many Winners are Good to Have? On Tournaments with Sabotage." *Journal of Economic Behavior and Organization*, 65(3–4): 682–702.

Harbring, Christine, Bernd Irlenbusch, Matthias Krackel, and Reinhard Selten. Forthcoming. "Sabotage in Asymmetric Contests: An Experimental Analysis." *International Journal of the Economics of Business***.**

Holland, Rob, Ree Meertens, and Mark van Vugt. 2002. "Dissonance on the Road: Self-esteem as a Moderator of Internal and External Self-Justification Strategies." *Personality and Social Psychology Bulletin*, 28(12): 1713–24.

Kane, Jeffrey S., and Edward E. Lawler. 1978. "Methods of Peer Assessment." *Psychological Bulletin*, 85(3): 555–86.

Konow, James. 2000. "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review*, 90(4): 1072–91.

Lazear, Edward P. 1989. "Pay Equality and Industrial Politics." *Journal of Political Economy*, 97(3): 561–80.

McEvoy, Glenn, and Paul Buller. 1987. "User Acceptance of Peer Appraisals in an Industrial Setting." *Personnel Psychology*, 40(4): 785–97.

Milgrom, Paul, and John Roberts. 1988. "An Economic Approach to Influence Activities in Organizations." *American Journal of Sociology*, 94: 154–79.

Nalbantian, Haig R., and Andrew Schotter. 1997. "Productivity under Group Incentives: An Experimental Study." *American Economic Review*, 87(3): 314–41.

Organ, Dennis. 1988. *Organizational Citizenship Behavior: The Good Soldier Syndrome*. Lanham, MD: Lexington Books.

Palsane, Madan N. 2005. "Self-Incongruent Behaviour, Stress and Disease." *Psychological Studies*, 50(4): 283–97.

Schultheiss, Oliver, Kenneth Campbell, and David McClelland. 1999. "Implicit Power Motivation Moderates Men's Testosterone Responses to Imagined and Real Dominance Success." *Hormones and Behavior*, 36(3): 234–41.

van Dijk, Frans, Joep Sonnemans, and Frans van Winden. 2001. "Incentive Systems in a Real Effort Experiment." *European Economic Review*, 45(2): 187–214.

Weber, Elke, Ann-Renée Blais, and Nancy Betz. 2002. "A Domain-Specific Risk Attitude Scale: Measuring Risk Perceptions and Risk Behaviors." *Journal of Behavioral Decision Making*, 15(4): 263–90.

PLEASE ANSWER ALL AUTHOR QUERIES (numbered with "AQ" in the margin of the page). Please disregard all Editor Queries (numbered with "EQ" in the margins). They are reminders for the editorial staff.

AQ#	Question	Response
1.	Units on y-axis?	
2.	Would you like to add column headers?	
3.	Are these correct?	
4.	Are these correct?	

Numbered with "EQ" in the page margin.

EQ#	Question	Response
1.	Add URL to footnote 1.	

PROOFREADERS' MARKS

SYMBOL	MEANING	EXAMPLE
	delete	take it out
	close up	per cent
	delete and close up	remove
^	insert something here	something is missing
#	space	too close
eq #	space evenly	space these consistently
stet	let stand	ignore marks and leave as was
tr or 	transpose	this backwards is
/	used to separate 2 or more marks in margin	; / ;
][	center	]this should center[
[	set farther to the left	[move left
]	set farther to the right	move right]
=	align horizontally	align with surrounding text
||	align vertically	align with surrounding text
	move to next line	
¶	begin new paragraph	
sp	spell out	set PA as Pennsylvania
cap or ≡	set in capitals	ALL CAPS
sm cap or s.c.	set in small capitals	SMALL CAPITALS
lc	set in lowercase	lower case
ital	set in italic (underline the text)	*italic*
rom	set in roman	roman or regular
bf or	set in bold (squiggly underline of text)	**BOLD**
-	hyphen-used to join words and to separate syllables	
en or 1/n	en dash–a connection between two things	2006–2007
em or 1/m	em (long) dash—indicates a sudden break in thought	
V	superscript or superior	$E=MC^2$
^	subscript or inferior	H_20
	centered	for a centered dot in p ⋅ q
;^	comma	red, white, and blue
'^	apostrophe	my sister's friend's investments
	period	the end.
;	semicolon	he said; she said
:	colon	what follows proves: clarifies
"^ "^ or '^ '^	quotations marks	"the economist"
(/)	parentheses	(like this)
[/]	brackets	[like this]
wf	wrong font	wrong size or style

COPYRIGHT TRANSFER AGREEMENT

From: **The American Economic Association** **Fax: +1 412-431-3014**
The American Economic Review
2403 Sidney Street, Suite 260
Pittsburgh, PA 15203

To: **Author** (please print name here) _____

The American Economic Association (hereinafter Association) is pleased to have the opportunity to publish your manuscript in the *American Economic Review*. In order that the Association, as Publisher, may obtain copyright protection for the contents of the Journal, it is necessary for you to execute this formal transfer of your copyrights in this manuscript to the Association.

The Association acknowledges the receipt of your manuscript titled _____

_____,

to appear in the March 2010 issue of *AER*.

Consent to Publish

In consideration of the publication by the Association of the above-named manuscript, the undersigned as Author(s) transfer(s) exclusively to the Association all rights, title and interest defined by the Copyright Law of the United States in and to the above-named manuscript in its entirety, including all subsidiary rights. The rights transferred herein shall remain the property of the Association for the full duration of these rights under the Copyright Law of the United States. If it should become necessary, the Author(s) agree(s) to assist the Association in registering and enforcing the Copyright in the name of the Association. The Association shall have the right to publish the above-named manuscript in print, sound or video recordings, magnetic media (i.e., computer disk, CD-ROM, etc. . . .) electronic media (including transmission via the Internet, or any other computerized communication network), or any other technology for publication of this work which may hereinafter be developed.

The Association, in turn, grants to the Author(s) the right to republication in any work in any form, including digital repositories in universities and other institutions subject only to giving proper credit of copyright. The Association further grants to the Author(s) the right to distribute the above-noted work in any classroom in which he or she is a teacher, subject only to the Author(s) giving proper credit in any such derivative work and on any copies distributed for classroom use. Proper notice may be given as follows: [Copyright _____, American Economic Association; reproduced with permission of the *American Economic Review*].

Permission to Reprint Policy

The Author(s) may specify the degree of access to which the Association grants others the right to reproduce the Author(s)' material. Check one:

_____ Implicit consent: Grants anyone permission to reprint in all places in all forms provided that the appropriate copyright information is included and the Association is notified that the work is being reprinted.

_____ Explicit consent: Requires direct consent of the Author(s) and the Association before any republication is allowed. The republisher must obtain from the Author(s) permission to reprint all or any major portion of the Author(s)' manuscript. Author(s) may charge a fee for reprint or translation rights.

Rights to translate are retained by the Author(s) and dealt with on a case by case basis.

Warranty of Authorship

The Author(s) warrant(s) that the above-named manuscript is his or her own original work of authorship and has not been published previously. If any material included by the Author(s) in the above-noted manuscript (including tables, charts, or figures) is the work of another author or is otherwise under prior copyright protection by another proprietor, the Author(s) undertake(s) to obtain permission from that copyright proprietor for the inclusion of such material in this manuscript to be published by the Association. The Author(s) further agree(s) to save and hold the Association harmless in any suit for infringement arising from the Author(s)' unauthorized use of copyrighted material. The Author(s) agree(s) to submit to the Editor of the Journal of the Association to whom the manuscript has been submitted, copies of all letters of permission to include copyrighted material of another author included in the subject manuscript by this Author or material written by Author(s) that is under prior copyright protection by another proprietor.

The Author(s) further warrant(s) that this manuscript was not written as part of his or her official duties as an employee(s) of the United States government. Since copyright protection is not available for a work of the United States government, the Author(s) agree(s) to disclose fully to the Association the circumstances of federal employment which might invoke this bar to copyright protection of the manuscript by signing below to confirm the author warranties.

The Author(s) further warrant(s) that this manuscript was not written as an employee so as to constitute a work-for-hire in which the ownership of the copyright is in that employer.

Please sign and date this agreement. *Return one copy* to the Editor of the *American Economic Review* promptly and *retain one copy.* A manuscript for which there is no valid Copyright Transfer Agreement cannot be published.

Accepted and approved: _____
Author(s)

Government Employees please sign here: _____ Date: _____

For the American Economic Association and the *American Economic Review*:

_____