Does sabotage attenuate the incentive effect of tournaments? Evidence from a real effort experiment

Jeffrey Carpenter Peter Hans Matthews John Schirm

May 22, 2007

— Preliminary and incomplete first draft —

1 Introduction

Despite firms' frequent adoption of, and economists' durable interest in, compensation schemes based on relative performance measures, there is still much to learn about the effects of tournaments on individual effort in practice. We understand in principle, for example, the effects of sabotage on tournament incentives (Lazear 1989, Chen 2003), even when workers are prosocial (Grund and Sliwka 2003), but are much less certain about both its practical causes and consequences. There are, to our knowledge, just two empirical studies of sabotage based on non-experimental data. In the now rich tradition of tests based on professional sports data, Garicano and Palacios-Huerta (2006) find that when soccer teams have more incentive to win, there is both more offensive effort and more dirty play, but no net change in scoring. Drago and Garvey (1998), on the other hand, conclude on the basis of a survey of Australian manufacturing workers that sharper promotion incentives lead to a reduction in "helping efforts."

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 (2005, 2007a, 2007b) who, in a series of papers, have explored the effects of variations in the number of competitors, the number of prizes, the prize spread and contract choices available to firms/principals 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, in which subjects

subjects were required to complete and print a series of form letters, and then stuff these into hand addressed envelopes. While there is now some debate about the differences between real and chosen effort designs (Bruggen and Strobel 2005, for example), we are sympathetic to the views of van Dijk, Sonnemans and van Winden (2001: 189), who suspect that the replacement of work, which "involves effort, fatigue, boredom, excitement and other affectations not present in the [more abstract notion of chosen effort]" influences tournament behavior and therefore outcomes.

Second, but 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 *blunt* and *diffuse*. It is diffuse in the sense that, with the notable exception of Harbring, Irlenbusch, Krackel and Selten (2000), 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. And it is blunt because what is represented in most, if not all, experiments is the actual destruction of output, the most extreme form of sabotage. To win a tournament, a worker need not produce more than her rivals, but rather create the impression, well-founded or otherwise, that she has produced more. In cases where individual effort levels are difficult to rank order, let alone measure, this is often a simpler, but more subtle, task. Much of what we call "office politics" is about the formation of beliefs about relative performance. Our design allows for two forms of directed sabotage: subjects counted and evaluated their quality of each competitor's output, and in the sabotage treatments, compensation depended, in part, on these evaluations.

[SUMMARY OF RESULTS]

2 Experimental Design

In all four treatments, each of 8 student subjects was provided with his or her 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 and was salient to our subjects. The task was not as simple as first seems: from start to finish,

each letter required between 60 and 90 seconds to complete. Subjects had 30 minutes to complete letters from a list that contained more names and addresses than could ever be exhausted.

After the production period, all of the subjects and a "supervisor," one of the experimenters, went around the room and examined all of the output boxes. Each counted and recorded the number of completed envelopes in each 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 experimenters nor subjects had a vested interest, was later needed to determine levels of sabotage, a member of the US Postal Service was hired to evaluate *all* of 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 based on Weber, Blais and Betz (2002) scale.

The treatments differ with respect to the method of compensation and the opportunities for sabotage. In the first treatment, denoted PN, subjects were paid a piece rate of \$1 per "quality adjusted envelope," where both the count and the evaluation of quality are the supervisor's. The second, TN, is the no sabotage tournament: in addition to the same piece rate available udner PN, the subject who produced the most quality adjusted envelopes received a bonus of \$25. The third, PS, is like PN except that the raw counts and quality ratings are the combined subject and supervisor averages. Other than inequality aversion, there is no obvious reason for subjects to undercount or underestimate the quality of other output boxes under PS and, if so, no reason to expect differences in adjusted output. Because this is what the data show, the PS and PN treatments have been consolidated. Last, in the tournament with sabotage or TS treatment, the \$25 prize is restored but determined on the basis of the combined subject and supervisor averages.

3 Descriptive Statistics

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

4 Results

As the first column in Table 2 reveals. when adjusted individual output is regressed on just the tournament and tournament with sabotage treatment indicators, both estimated coefficients are significant at the 10 percent level. Relative to the combined piece rate treatments, adjusted output is estimated to rise 1.171 per person, or $9.368 = 8(1.171)$ per team, in the tournament. (Since labor costs in the tournament are therefore $34.368 = 25 + 9.368$ more, it will also be more profitable if $9.368p \geq 34.368$ or $p \geq 3.57$, where p is the price per finished envelope.) When sabotage is possible, on the other hand, adjusted output per worker falls not just relative to the tournament, but to the combined piece rate treatments: if $p = 3.57$, for example, so that the tournament and piece rates are equally profitable, the firm loses 55.716, 42.668 of which is the result of lower revenues, or the equivalent of almost 16 envelopes.)

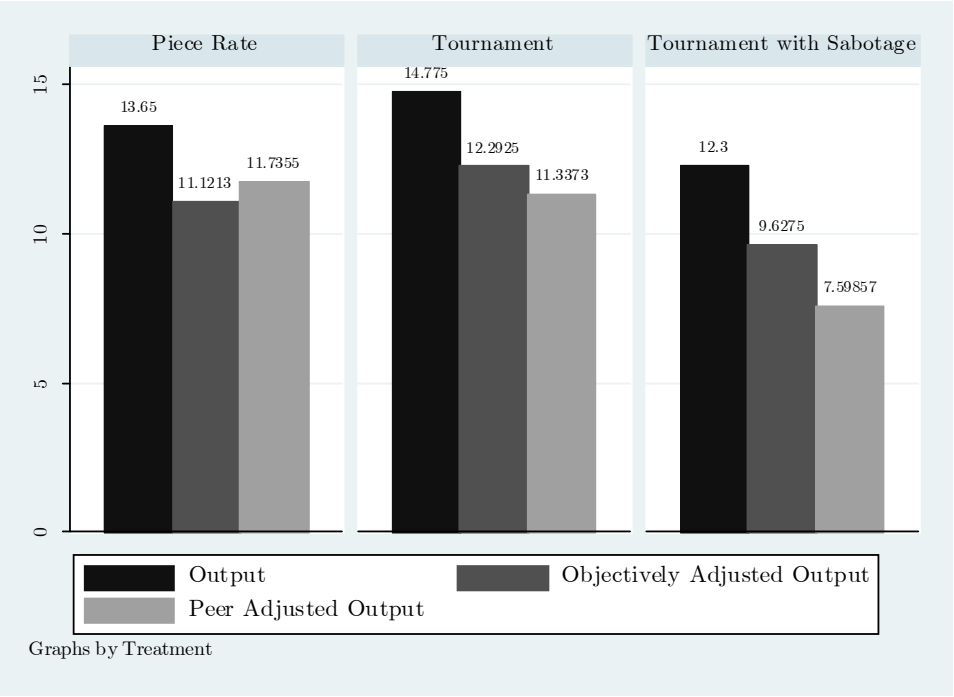

Figure 1

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 skills, neat handwriting and facility 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, requires a more subtle interpretation. It is our hypothesis that risk takers are more likely than other subjects to skimp on quality in an effort to finish more envelopes. We find support for this view in separate but unreported quantity and quality regressions and in the third column, which reveals that when another indicator, whether subjects expect their output to be reported correctly, is added to the model, the risk taker coefficient shrinks in size and significance.

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.

More important, the addition of this expectations variable also causes the coefficient for tournaments with sabotage to vanish. Quality adjusted output is lower in tournaments with sabotage, in other words, because fewer individuals expect their co-workers to be truthful about their contributions to production and, as a result, effort is scaled back. In more evocative terms, firms in which promotion is the result, at least in part, of office politics, the response of workers will be to produce less.

Furthermore, as the results in the fourth column reveal, this observation is robust with respect to the addition of other demographic controls and even the number of other subjects known.

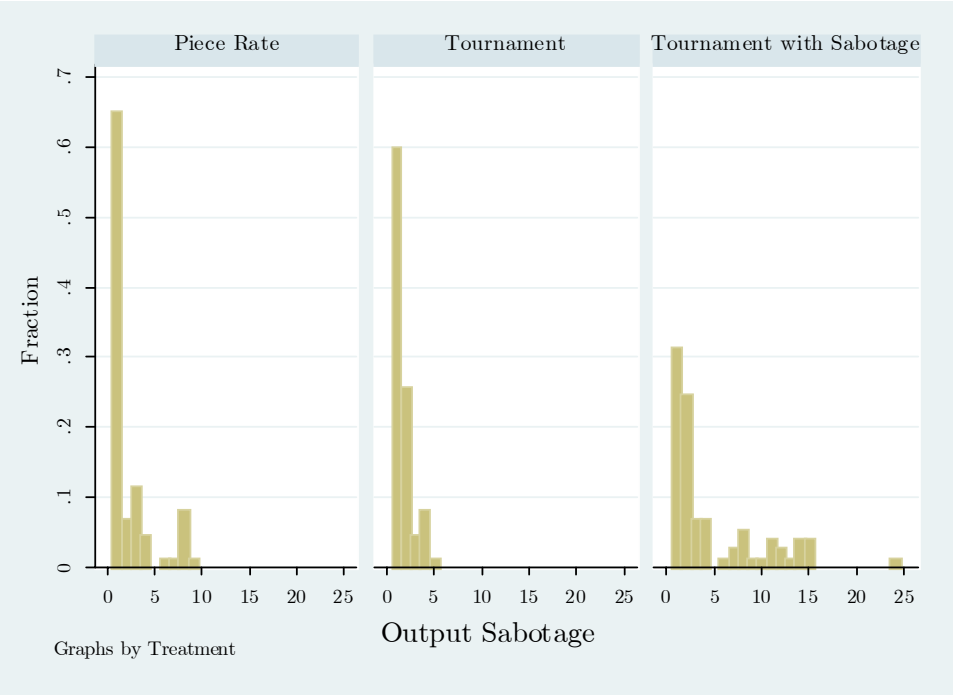

Figure 2

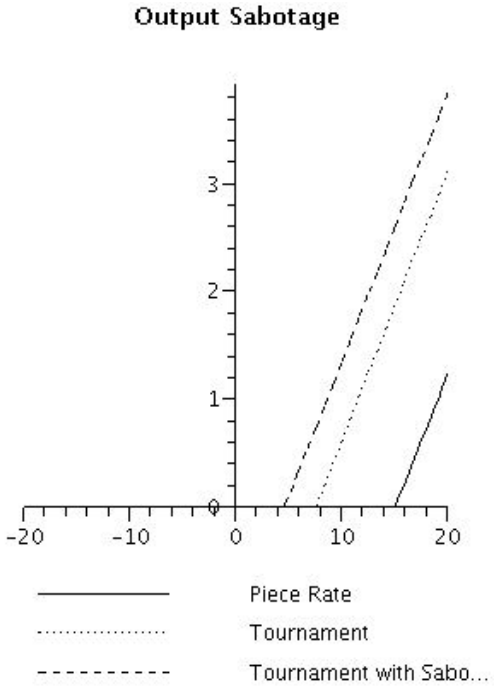

Figure 3

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? Table 3 reports the estimates for a series of panel tobit models for the subsample of cases in which either

positive or no output sabotage was observed. The first specification, for example, is limited to the two treatment indicators (the combined piece rate treatment serves as the default) and the difference between the target's output and the saboteur's. All three coefficients are significant at the 1 percent level, and produce the "sabotage functions" depicted in Figure 3. In particular, subjects who produce more than their co-workers do not undercount their output under any compensation scheme.

TABLE 3: Analysis of Output Sabotage

Tournament	1.882***	1.874***	1.824***	1.848***
	(0.56)	(0.55)	(0.54)	(0.55)
Tournament with Sabotage	2.612***	2.571***	1.612**	1.708**
	(0.57)	(0.57)	(0.68)	(0.70)
(Target's Output - Saboteur's Output)	0.250***	0.265***	0.245***	0.257***
	(0.05)	(0.05)	(0.05)	(0.05)
Male		0.448	0.415	0.596
		(0.46)	(0.45)	(0.48)
International Student		-2.700***	-2.636***	-2.721***
		(0.80)	(0.80)	(0.84)
Risk Taker		0.123	-0.415	-0.168
		(0.76)	(0.78)	(0.82)
E(Teammates to correctly report my output)			-2.261***	-1.956**
			(0.86)	(0.88)
Intercept	-3.757***	-3.728***	-1.426	-11.150**
	(0.39)	(0.46)	(0.98)	(4.93)
Includes demographic controls	No	No	No	Yes
rho	0.289	0.271	0.253	0.242
Wald Chi-squared	54.842	66.622	75.683	89.804
N	945	945	945	925

Note: Tobit with lower limit of zero; 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.

Furthermore, subjects who produce less than their co-workers do not undercount their output under piece rates unless the difference is *very* large.

Output differences do lead to significant output sabotage in both tournament treatments, more so, as one would expect, when it has real consequences: in the tournament with

sabotage treatment, a difference of even four envelopes is sufficient to induce undercounts. Furthermore, it seems that once this threshold is reached, one out of every four additional envelopes will not be counted. Given the existence of a second, more covert, method of sabotage, the size of these effects comes as a surprise.

Here, too, the addition of controls for sex, international student status and risk preferences has little or no effect on these coefficients. In this case, however, neither sex nor risk preferences seem to matter. It also appears, however, that international students engage in much less output sabotage.

If expectations about teammates' behavior is included, the treatment effect for tournaments with sabotage is still significant, but much smaller in size. In particular, there is now little or no difference in the two tournament treatments, even if both produce more "misbehavior" than the combined piece rate treatments. Furthermore, the coefficient on the expectations term is negative and large, in both the statistical and economic senses: even for the tournament with sabotage, it is sufficient to eliminate output sabotage for all but the most extreme differences in output. Team members, it seems, practice a form of negative reciprocity: if A expects B to correctly report A's output, then A will correctly report B's, but if not, A will undercount B whenever B produces more than A. In tournaments where there is a pronounced incentive to undercount, few(er) teammates expect to have their output counted correctly, which leads them to undercount their teammates' output.

Surprisingly, perhaps, this result is robust with respect to not just the usual additional controls, but to the number of other subjects known.

Because quality sabotage is much less blatant, our prior was that it would have a lower threshold, and be more responsive to differences, than output sabotage, but otherwise exhibit the same broad patterns. The results in Table 4, which report random effects tobit estimates for four models obtained for the subsample of cases in which non-negative quality sabotage is observed, are consistent with this. In the simplest specification, which includes the two tournament treatment indicators and the difference in the target's objective quality of output and the subject's, all three coefficients are significant at the 10 percent level or better. To understand their implications, consider the implied sabotage functions, as depicted in Figure 5. Any positive difference is now sufficient to induce sabotage, even in the piece rate treatments, and sabotage in the first tournament treatment, in which peer evaluations have at most psychological consequences, exceeds this by a smallish but significant amount.

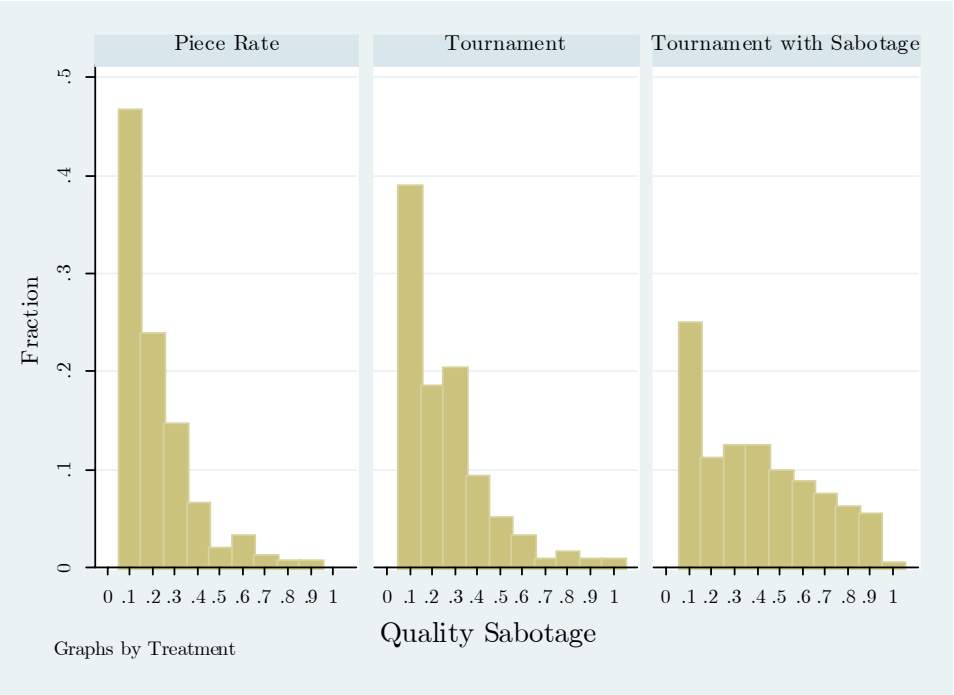

Figure 4

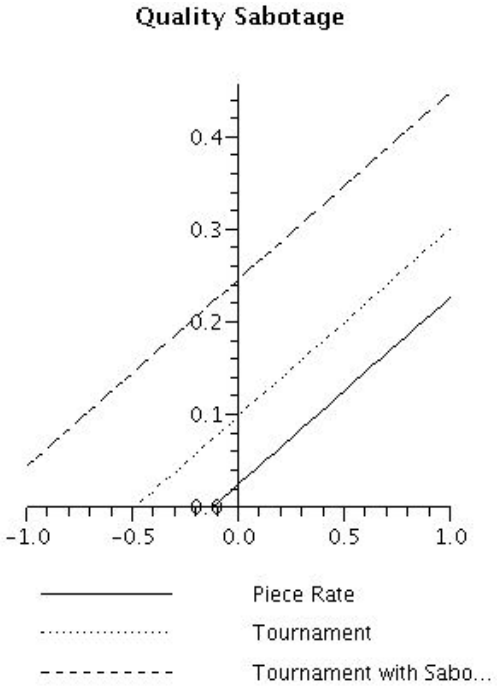

Figure 5

In tournaments with sabotage, on the other hand, underestimation is pervasive, even when the quality of the subject's output is much higher than the target's. As Figure 4 reveals, even when the two produce output of the same objective quality, the benchmark

model predicts that each will underestimate the quality of the other by about 0.25.

TABLE 4: Analysis of Quality Sabotage

Tournament	0.074*	0.071*	0.068	0.054
	(0.04)	(0.04)	(0.04)	(0.04)
Tournament with Sabotage	0.221***	0.225***	0.183***	0.184***
	(0.04)	(0.04)	(0.05)	(0.05)
(Target's Quality - Saboteur's Quality)	0.202***	0.189***	0.144*	0.135*
	(0.07)	(0.07)	(0.07)	(0.08)
Male		-0.037	-0.038	-0.028
		(0.04)	(0.04)	(0.04)
International Student		-0.018	-0.01	-0.024
		(0.06)	(0.06)	(0.06)
Risk Taker		0.123*	0.103	0.112*
		(0.06)	(0.06)	(0.06)
E(Teammates to correctly report my output)			-0.112*	-0.104
			(0.07)	(0.07)
Intercept	0.026	0.035	0.151**	-0.251
	(0.03)	(0.03)	(0.08)	(0.35)
Includes demographic controls	No	No	No	Yes
rho	0.407	0.393	0.373	0.35
Wald Chi-squared	37.098	42.709	47.476	54.172
N	627	627	627	622

Note: Tobit with lower limit of zero; 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.

When the model is expanded to include the now familiar sex, international student and risk preference indicators, there is no effect on either the size or the significance of the treatment or the quality difference coefficients, and only the coefficient on risk takers is significant at the 10 percent level or better.

Once more, it is the addition of the expectations term that produces the most dramatic effects. Its own large, significant effect suggests that negative reciprocity is not limited to quantity sabotage: if A expects B to underestimate the quality of A's work, then A is much more likely to underestimate the quality of B's work. The further observation that the tournament with sabotage coefficient is smaller in size, if not significance, tells us that few(er) teammates are expected to tell the truth when the stakes are increased.

Last, and as before, these results are (almost, in this case) robust with respect to the addition of the other demographic controls.

5 References

Bruggen and Strobel 2005

Bull, Schotter and Weigelt 1987

Chen 2003

Drago and Garvey 1998

Falk and Fehr 2003

Garicano and Palacios-Huerta 2006

Grund and Sliwka 2002

Harbring and Irlenbusch (2005, 2007a, 2007b)

Harbring, Irlenbusch, Krackel and Selten 2000

Lazear 1989

van Dijk, Sonnemans and van Winden 2001

Weber, Blais and Betz 2002