

Rational Self-Sabotage

Oliver Gürtler*and Johannes Münster†

November 1, 2010

Abstract

We consider a dynamic tournament where contestants choose productive effort and to help or sabotage their opponents. Because the current leader is sabotaged most strongly in the final stage, players help other players and even sabotage themselves in earlier stages.

Keywords: Tournament, sabotage, help, self-sabotage
JEL: C73, M52

1 Introduction

It is well known that contestants in tournaments do not have an incentive to help each other since relative performance is decisive. Contestants may even try to sabotage their opponents, i.e., to decrease their opponents' performances (see, e.g., Lazear 1989). Sabotage is usually modelled as a reduction in performance or output. Sabotage has, however, additional consequences that have not been taken into consideration in the economics literature so far. Often, contestants that get sabotaged suffer from considerable psychic costs. For instance, mobbing in the workplace seems to cause severe trauma for the victims, as many studies indicate (e.g., Groeblichhoff and Becker 1996, Leymann and Gustafsson 1996, Shallcross et al. 2008).

*University of Cologne, Albertus-Magnus Platz, D-50923 Cologne, Germany, phone: +49 221 4701450, fax: +49 221 4701816, e-mail: oliver.guertler@uni-koeln.de.

†Free University of Berlin, Boltzmannstrasse 20, 14195 Berlin, Germany, phone: +49 30 83855223, e-mail: johannes.muenster@fu-berlin.de

Typically, sabotage has been analyzed in static, one-period tournament models.¹ In practice, however, tournaments often last for some time and are thus dynamic in nature. The aim of the current note is to consider a dynamic tournament model with sabotage. We show that some insights from static models can be reversed in the dynamic case when the psychic costs of being sabotaged are taken into account.

In the final period of our model, we replicate a finding from static tournaments that leading players are sabotaged more strongly than trailing ones, as in Chen (2003) and Münster (2007). To be more precise, in any interior equilibrium of the final stage, sabotage efforts are chosen such that winning-probabilities of all contestants are equalized. In addition, effort costs in the final stage are independent of choices in previous stages. As a consequence, incentives to put forth productive effort or to sabotage the opponents are nonexistent in earlier stages. Moreover, if being sabotaged entails a psychic cost, contestants may even want to help their opponents in order to reduce the amount of sabotage directed against them in the final stage. Similarly, contestants are willing to incur a cost to *lower* their own performance - they engage in rational self-sabotage.

This note is related to Gürtler and Münster (2010), where we have shown that, in a dynamic tournament, the possibility of sabotage by rivals may decrease incentives to work productively in an early stage. However, in our previous model there are no psychic costs of being sabotaged so that players have no incentive to help each other or to self-sabotage, which is the key point of this note. Moreover, Gürtler and Münster (2010) assume that there are exactly three players who interact in two stages, and that the action spaces are discrete. This naturally raises the question of how general the results are. Here we demonstrate that similar results can be obtained in a model with $N \geq 3$ possibly heterogeneous players, $T \geq 2$ stages, and continuous strategy spaces. Self-sabotage also arises in Sappington and Weisman (2005), but for a different reason. They study a vertically-integrated producer operating under a parity requirement that requires it to offer the same services to its competitors as to its downstream affiliate, and show that the vertically-integrated producer can gain from raising costs or reducing quality if this harms the competitors more than it harms itself.

¹For example in Lazear (1989), Chen (2003), Münster (2007). Two notable exceptions are Amegashie and Runkel (2005) on sabotage in a sequential elimination tournament, and Ishida (2006) on the optimal design of sabotage-proof contracts.

2 Model

$N \geq 3$ players compete in a tournament with T stages. In stage $t = 1, \dots, T$, each player chooses effort $e_i^t \in \mathbb{R}$. Here $e_i^t > 0$ is interpreted as productive effort, $e_i^t < 0$ is self-sabotage. We think of a situation where a tournament is used as an incentive mechanism and where players would choose some regular effort normalized to zero in the absence of an explicit incentive scheme. In this interpretation, self-sabotage means that a player chooses a lower productive effort than without any explicit incentives. Similarly, in stage $t = 1, \dots, T$, each player chooses how much to help or sabotage each rival. Let $h_{ij}^t \in \mathbb{R}$ be the amount of help of i for $j \in \{1, \dots, n\} \setminus \{i\}$ in stage t . Again, $h_{ij}^t > 0$ is real help, whereas $h_{ij}^t < 0$ is sabotage.

The final output of i is determined by the sum of his efforts, the help received, some exogenous constants a_i^t and a noise term ε_i that is realized after the final stage of the game has been played:

$$q_i = \sum_{t=1}^T \left(a_i^t + e_i^t + \sum_{l \neq i} h_{li}^t \right) + \varepsilon_i$$

Here, the a_i^t can be interpreted as reflecting the abilities of the players. The noise terms ε_i are identically and independently distributed according to a continuously differentiable distribution function F . We assume that F has full support and is log-concave.² The highest output wins a winner prize $V > 0$, all others get nothing. Let p_i denote the probability that i wins. To ease notation, define

$$y_{ij} = \sum_{t=1}^T \left(a_i^t + e_i^t + \sum_{l \neq i} h_{li}^t \right) - \sum_{t=1}^T \left(a_j^t + e_j^t + \sum_{l \neq j} h_{lj}^t \right).$$

Thus y_{ij} is the difference in expected output between player i and j , given all decisions taken in the game. Then

$$\begin{aligned} p_i &= \Pr \{ q_i > q_j, \forall j \neq i \} \\ &= \Pr \{ y_{ij} + \varepsilon_i > \varepsilon_j, \forall j \neq i \} \\ &= \int_{-\infty}^{\infty} (\prod_{j \neq i} F(y_{ij} + \varepsilon_i)) F'(\varepsilon_i) d\varepsilon_i. \end{aligned} \tag{1}$$

²The assumption of log-concavity is fulfilled by many commonly studied parametric distribution functions, see Bagnoli and Bergstrom (2005).

The objective function of player i is

$$u_i = p_i V - \sum_{t=1}^T \left[C_{it} \left((e_i^t)^+, (e_i^t)^- \right) + H_{it} \left(\sum_{j \neq i} (h_{ij}^t)^+, \sum_{j \neq i} (h_{ij}^t)^- \right) + \psi_{it} \left(\sum_{j \neq i} (h_{ji}^t)^- \right) \right]$$

Here, $C_{it} : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ and $H_{it} : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ are cost functions, assumed to be strictly increasing in each argument and convex. The notation $(x)^+ = \max\{x, 0\}$ and $(x)^- = |\min\{x, 0\}|$ denotes the positive and negative part of x , respectively. Thus the first argument of C_{it} is productive effort, the second is self-sabotage. Similarly, the first argument of H_{it} is the total help chosen by i in t , whereas the second argument is the total sabotage chosen by i in t . Note that both effort and self-sabotage is costly;³ similarly, helping and sabotaging is costly. The potential psychic costs from being sabotaged are captured by ψ_{it} . We assume that ψ_{it} is increasing. We allow players to be heterogeneous and have different cost functions. In addition, we allow the cost functions to differ across stages; thereby we can also capture discounting.

Note that in the final stage T , no player will ever help another player or sabotage himself. In contrast, each player has an incentive to sabotage each rival, in order to increase his chances to win the tournament. To keep things as simple as possible, we maintain the following assumption throughout:

A1. For any decisions taken in earlier rounds that are not prohibitively costly, the final stage of the game has only interior equilibria⁴ where all players sabotage all rivals.

Assumption A1 implicitly limits the degree of heterogeneity between players, both concerning the exogenous differences in ability captured in the cost functions and in the production functions, and the endogenous differences due to the play in previous stages.⁵ It ensures that, in the final stage, sabotage is sufficiently easy and no player is beyond the reach of his competitors.

³For instance, self-sabotage could be costly, as the players may get bored if doing nothing at all.

⁴By "interior equilibria" we mean equilibria which do not involve "corner solutions" at $e_i^T = 0$ or $h_{ij}^T = 0$.

⁵As shown in Münster (2007), when players are very different, weaker players may not be sabotaged at all. Moreover, there has to be sufficiently much noise in order that a pure strategy equilibrium exists; this is a common feature of most models of tournaments. See Münster (2007), for a more detailed discussion of the assumption of interior equilibria.

3 Results

As in Münster (2007, Proposition 2), in any interior equilibrium of the last stage of the game, each player will win with the same probability. The reason behind this *prima facie* surprising result is straightforward. Suppose to the contrary that player i wins with a higher probability than j . Then player k should sabotage i more and reduce his sabotage against j by the same amount. Since the cost of sabotage depends only on the sum of sabotage activities, the costs of k are unchanged. However, k 's probability of winning is increased; thus the initial situation cannot have been an equilibrium. Therefore, in equilibrium all players win with the same probability.

It follows that, by gaining a lead in the previous stages, a player does not improve his probability of winning. Moreover, decisions taken in previous stages have no impact on the equilibrium effort and the total amount of sabotage chosen by a player in the final stage. To see this, consider the necessary first order conditions for the optimal decisions in stage T . As argued above, $e_i^T \geq 0$, and $h_{ij}^T \leq 0$ for all $j \neq i$. By A1, these inequalities are strict. Thus we have $(e_i^T)^+ = e_i^T > 0$ and $(e_i^T)^- = 0$, similarly $(h_{ij}^T)^+ = 0$ and $(h_{ij}^T)^- > 0$. In equilibrium, the following first order conditions have to hold:

$$\begin{aligned} \frac{\partial p_i}{\partial e_i^T} V &= \frac{\partial C_{iT}(e_i^T, 0)}{\partial (e_i^T)^+} \\ \frac{\partial p_i}{\partial (h_{ij}^T)^-} V &= \frac{\partial H_{iT}\left(0, \sum_{l \neq i} (h_{il}^T)^-\right)}{\partial (h_{ij}^T)^-}, \quad \forall j \in \{1, \dots, n\} \setminus \{i\} \end{aligned}$$

Using (1),

$$\begin{aligned} \frac{\partial p_i}{\partial e_i^T} &= \int_{-\infty}^{\infty} \left(\sum_{j \neq i} (F'(y_{ij} + \varepsilon_i) \Pi_{l \neq i, j} F(y_{il} + \varepsilon_i)) \right) F'(\varepsilon_i) d\varepsilon_i, \\ \frac{\partial p_i}{\partial (h_{ij}^T)^-} &= \int_{-\infty}^{\infty} F'(y_{ij} + \varepsilon_i) (\Pi_{l \neq i, j} F(y_{il} + \varepsilon_i)) F'(\varepsilon_i) d\varepsilon_i. \end{aligned}$$

Since every player wins with the same probability, $y_{il} = 0$ for all i and $l \neq i$.

It follows that

$$\begin{aligned}\frac{\partial p_i}{\partial e_i^T} &= (n-1) \int_{-\infty}^{\infty} F'(\varepsilon_i) (F(\varepsilon_i))^{n-2} F'(\varepsilon_i) d\varepsilon_i, \\ \frac{\partial p_i}{\partial (h_{ij}^T)^-} &= \int_{-\infty}^{\infty} F'(\varepsilon_i) (F(\varepsilon_i))^{n-2} F'(\varepsilon_i) d\varepsilon_i.\end{aligned}$$

Define

$$g := (n-1) \int_{-\infty}^{\infty} F'(\varepsilon_i) (F(\varepsilon_i))^{n-2} F'(\varepsilon_i) d\varepsilon_i.$$

and note that g depends only on the fundamentals of the model, and is independent of the endogenous variables. Equilibrium effort, and the sum of sabotage chosen by i , are determined by

$$\begin{aligned}gV &= \frac{\partial C_{iT}(e_i^T, 0)}{\partial (e_i^T)^+} \\ \frac{g}{n-1}V &= \frac{\partial H_{iT}(0, \sum_{l \neq i} (h_{il}^T)^-)}{\partial (h_{ij}^T)^-}, \quad \forall j \in \{1, \dots, n\} \setminus \{i\}\end{aligned}$$

Therefore, e_i^T and $\sum_{l \neq i} (h_{il}^T)^-$, and hence the costs C_{iT} and H_{iT} that i incurs in the final stage, are independent of the decisions taken in previous stages. Of course, the previous decisions will impact on which rival i sabotages how much; only the sum of sabotage activities chosen by i is independent of previous decisions.

We now turn to the decisions taken earlier in the game. Consider stage $T-1$. As we have just argued, the choices of player i neither change his probability of winning in the end, nor change the costs of effort and sabotage he will incur in the equilibrium of the final stage. It will, however, change the amount of sabotage received in the final round. This does not matter, however, if receiving sabotage does not bring a direct psychic cost. Then the only impact of i 's choices in $T-1$ on his payoff are through the cost C_{iT-1} and H_{iT-1} . Therefore it is optimal to choose zero effort and help. Proceeding with backward induction, the argument carries over to all previous stages.

Proposition 1 *Suppose there are no psychic costs of suffering from sabotage (i.e. $\psi_{it} \equiv 0$). In any subgame perfect Nash equilibrium,*

$$e_i^t = h_{ij}^t = 0$$

for all $t \in \{1, \dots, T-1\}$, all $i \in \{1, \dots, n\}$, and all $j \in \{1, \dots, n\} \setminus \{i\}$.

Proposition 1 shows that, when there is no psychic cost from being sabotaged, players choose zero effort, zero help, and zero sabotage in all stages except the final. Next, we consider how much sabotage is inflicted on player i in the final stage. Let

$$z_i = \sum_{t=1}^{T-1} \left(a_i^t + e_i^t + \sum_{j \neq i} h_{ji}^t \right).$$

In equilibrium of the final stage, all players win with the same probability; hence the expected output of all players is equalized. Thus for all $i, k \in \{1, \dots, n\}$,

$$z_i + a_i^T + e_i^T + \sum_{j \neq i} h_{ji}^T = z_k + a_k^T + e_k^T + \sum_{j \neq k} h_{jk}^T.$$

Summing these equations and rearranging, we obtain

$$\sum_{j \neq k} h_{jk}^T = -\frac{n-1}{n} (z_k + a_k^T + e_k^T) + \frac{1}{n} \sum_{i \neq k} (z_i + a_i^T + e_i^T) + \frac{1}{n} \sum_{i=1}^n \sum_{j \neq i} h_{ji}^T.$$

Since $h_{ij}^T \leq 0$ for all i and $j \neq i$, total sabotage against player k in stage T is

$$\begin{aligned} \sum_{j \neq k} (h_{jk}^T)^- &= \frac{n-1}{n} (z_k + a_k^T + e_k^T) - \frac{1}{n} \sum_{i \neq k} (z_i + a_i^T + e_i^T) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \sum_{j \neq i} (h_{ji}^T)^-. \end{aligned} \tag{2}$$

Proposition 2 *Sabotage received in T is strictly increasing in productive effort and sabotage against others in $T-1$, and strictly decreasing in self-sabotage and help for others in $T-1$.*

Proof. As argued above, the total amount of sabotage chosen by any player i in the final stage, $\sum_{j \neq i} (h_{ij}^T)^-$, does not depend on the previous history of the game. Since $\sum_{i=1}^n \sum_{j \neq i} (h_{ji}^T)^- = \sum_{i=1}^n \sum_{j \neq i} (h_{ij}^T)^-$, it follows that the third term on the right hand side of equation (2) is independent of actions taken in earlier stages. Moreover, e_i^T is also independent of the previous history of the game, and a_i^T is exogenous. Thus $\sum_{j \neq k} (h_{jk}^T)^-$ is strictly increasing in

z_k , and hence strictly increasing in e_k^{T-1} . Similarly, $\sum_{j \neq k} (h_{jk}^T)^-$ is strictly decreasing in z_i for $i \neq k$ and hence strictly decreasing in h_{ki}^{T-1} . ■

We now turn to the case where there is a direct psychic cost of being sabotaged. We focus on the semifinal stage, which is sufficient to make our main points. Clearly, players have an no incentive to work or sabotage others in $T - 1$. On the contrary, they have an incentive to help others and to sabotage themselves in $T - 1$, as this decreases the psychic costs suffered because of being sabotaged in the final stage. Whether players actually do sabotage themselves or help others depends, of course, on the cost functions for these activities as well. Proposition 3, which is immediate, states sufficient conditions.

Proposition 3 *Suppose that ψ_{iT} is strictly increasing. If*

$$\frac{\partial C_{iT-1}(0,0)}{\partial (e_{iT-1})^-} = 0$$

then player i sabotages himself in stage $T - 1$ in any subgame perfect Nash equilibrium. Similarly, if

$$\frac{\partial H_{iT-1}(0,0)}{\partial (h_{ij})^+} = 0,$$

player i helps other players in stage $T - 1$.

Hence in the semifinal stage, players may rationally help each other and sabotage themselves. Players leading in the tournament are more dangerous rivals and, thus, sabotaged more strongly in the final stage. If players suffer psychic costs from being sabotaged, they try to avoid being in front of their opponents. Thus, they have an incentive to help their opponents and to lower their own effort below the regular level normalized to zero, i.e. to self-sabotage.

References

- [1] Amegashie, J. A. and M. Runkel, 2005, "Sabotaging potential rivals", *Social Choice and Welfare*, 28, 143-162.

- [2] Bagnoli, M. and T. Bergstrom, 2005, “Log-concave probability and its applications”, *Economic Theory*, 26, 445-469.
- [3] Chen, K.-P., 2003, “Sabotage in Promotion Tournaments”, *Journal of Law, Economics and Organization*, 19, 119-140.
- [4] Groeblinghoff, D. and M. Becker, 1996, “A case study of mobbing and the clinical treatment of mobbing victims”, *European Journal of Work and Organizational Psychology*, 5, 277-294.
- [5] Gürtler, O. and J. Münster, 2010, “Sabotage in dynamic tournaments”, *Journal of Mathematical Economics*, 46, 179-190.
- [6] Ishida, J., 2006, “Dynamically Sabotage-Proof Tournaments”, *OSIPP Discussion Paper DP-2006-E-001*.
- [7] Lazear, E., 1989, “Pay Equality and Industrial Politics”, *Journal of Political Economy*, 97, 561-580.
- [8] Leymann, H. and A. Gustafsson, 1996, “Mobbing at work and the development of post traumatic stress disorders”, *European Journal of Work and Organizational Psychology*, 5, 251-275.
- [9] Münster, J., 2007, “Selection Tournaments, Sabotage, and Participation”, *Journal of Economics and Management Strategy*, 16, 943-970.
- [10] Sappington, D.E.M. and D.L. Weisman, 2005, “Self-sabotage”, *Journal of Regulatory Economics*, 27, 155-175.
- [11] Shallcross, L., Sheehan, M. and S. Ramsay, 2008, “Workplace Mobbing: Experiences in the public sector”, *International Journal of Organizational Behaviour*, 13, 56-70.