



Rational self-sabotage

Oliver Gürtler^{*}, Johannes Münster¹

University of Cologne, Albertus-Magnus-Platz, D-50923 Cologne, Germany

ARTICLE INFO

Article history:

Received 27 March 2011

Received in revised form

20 January 2012

Accepted 15 June 2012

Available online 20 August 2012

ABSTRACT

We consider a dynamic tournament where contestants choose a productive effort and to help or sabotage their opponents. Sabotage lowers the output of the victim. Moreover, sabotage imposes an additional direct psychic cost on the victim. Because the current leader is sabotaged most strongly in the final stage, players help other players and even sabotage themselves early on.

© 2012 Elsevier B.V. All rights reserved.

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 modeled 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 or bullying in the workplace seems to cause severe trauma for the victims, as many studies indicate (e.g., Groeblinghoff and Becker, 1996, Leymann and Gustafsson, 1996, Shallcross et al., 2008).²

Typically, sabotage has been analyzed in static, one-period tournament models.³ 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 earlier choices. As a consequence, incentives to put forth productive effort or to sabotage the opponents are nonexistent early on. 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.

These findings have implications for the design of incentive systems in firms. Often, inner-firm incentives are induced by promotion tournaments, in which several employees compete for a promotion to a higher layer in the firm's hierarchy. These promotion tournaments often last for a considerable period of time and are thus dynamic in nature. Moreover, employees are often able to observe each others' performances and, since they meet regularly, mobbing opportunities are ubiquitous. The note highlights the potential problems of these kinds of tournaments and indicates that employees may want to lower their own performance to avoid being mobbed by their colleagues. There are several ways in which own performance could be lowered. As e.g. argued by Holmstrom and Milgrom (1991) employees derive some pleasure from working up to some limit so that incentives are only required to induce work beyond that limit. Self-sabotage could then mean that employees decide to work less than the limit in spite of deriving pleasure from working more. Similarly, employees may intentionally make mistakes, miss meetings, or come late to the office.⁴

^{*} Corresponding author. Tel.: +49 221 4701450; fax: +49 221 4701816.

E-mail addresses: oliver.guertler@uni-koeln.de (O. Gürtler), johannes.muenster@uni-koeln.de (J. Münster).

¹ Tel.: +49 221 4704410; fax: +49 221 4704920.

² See Hogg et al. (2011) for a recent survey. They conclude that "exposure to bullying may have highly detrimental effects on the target's health and well-being" (p. 121).

³ For example in Lazear (1989), Chen (2003) and 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.

⁴ In Robert Graves' famous historical novel *I, Claudius*, self-sabotage is recommended by Pollio to Claudius, who goes on to become the emperor since he is the only male member of his family who survives:

'Do you want to live a long busy life, with honour at the end of it?'
'Yes'.

To prevent such nonproductive behavior, firms must take serious actions to tackle the sabotage or mobbing problem. A possible way to achieve this is to impose strong punishments for detected sabotage. Alternatively, the firm may try to make it more difficult for employees to sabotage each other, for instance, by organizing tournaments between employees who are locally separated.

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. Gürtler et al. (2011) point out that this effect hinges on the observability of decisions taken in an early stage, and present experimental evidence. However, in these papers 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. Another related paper is Amegashie and Runkel (forthcoming). They introduce the possibility of revenge into a conflict setting and find that players lower their conflict efforts, because the other player retaliates for high effort by increasing effort as well. The commonality between their model and the current one is that actions chosen early in the game that are essentially productive may induce harmful reactions by the other players later on. Whereas Amegashie and Runkel (forthcoming) introduce the concept of revenge to trigger such harmful reactions, in our model it is a rational response to focus most final-stage sabotage activities at the leading player. The two papers therefore consider different mechanisms by which certain actions are punished in later stages of the game. Self-sabotage also arises in Sappington and Weisman (2005), but for a very 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.

2. Model

$N \geq 3$ players compete in a tournament with 2 stages. In stage $t = 1, 2$, 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. As indicated before, we may 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. Similarly, in stage $t = 1, 2$, 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. Within a stage, all players choose their actions simultaneously. Players are able to observe the first-stage decisions of their rivals before the second stage begins.

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:

$$q_i = \sum_{t=1}^2 \left(a_i^t + e_i^t + \sum_{l \neq i} h_{il}^t \right) + \varepsilon_i.$$

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.⁵ 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

$$y_{ij} = \sum_{t=1}^2 \left(a_i^t + e_i^t + \sum_{l \neq i} h_{il}^t \right) - \sum_{t=1}^2 \left(a_j^t + e_j^t + \sum_{l \neq j} h_{jl}^t \right).$$

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

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

The objective function of player i is

$$\begin{aligned} u_i &= p_i V - \sum_{t=1}^2 \left[C_{it} \left((e_i^t)^+, (e_i^t)^- \right) \right. \\ &\quad \left. + H_{it} \left(\sum_{j \neq i} (h_{ij}^t)^+, \sum_{j \neq i} (h_{ij}^t)^- \right) + \psi_{it} \left(\sum_{j \neq i} (h_{ji}^t)^- \right) \right] \end{aligned}$$

$C_{it} : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ 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 costly⁶; similarly, helping and sabotaging is costly. We impose the following Inada condition that the first unit of effort, of self-sabotage, of help, and of sabotage has zero marginal cost:

$$\frac{\partial C_{it}(0, 0)}{\partial (e_i^t)^+} = \frac{\partial C_{it}(0, 0)}{\partial (e_i^t)^-} = \frac{\partial H_{it}(0, 0)}{\partial (h_{ij}^t)^+} = \frac{\partial H_{it}(0, 0)}{\partial (h_{ij}^t)^-} = 0 \quad (2)$$

for $t = 1, 2$. A simple specification that is nested in our model is when the cost of effort and self-sabotage depends only on its absolute value $|e_i|$, and the cost of help and sabotage depends only on the sum of absolute values $\sum_j |h_{ij}|$. However, this specification is restrictive since it implies a certain symmetry: for example, one unit of effort ($e_i = 1$) has the same cost as one unit of self-sabotage ($e_i = -1$), and one unit of help ($h_{ij}^t = 1$) has the same cost as one unit of sabotage ($h_{ij}^t = -1$). 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, 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:

⁵ The assumption of log-concavity is fulfilled by many commonly studied parametric distribution functions, see Bagnoli and Bergstrom (2005). Existence of pure strategy equilibria requires that there is sufficient noise in the tournament. This is a common feature of most models of tournaments. We assume throughout that this requirement is fulfilled.

⁶ For instance, self-sabotage could be costly, as the players may get bored if doing nothing at all.

⁷ Then exaggerate your limp, stammer deliberately, sham sickness frequently, let your wits wander, jerk your head, and twitch with your hands on all public or semi-public occasions. If you could see as much as I can see you would know that this was your only hope of safety and eventual glory.' Graves (1934–2006, cited from the Penguin Classics edition 2006), pp. 111–112.

A1. For any decisions taken in $t = 1$ that are not prohibitively costly, the final stage of the game has an interior equilibrium⁷ where all players sabotage all rivals. In the following, we confine attention to such an interior equilibrium of the final stage.

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 the previous stage.⁸ It ensures that, in the final stage, sabotage is sufficiently easy and no player is beyond the reach of his competitors. In terms of the fundamentals of the model, existence of an interior equilibrium is guaranteed whenever the players are symmetric ex-ante, the cost functions are sufficiently steep, and players discount future payoffs enough. Then discounting limits the incentives to work or help or sabotage in the first stage. Hence the asymmetries between the players in the final stage cannot become too big.

Note that the equilibrium of the subgame in stage 2 is not unique. The Inada condition (2) ensures that in any equilibrium of stage 2, every player i will choose a strictly positive effort $e_i^2 > 0$ and a strictly positive sum of sabotage activities $\sum_{j \neq i} (h_{ij}^2)^- > 0$. However, only the sum of the sabotage chosen, and the sum of sabotage received, is determined. To see this, note that all payoffs depend only on these sums. For this reason, there also exist equilibria with corner solutions where some players do not sabotage specific rivals at all; these equilibria have similar properties as the equilibria that we focus on.⁹

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 stage 1, a player does not improve his probability of winning. Moreover, decisions taken in stage 1 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 2. As argued above, $e_i^2 \geq 0$, and $h_{ij}^2 \leq 0$ for all $j \neq i$. By A1, these inequalities are strict. Thus we have $(e_i^2)^+ = e_i^2 > 0$ and $(e_i^2)^- = 0$, similarly $(h_{ij}^2)^+ = 0$ and $(h_{ij}^2)^- > 0$. In equilibrium, the following first order conditions have to hold:

$$\frac{\partial p_i}{\partial e_i^2} V = \frac{\partial C_{i2}(e_i^2, 0)}{\partial (e_i^2)^+}$$

$$\frac{\partial p_i}{\partial (h_{ij}^2)^-} V = \frac{\partial H_{i2}\left(0, \sum_{l \neq i} (h_{il}^2)^-\right)}{\partial (h_{ij}^2)^-}, \quad \forall j \in \{1, \dots, n\} \setminus \{i\}.$$

⁷ By “interior equilibrium” we mean an equilibrium which does not involve “corner solutions” at $e_i^2 = 0$ or $h_{ij}^2 = 0$.

⁸ As shown in Münster (2007), when players are very different, weaker players may not be sabotaged at all. See Münster (2007) for a more detailed discussion of the assumption of interior equilibria.

⁹ See Münster (2007, Appendix A.3).

Using (1),

$$\frac{\partial p_i}{\partial e_i^2} = \int_{-\infty}^{\infty} \left(\sum_{j \neq i} F'(y_{ij} + \varepsilon_i) \Pi_{l \neq i, j} F(y_{il} + \varepsilon_i) \right) F'(\varepsilon_i) d\varepsilon_i,$$

$$\frac{\partial p_i}{\partial (h_{ij}^2)^-} = \int_{-\infty}^{\infty} F'(y_{ij} + \varepsilon_i) (\Pi_{l \neq i, j} F(y_{il} + \varepsilon_i)) F'(\varepsilon_i) d\varepsilon_i.$$

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

$$\frac{\partial p_i}{\partial e_i^2} = (n-1) \int_{-\infty}^{\infty} F'(\varepsilon_i) (F(\varepsilon_i))^{n-2} F'(\varepsilon_i) d\varepsilon_i,$$

$$\frac{\partial p_i}{\partial (h_{ij}^2)^-} = \int_{-\infty}^{\infty} F'(\varepsilon_i) (F(\varepsilon_i))^{n-2} F'(\varepsilon_i) d\varepsilon_i.$$

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

$$gV = \frac{\partial C_{i2}(e_i^2, 0)}{\partial (e_i^2)^+}$$

$$\frac{g}{n-1} V = \frac{\partial H_{i2}\left(0, \sum_{l \neq i} (h_{il}^2)^-\right)}{\partial (h_{ij}^2)^-}, \quad \forall j \in \{1, \dots, n\} \setminus \{i\}.$$

Therefore, e_i^2 and $\sum_{l \neq i} (h_{il}^2)^-$, and hence the costs C_{i2} and H_{i2} that i incurs in the final stage, are independent of the decisions taken in stage 1. 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 in stage 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.

If receiving sabotage does not bring a direct psychic cost, this effect does not matter. Then the only impact of i 's choices in 1 on his payoff are through the cost C_{i1} and H_{i1} . Therefore it is optimal to choose zero effort, zero help, and zero sabotage in the first stage.¹⁰ Note that if there are $T > 2$ stages, the argument carries over to all previous stages by backward induction.

Next, we consider how much sabotage is inflicted on player i in the final stage. Let

$$z_i = a_i^1 + e_i^1 + \sum_{j \neq i} h_{ji}^1.$$

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 $i, k \in \{1, \dots, n\}$,

$$z_i + a_i^2 + e_i^2 + \sum_{j \neq i} h_{ji}^2 = z_k + a_k^2 + e_k^2 + \sum_{j \neq k} h_{jk}^2.$$

¹⁰ Recall that the regular effort level in the absence of an explicit incentive scheme has been normalized to zero. Hence, our results imply that players choose their regular level of effort in the first stage.

Summing these equations and rearranging, we obtain

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

Since $h_{ij}^2 \leq 0$ for all i and $j \neq i$, total sabotage against player k in stage 2 is

$$\sum_{j \neq k} (h_{jk}^2)^- = \frac{n-1}{n} (z_k + a_k^2 + e_k^2) - \frac{1}{n} \sum_{i \neq k} (z_i + a_i^2 + e_i^2) + \frac{1}{n} \sum_{i=1}^n \sum_{j \neq i} (h_{ji}^2)^-. \quad (3)$$

Proposition 1. *Under Assumption A1, sabotage received in the final stage $t = 2$ 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, $\sum_{j \neq i} (h_{ij}^2)^-$, does not depend on the previous history of the game. Since $\sum_{i=1}^n \sum_{j \neq i} (h_{ji}^2)^- = \sum_{i=1}^n \sum_{j \neq i} (h_{ij}^2)^-$, it follows that the third term on the right hand side of Eq. (3) is independent of actions taken in earlier stages. Moreover, e_i^2 is also independent of the previous history of the game, and a_i^2 is exogenous. Thus $\sum_{j \neq k} (h_{jk}^2)^-$ is strictly increasing in z_k , and hence strictly increasing in e_k^1 . Similarly, $\sum_{j \neq k} (h_{jk}^2)^-$ is strictly decreasing in z_i for $i \neq k$ and hence strictly decreasing in h_{ki}^1 . $\square$

We now turn to the case where there is a direct psychic cost of being sabotaged. Players have 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. The following proposition follows immediately from Proposition 1.

Proposition 2. *Suppose that ψ_{i2} is strictly increasing. Under Assumption A1, player i sabotages himself and helps other players in stage $t = 1$.*

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, in the first stage 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.

4. Concluding discussion

The current note has shown that the results on sabotage in static tournaments may be completely reversed in a dynamic setting. While in a static tournament players have an incentive to work productively and to sabotage each other, they may want to help

each other or even to self-sabotage in a dynamic tournament. By doing so, they reduce the amount of sabotage that is inflicted on them in the final tournament stage.

To ease the derivation of our results, we have introduced two simplifying assumptions. First, we have assumed that sabotage costs depend on total sabotage only. Second, we have assumed that in the final stage of the game an interior equilibrium exists for all histories of the game. If either of the two assumptions were violated, players' winning-probabilities were not equalized in the final stage of the game. Then the effects driving our main results would become weaker. In this sense, we have imposed assumptions that help to make the effects as strong as possible and thus to simplify the illustration of our findings.

The assumed information structure and the implied order of moves is important for our results as well. If players were not able to observe other players' decisions at the end of the first stage, the game would be strategically equivalent to a static game, and players would not choose to self-sabotage or to help other players.¹¹

Notice finally that the model results also depend on the assumption of a finite and commonly known number of stages. In an infinitely repeated game it is well known that many different action profiles can be implemented if the discount factor is sufficiently high. In particular, the players may coordinate on an equilibrium, in which they all choose zero effort and zero help, thereby minimizing their costs. The possibility of sabotage may in fact make it easier for players to sustain such an equilibrium, because sabotage could be used as a punishment for a deviation.

References

- Amegashie, J.A., Runkel, M., 2005. Sabotaging potential rivals. *Social Choice and Welfare* 28, 143–162.
- Amegashie, J.A., Runkel, M., 2012. The paradoxes of revenge in conflicts. *Journal of Conflict Resolution* 56, 313–330.
- Bagnoli, M., Bergstrom, T., 2005. Log-concave probability and its applications. *Economic Theory* 26, 445–469.
- Chen, K.-P., 2003. Sabotage in promotion tournaments. *Journal of Law, Economics and Organization* 19, 119–140.
- Graves, R., 1934–2006. I. Claudius, Penguin Classics Edition.
- Groebelinghoff, D., Becker, M., 1996. A case study of mobbing and the clinical treatment of mobbing victims. *European Journal of Work and Organizational Psychology* 5, 277–294.
- Gürtler, O., Münster, J., 2010. Sabotage in dynamic tournaments. *Journal of Mathematical Economics* 46, 179–190.
- Gürtler, O., Münster, J., Nieken, P., 2011. Information policy in tournaments with sabotage. Discussion Paper.
- Hogh, A., Mikkelsen, E.G., Hansen, A.M., 2011. Individual consequences of workplace bullying/mobbing. In: Einarsen, S., Hoel, H., Zapf, D., Cooper, C.L. (Eds.), *Bullying and Harassment in the Workplace: Developments in Theory, Research, and Practice*, second ed. CRC Press Taylor & Francis Group, pp. 107–128.
- Holmstrom, B., Milgrom, P., 1991. Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics and Organization* 7, 24–52.
- Ishida, J., 2006. Dynamically sabotage-proof tournaments. OSIPP Discussion Paper DP-2006-E-001.
- Lazear, E., 1989. Pay equality and industrial politics. *Journal of Political Economy* 97, 561–580.
- Leymann, H., Gustafsson, A., 1996. Mobbing at work and the development of post traumatic stress disorders. *European Journal of Work and Organizational Psychology* 5, 251–275.
- Münster, J., 2007. Selection tournaments, sabotage, and participation. *Journal of Economics and Management Strategy* 16, 943–970.
- Sappington, D.E.M., Weisman, D.L., 2005. Self-sabotage. *Journal of Regulatory Economics* 27, 155–175.
- Shallcross, L., Sheehan, M., Ramsay, S., 2008. Workplace mobbing: experiences in the public sector. *International Journal of Organisational Behaviour* 13, 56–70.

¹¹ See Gürtler et al. (2011).