Rational Sabotage in Cooperative Production*

Carmen Beviá†

Universitat Autònoma de Barcelona-CODE

Luis C. Corchón

Departamento de Economía. Universidad Carlos III de Madrid

November, 2004.

Abstract: We consider a model of cooperative production in which rational agents have the possibility of carrying out sabotage activities that decrease output. It is shown that sabotage depends on the interplay between returns to scale, the technology of sabotage, the number of agents, the degree of substitutability and the degree of meritocracy. In particular it is shown that, ceteris paribus, meritocratic systems give more incentives to sabotage than egalitarian systems. We address two questions: The degree of meritocracy that is compatible with absence of sabotage and the existence of a Nash equilibrium without sabotage.

Key words: Cooperative production, sharing rules, sabotage.

JEL Classification Number(s): D20, D72, D78, J54.

---

* We thank Paco Alcala, Clara Eugenia Garcia, Shasikanta Nandeibam, Ines Macho-Stadler, David Perez-Castrillo, Clara Ponsati an associate editor and two anonymous referees for their helpful comments. The first author acknowledges financial support from BEC2002-02130, HI2001-0039, 2000SGR-00054, and the support of the Barcelona Economics program (CREA). The second author acknowledges financial support from BEC2002-02194.

† Address for correspondence: Carmen Beviá. Universitat Autonoma de Barcelona. Edifici B. 08193. Bellaterra, Barcelona. Spain. e-mail: Carmen.Bevia@uab.es phone: 935812190

Rational Sabotage in Cooperative Production

1. Introduction

Economics is the study of resource allocation under certain constraints. Historically, the first constraint considered was feasibility, i.e. the consumption of a good can not exceed its available quantity. As economics developed, other constraints were considered, for instance Incentive Constraints: Any economic system must not give incentives to agents to transmit wrong information -adverse selection- or to take socially unwanted actions -moral hazard.

In this paper we want to broaden the scope of incentive theory by considering situations in which agents can sabotage production by destroying other people´s inputs. Thus, the input supplied by an agent reflects her effort and the sabotage done by others on this agent. We assume that sabotage is undetectable because the effort and the technology of sabotage are not contractible. On the contrary, inputs are contractible. Traditionally, sabotage has been associated with capitalism, at least from the days of the Luddite revolt. In this paper we show that sabotage might arise in the framework of cooperative production, in which the output is entirely distributed to workers.[1] An example of how sabotage may arise as a rational action follows.

Two people are collecting grapes. Andy collects white grapes -whose quantity is denoted by R_1- and Beth collects red grapes, whose quantity is denoted by R_2. These grapes are transformed in wine -denoted by $Y-$ according to the production function $Y = (R_1 + R_2)^{1/2}$. The consumption of wine allocated to each worker -C_1 and C_2 respectively- is determined by the *Proportional Sharing*

[1] We do not deny that sabotage is, sometimes, an irrational action, taken for revenge, etc. Our purpose here is to show that even if such feelings do not exist, there is room for sabotage.

Rule, i.e.

$$C_i = \frac{R_i}{R_1 + R_2}(R_1 + R_2)^{1/2},\ i = 1, 2.$$

For future reference we notice that this sharing rule is meritocratic, in the sense that it allocates wine depending on relative inputs. Suppose that when the working day is about to finish, $R_1 = R_2 = 50$. Thus, $Y = 10$, $C_1 = C_2 = 5$. Now an unexpected event forces Beth to leave. Choices for Andy are to remain faithfully devoted to his own work, in which case he would obtain 21 extra units of grape or to destroy the crop assembled by Beth and pretend that somebody stole it.[2] In the first case his consumption of wine -the only thing he cares about- is

$$C_1 = \frac{71}{121}(121)^{1/2} \simeq 6,45.$$

In the second case, Andy's consumption of wine is

$$C_1 = \frac{50}{50}(50)^{1/2} \simeq 7,07.$$

Therefore, if Andy is rational, he will destroy Beth's crop. Suppose now that the sharing rule is *Egalitarian*, i.e.

$$C_i = \frac{(R_1 + R_2)^{1/2}}{2},\ i = 1, 2.$$

We notice that this rule is not meritocratic at all, in the sense that it allocates wine irrespectively of relative inputs. In this case, faithful work yields to Andy $C_1 = 5.5$ and sabotage $C_1 = 3.5$, i.e. sabotage is not a rational action.

What is going on in these examples? When an agent decides to sabotage another agent's crop, there are two effects. On the one hand output falls reflecting the fall in the quantity of input supplied by both the saboteur and the agent who

[2]The fact that Andy's output still 50 at the end of the day can be explained by saying that he spent the remaining labor time chasing the thief.

has been sabotaged. This is bad from the saboteur's point of view because there is less to be distributed. The fall in output depends on the returns embodied in the production function; the more "curvature" of the production function, the less fall in output. Since this curvature reflects the degree of congestion we may say that the fall in output depends inversely on the degree of congestion. On the other hand, the relative ranking of the saboteur rises and this is good for him. The importance of the second effect depends on how meritocratic the sharing rule is; for instance in the egalitarian sharing rule this effect does not exist. When the rule is meritocratic and there is congestion, the second effect may dominate and sabotage is a rational action, as in the case of the proportional sharing rule above.[3] The relationship between congestion, meritocracy and sabotage will be a central theme throughout our paper.

The model is presented in Section 2. In order to make the model tractable we make a number of simplifications. First we assume that the total quantity of labor supplied by each agent is fixed. Thus, labor can be spent on the production of an intermediate input (by exerting effort) or on the destruction of the inputs of other agents (sabotage). This assumption is made in order to focus attention on the choice between productive and sabotage activities. It is appropriated when length of working time is fixed exogenously by law, custom, etc. Second, we assume that the production function is symmetric and the sharing rules satisfy an anonymity condition.

[3] A related example may help to further understand the situation. Suppose that some runners have to compete in a, say, 1.500 meters run. Runners can devote their energies either to run or to step into other people shoes, i.e. to sabotage other runners. One would expect that, in a given moment of time, the amount of sabotage depends on how runners are rewarded -the more meritocratic the reward, the more sabotage- and the degree of congestion i.e. if runners form a compact group or they are scattered, etc.

With these assumptions in hand we study the existence and properties of Nash equilibria. In Section 3 we present a necessary condition for sabotage not to arise in a Nash equilibrium that says that either the possibilities of destruction are small relative to the number of agents, or if this is not the case, the degree of meritocracy should be bounded. This bound depends on the degree of congestion, the number of agents and the damage caused by sabotage. In contrast, complementarity or substitutability among inputs does not have a role here, at least if the production function is of CES type and sharing rules are suitably parametrized.

In Section 4 we prove that, under some conditions there is a Nash equilibrium with zero sabotage.[4] As it happens with the necessary condition, we have two cases. When the possibilities of destruction are small our conditions are mild and include most sharing rules that have been considered in the literature. Again, complementarity or substitutability among inputs does not play any role here. However, when the possibilities of destruction are not small we need an extra condition that is stronger -but it has the same flavor- than the necessary condition. In this case the relationship among inputs plays a role. Finally, Section 5 concludes and suggest some possibilities of research.

Let us now comment on other papers dealing with similar issues. In the case of a capitalistic firm, Lazear (1989) was the first to point out that, in the presence of sabotage, large differences in salaries become dysfunctional. In his model, agents are paid according to the position achieved in a contest.[5] In a model of

[4]In our model, since labor supply is fixed, zero sabotage implies efficiency. This contrast with the results obtained by Holmstrom (1982). There are several differences between our model and Holmstrom's, the most important being that in our model agents produce an intermediate input that is contractible. As has been shown by Nandeibam (2002) in a model where sabotage is not possible, efficiency might arise in equilibrium if there are contractible intermediate inputs. Thus, our results and Nandeibam's show the importance of contractible intermediate inputs.

[5]Other types of dysfunctional behavior when only some aspects of performance are rewarded

Rent-Seeking, Konrad (2000) considered that the effort of an agent reduces rival´s performance by sabotaging her activities. He shows that, in equilibrium, sabotage disappears if the number of agents is sufficiently large. In all these models agents compete directly so the existence of sabotage is quite natural. It is not that clear that sabotage may arise in a model of cooperative production which is the focus of our analysis.[6]

Summing up, our paper finds that the possibility of sabotage depends on the following factors: (1) The degree of meritocracy (as in capitalistic firms), (2) the number of agents (as in rent-seeking contests), (3) the degree of congestion, (4) the relationship among inputs and (5) the technology of sabotage. Thus, our analysis of the necessary and sufficient conditions for absence of sabotage produces a picture that is more complex than the one we had before.

2. The Model

The setting is one of cooperative production, see, e.g., Sen (1986), Fabella (1988) and Roemer and Silvestre (1993) for examples and applications. There are n agents. The input provided by agent i is denoted by $R_i \in \mathbb{R}_+$. Let Y be the total output. The production function is written $Y = f(R_1, .., R_n)$, where $f : \mathbb{R}^n_+ \to \mathbb{R}_+$ is C^1, concave, increasing in all its arguments and symmetric. Total output is shared among agents by means of a sharing rule, i.e. a list of functions $s_i : \mathbb{R}^n_+ \to [0, 1]$, $i = 1, ..., n$ such that,

$$\sum_{i=1}^{n} s_i(R_1, .., R_n) = 1 \text{ for all } (R_1, .., R_n) \in \mathbb{R}^n_+$$

were studied by Holmstrom and Milgrom (1991) in a Principal- Multi Task-Agent model. This literature is surveyed in Gibbons (1998) and Prendergast (1999).

[6] Itoh (1991) and Macho-Stadler and Pérez-Castrillo (1993) have analyzed the polar case where cooperation among agents is possible.

We will assume that $s_i(\ )$ is C^1, non decreasing on R_i, $s_i(R_1, .., R_n) > 0$ if $R_i > 0$, and that sharing rules are anonymous in the following sense:

$$\begin{aligned} \{R_1 &= ... = R_n = R\} \Rightarrow \{s_i(R, .., R) = s_j(R, .., R) \text{ and} \\ \frac{\partial s_i(R, .., R)}{\partial R_j} &= \frac{\partial s_j(R, .., R)}{\partial R_i} \ \forall i, j \text{ with } i \neq j\}. \end{aligned}$$

This assumption holds if, for example, $s_i(R_1, .., R_n) = s(R_i, \sum_{k=1}^{n} g(R_k))$, i.e. if the share of i is a function independent of i, with arguments the input of i and the sum of a function of all inputs. An example of a class of sharing rules fulfilling these conditions is:

$$s_i(R_1, .., R_n) = (\frac{\alpha R_i}{\sum_{k=1}^{n} R_k} + \frac{1-\alpha}{n}), \alpha \in [0, 1], i = 1, ..., n.$$

This class of sharing rules is parametrized by α. If $\alpha = 0$ we get the egalitarian sharing rule and if $\alpha = 1$, we have the proportional sharing rule. The parameter α is a measure of how relative effort is valued and thus measures the degree of meritocracy. The interested reader can find in Moulin (1987) and Pfingsten (1991) other examples of sharing rules fulfilling our conditions.

Agents care only about their own consumption. As we remarked in the Introduction, the quantity of labor time is fixed. An agent, say i, can divide her working time, denoted by T, between productive labor, denoted by l_i^P and sabotage activities. Let l_{ij} be the quantity of labor allocated by i to sabotage the input of agent j. The time constraint reads, $T = l_i^P + \sum_{j \neq i} l_{ij}$. The input provided by agent i depends on her own productive effort, l_i^P, and the amount of time devoted by the remaining agents to sabotaging the input of i, i.e.

$$R_i = R(l_i^P, l_{1i}, ...l_{(i-1)i}, l_{(i+1)i}, ...l_{ni})$$

where $R(\ )$ is a $\mathcal{C}^1$ function such that

$$\frac{\partial R}{\partial l_i^P} > 0 \text{ and } \frac{\partial R}{\partial l_{ji}} < 0.$$

Given these elements, we describe the *sabotage game* as follows: for each agent i, a strategy is the time devoted to sabotage activities, i.e. the vector $l_i = (l_{i1}, l_{i2}, l_{i(i-1)}, l_{i(i+1)}, l_{in})$. Time devoted to productive activities is determined by the constraint $l_i^P = T - \sum_{j \neq i} l_{ij}$. By l_{-i} we denote the vector $(l_1, ..., l_{i-1}, l_{i+1}, .., l_n)$. For each agent i, given a vector of strategies (l_i, l_{-i}), the payoff function is given by

$$\begin{aligned} \pi_i(l_i, l_{-i}) &\equiv s_i(R_1(l_i, l_{-i}), ..., R_n(l_i, l_{-i}))f(R_1(l_i, l_{-i}), ..., R_n(l_i, l_{-i})) \text{ where} \\ R_j(l_i, l_{-i}) &\equiv R(T - \sum_{j \neq i} l_{ji}, l_{1j}, ...l_{(j-1)j}, l_{(j+1)j}, ...l_{nj}),\ j = 1, ..., n. \end{aligned}$$

A *Nash equilibrium of the sabotage game,* denoted by *NE,* is a vector of strategies $(l_1, .., l_n)$ such that for all agent i, $\pi_i(l_i, l_{-i}) \geq \pi_i(l_i', l_{-i})$ for all l_i'.

We postpone until Section 4 the problem of the existence of a NE. In Section 3 below we concentrate on the implications of guaranteeing that no agent has incentives to engage in sabotage.

3. A Necessary Condition for No Sabotage

If all working time is devoted to productive activities, $(l_i, l_{-i}) = (0, 0)$. Define $R_j^0 \equiv R_j(0, 0)$ for agent j, and let R^0 denote the vector of inputs evaluated at the point of zero sabotage, that is, $R^0 = (R_1^0, ..., R_n^0)$. If no agent has incentive to sabotage when all other agents do not sabotage, it must be that $\forall i, j$, $\frac{\partial \pi_i(0,0)}{\partial l_{ij}} \leq 0$, where

$$\begin{aligned} \frac{\partial \pi_i(0,0)}{\partial l_{ij}} &= f(R^0)(\frac{\partial s_i(R^0)}{\partial R_j}\frac{\partial R_j(0,0)}{\partial l_{ij}} - \frac{\partial s_i(R^0)}{\partial R_i}\frac{\partial R_i(0,0)}{\partial l_i^P}) \\ &\quad + s_i(R^0)(\frac{\partial f(R^0)}{\partial R_j}\frac{\partial R_j(0,0)}{\partial l_{ij}} - \frac{\partial f(R^0)}{\partial R_i}\frac{\partial R_i(0,0)}{\partial l_i^P}) \end{aligned}$$

If this requirement is not fulfilled, all NE imply sabotage. Let

$$M \equiv -\frac{\frac{\partial R_j(0,0)}{\partial l_{ij}}}{\frac{\partial R_i(0,0)}{\partial l_i^P}}$$

M is a measure of the relative impact of change in inputs induce by an allocation of labor time of i from productive activities to sabotage activities evaluated at the point of zero sabotage. Thus, M is a measure of the power of destruction versus production capabilities. Abusing language, we will say that M is a measure of the possibilities of destruction. From our assumptions it follows that $M > 0$. Furthermore, because our symmetry assumption on $R(\ )$, M is independent of i and j.

Since the production function f is symmetric, $\frac{\partial f(R^0)}{\partial R_j} = \frac{\partial f(R^0)}{\partial R_i}$. Using the definition of M and dividing by $f(R^0)$, the necessary condition reads

$$-\frac{\partial s_i(R^0)}{\partial R_j}M - \frac{\partial s_i(R^0)}{\partial R_i} + s_i(R^0)\frac{\partial f(R^0)}{\partial R_i}\frac{1}{f(R^0)}(-M-1) \leq 0.$$

Differentiating $\sum_{i=1}^n s_i(R_1, ..., R_n) = 1$ and using our anonymity assumption,

$$\frac{\partial s_i(R^0)}{\partial R_i} = -(n-1)\frac{\partial s_j(R^0)}{\partial R_i} = -(n-1)\frac{\partial s_i(R^0)}{\partial R_j}. \text{ Thus,}$$

$$(\frac{M-n+1}{n-1})\frac{\partial s_i(R^0)}{\partial R_i}\frac{1}{s_i(R^0)} \leq \frac{\partial f(R^0)}{\partial R_i}\frac{1}{f(R^0)}(M+1).$$

Multiplying by R_i^0, the above inequality implies that

$$(\frac{M-n+1}{n-1})\frac{\partial s_i(R^0)}{\partial R_i}\frac{R_i^0}{s_i(R^0)} \leq \frac{\partial f(R^0)}{\partial R_i}\frac{R_i^0}{f(R^0)}(M+1). \tag{3.1}$$

Now we have two cases. If $M \leq n-1$, the inequality (3.1) always holds. In words, if M is small in relationship with the number of agents, the necessary condition always holds. However, if $M > n-1$,

$$\frac{\partial s_i(R^0)}{\partial R_i}\frac{R_i^0}{s_i(R^0)} \leq \frac{\partial f(R^0)}{\partial R_i}\frac{R_i^0}{f(R^0)}\left(\frac{(M+1)(n-1)}{M+1-n}\right) \tag{3.2}$$

Summarizing, if zero sabotage is a Nash equilibrium of the sabotage game, then either

1. The possibilities of destruction are small relative to the number of agents, $M \leq n-1$, or

2. $M > n-1$. In this case, in the point of zero sabotage, the elasticity of the share with respect to the input of agent i is bounded by the elasticity of the production function with respect to the input of agent i, multiplied by a certain factor. These elasticities can be interpreted respectively as the degree of meritocracy -because it measures how one' share responds to one's efforts- and the degree of congestion -because it measures how output responds to variations in inputs. The factor depends on the possibilities of destruction and the number of agents (inequality 3.2).

Example 1. *Suppose a constant elasticity of substitution production function* $Y = (\sum (R_i)^{\rho})^{\frac{r}{\rho}}$ *with* $\rho \leq 1$ *and* $r \leq 1$, *i.e. non increasing returns to scale. Let* $s_i(R_1, .., R_n) = (\frac{\alpha R_i}{\sum_{k=1}^{n} R_k} + \frac{1-\alpha}{n})$. *If* $M > n-1$, *the necessary condition for no sabotage reads:*

$$\alpha \leq r \left[\frac{M+1}{M+1-n} \right]$$

Notice that the degree of substitutability, ρ, *does not play any role in the above condition. This equation implies that the degree of congestion* ($\frac{1}{r}$), *the number of agents and the possibilities of destruction determined the upper bound on the degree of meritocracy* (α), *if sabotage is to be avoided. If this bound is not respected, there is sabotage in all Nash equilibria.*

4. A Sufficient Condition for No Sabotage.

In this section we study under what conditions a Nash equilibrium with no sabotage exists. To make the problem tractable, we assume that the individual's input is given by

$$R_i(l_i, l_{-i}) = \max(T - \sum_{j \neq i} l_{ij} - K \sum_{j \neq i} l_{ji}, 0),\ i \in \{1, .., n\}.$$

Where the parameter K is a positive constant. Notice that in this case, if agent i uses one unit of labor time to sabotage the input of agent j, he reduces his input in one unit and the input of agent j in K units. Thus,

$$-\frac{\frac{\partial R_j(l_i, l_{-i})}{\partial l_{ij}}}{\frac{\partial R_i(l_i, l_{-i})}{\partial l_i^P}} = K \text{ for all } (l_i, l_{-i}).$$

Let us add the following assumptions:

$A1$. For all $i \in \{1, .., n\}$, $s_i(R_1, .., R_n) = s(R_i, \sum_{k=1}^n R_k)$.

$A2$. Let $x_i = R_i$ and $y = \sum_{k=1}^n R_k$. Then $\frac{\partial s(x_i, y)}{\partial y} \leq 0$, and $\frac{\partial^2 s(x_i, y)}{\partial x_i \partial y} \leq 0$.

Assumption $A1$ says that the share allocated to i depends on the input supplied by i and the sum of inputs supplied by all agents. It is a special case of the anonymity assumption introduced in Section 2. The first part of $A2$ just means that the share of agent i is decreasing with the input supply by agent j. The second part of $A2$ says that the increase of the share with one's input decreases with total input. This is just a technical assumption. Assumptions $A1$ and $A2$ are satisfied by our example of a sharing rule.[7]

The following Lemma says that, under $A1$ we can restrict our attention to symmetric best responses.

[7] See Beviá and Corchón (2003) for more examples of sharing rules satisfying these assumptions.

Lemma 1. *Under Assumption* $A1$, *if* $l_i = (l_{i1}, ..., l_{ii-1}, l_{ii+1}, .., l_{in})$ *is a best response for agent* i *to* $l_{-i} = 0$, *then* $\hat{l}_i = (l, ..., l)$ *with* $l = \frac{\sum_{j=1, j\neq i}^{n} l_{ij}}{n-1}$ *is also a best response for agent* i *to* $l_{-i} = 0$.

Proof. Notice first that if l_i is a best response to $l_{-i} = 0$, then $T - Kl_{ij} \geq 0$ for all j. Suppose not, that is, suppose that there is an agent j such that $T - Kl_{ij} < 0$. Then agent i can decrease the time dedicated to sabotaging agent j up to a point such that $T - Kl'_{ij} = 0$. Thus agent i will increase her input without affecting the input of the other agents, which implies that she will be better off and will contradict that l_i is a best response against $l_{-i} = 0$.

Since $T - Kl_{ij} \geq 0$ for all j, the total input of all agents but i is: $\sum_{j\neq i} R_j(l_i, l_{-i}) = (n-1)T - K \sum_{j=1, j\neq i}^{n} l_{ij}$. Notice also that $T - Kl \geq 0$, otherwise for the agent k such that $l_{ik} = \max_j l_{ij}$, $T - Kl_{ik} < 0$. Then, the input of all agents but i under $\hat{l}_i$ is $R_j(\hat{l}_i, l_{-i}) = T - Kl$, and $\sum_{j\neq i} R_j(\hat{l}_i, l_{-i}) = (n-1)T - K(n-1)l = \sum_{j\neq i} R_j(l_i, l_{-i})$. Furthermore, $R_i(l_i, l_{-i}) = R_i(\hat{l}_i, l_{-i})$. Since by $A1$ the share of agent i only depends on his input an the sum of total inputs, $s_i(l_i, l_{-i}) = s_i(\hat{l}_i, l_{-i})$. Finally, notice that for all $j \neq i$, $R_j(\hat{l}_i, l_{-i}) = \frac{1}{(n-1)} \sum_{j\neq i} R_j(l_i, l_{-i})$, so the vector $(R_1(\hat{l}_i, l_{-i}), .., R_n(\hat{l}_i, l_{-i}))$ is just a convex combination of $(n-1)$ vectors that are permutations of the vector $(R_1(l_i, l_{-i}), .., R_n(l_i, l_{-i}))$ and where in each of this permutations the image of i is i. Symmetry and concavity of f implies that $f(R_1(\hat{l}_i, l_{-i}), .., R_n(\hat{l}_i, l_{-i})) \geq f(R_1(l_i, l_{-i}), .., R_n(l_i, l_{-i}))$, which implies that $\pi_i(l_i, l_{-i}) \leq \pi_i(\hat{l}_i, l_{-i})$. ■

In the last section we proved that if $M \leq n-1$, the necessary condition for no sabotage holds. Notice that, given our assumption on the individual's input, $M = K$. The next Proposition shows that if $K \leq n-1$ and $A1$ and $A2$ hold, zero sabotage is a Nash equilibrium. The intuition is that since the damage that agents can inflict on each other is small, sabotage does not pay off.

Proposition 1. *Assume A1, A2, and $K \leq n-1$. Then $l_{ij} = 0$ for all $i, j \in \{1, .., n\}$ is a Nash equilibrium.*

Proof. Let us see that if $l_{-i} = 0$, the best response for agent i is $l_i = 0$. Because of the previous Lemma, if we prove that for all $j \neq i$, $\frac{\partial \pi_i(l_i,0)}{\partial l_{ij}} \leq 0$ for all $l_i = (l, ..., l)$, then $l_i = 0$ is a best response for this agent. Let $R_j^l = R_j(l_i, 0) = T - Kl$, $R_i^l = R_i(l_i, 0) = T - (n-1)l$, and let $R^l = (R_1^l, .., R_n^l)$. To see that $\frac{\partial \pi_i(l_i,0)}{\partial l_{ij}} \leq 0$, it is enough to show that

$$\frac{\partial s_i(R^l)}{\partial R_j} \frac{\partial R_j(l_i,0)}{\partial l_{ij}} + \frac{\partial s_i(R^l)}{\partial R_i} \frac{\partial R_i(l_i,0)}{\partial l_{ij}} \leq 0. \tag{4.1}$$

Since $\frac{\partial R_j(l_i,0)}{\partial l_{ij}} = -K$, and $\frac{\partial R_i(l_i,0)}{\partial l_{ij}} = -1$, we have to prove that

$$\frac{\partial s_i(R^l)}{\partial R_i} + K \frac{\partial s_i(R^l)}{\partial R_j} \geq 0.$$

Notice that for all vectors $l_i = (l, .., l)$ and because $K \leq n-1$, $R_i(l_i, 0) \leq R_j(l_i, 0)$ for all $j \neq i$. By Assumption $A2$,

$$\frac{\partial s_i(R^l)}{\partial R_j} = \frac{\partial s_i(R_i^l, \sum_{j=1}^n R_j^l)}{\partial R_j} = \frac{\partial s(R_i^l, y^l)}{\partial y},$$

where $y^l = \sum_{j=1}^n R_j^l$. Furthermore,

$$\frac{\partial s(R_i^l, y^l)}{\partial y} \geq \frac{\partial s(R_j^l, y^l)}{\partial y} = \frac{\partial s_j(R_j^l, \sum_{j=1}^n R_j^l)}{\partial R_i} = \frac{\partial s_j(R^l)}{\partial R_i}.$$

Thus,

$$\frac{\partial s_i(R^l)}{\partial R_i} + K \frac{\partial s_i(R^l)}{\partial R_j} \geq \frac{\partial s_i(R^l)}{\partial R_i} + K \frac{\partial s_j(R^l)}{\partial R_i}.$$

Since $\sum_{k=1}^n s_k = 1$, then $\sum_{k=1}^n \frac{\partial s_k(R^l)}{\partial R_i} = 0$. Furthermore, we know that $\frac{\partial s_k(R^l)}{\partial R_i} = \frac{\partial s_j(R^l)}{\partial R_i}$ for all $j, k \neq i$. Therefore,

$$\frac{\partial s_i(R^l)}{\partial R_i} + (n-1) \frac{\partial s_j(R^l)}{\partial R_i} = 0.$$

Since $K \leq n-1$,

$$\frac{\partial s_i(R^l)}{\partial R_i} + K\frac{\partial s_j(R^l)}{\partial R_i} \geq \frac{\partial s_i(R^l)}{\partial R_i} + (n-1)\frac{\partial s_j(R^l)}{\partial R_i} = 0,$$

as we wanted to prove. ■

The previous equilibrium is not always unique. If $n-1 \geq K \geq 1$, there are, at least, two kinds of Nash equilibria additional to that with no sabotage:

I) In the first kind, no positive output is produced. This equilibrium can be sustained with the following strategies: For each i, $l_{ii+1} = T$ (modulo n) and $l_{ij} = 0$ otherwise. Clearly $R_i(l_i, l_{-i}) = 0$ for all i. It is also clear that no agent can deviate profitably.

II) In the second kind (only possible if $n > 2$), only one agent produces a positive input. This equilibrium can be sustained with the following strategies: For each $i \neq n$, let $l_{ii+1} = T$ (modulo $n-1$) and $l_{ij} = 0$ otherwise, $l_{ni} = 0$ for all i. Clearly $R_i(l_i, l_{-1}) = 0$ for all $i \neq n$, and $R_n(l_i, l_{-i}) > 0$. It is also clear that no agent can deviate profitably.

However when $K < 1$, we can guarantee uniqueness of equilibrium.

Proposition 2. *Assume A1, A2, and $K < 1$. Then $l_{ij} = 0$ for all $i, j \in \{1, .., n\}$ is the unique Nash equilibrium.*

Proof. The proof is left to the Appendix.

Let us now consider the case when the possibilities of destruction are large, i.e. $K > n-1$. In contrast with the previous case, the necessary condition is no longer sufficient. For instance if in Example 1 we set $n = 4, T = 10, M = 5, \alpha = 0.5, r = 1/3, \rho = 1$ the necessary condition holds, but in this case no sabotage is not an equilibrium: the payoff of agent i in the point of zero sabotage is $\pi_i(0, 0) = 0.854$ but, for $l_i = (2, 2, 2)$, $\pi_i(l_i, 0) = 0.992$. So an increase in sabotage activities pays off.

The problem in this case is that the share of an agent can increase with sabotage, contrarily to what happened when $K \leq n-1$, (see 4.1). To guarantee that zero is a Nash equilibrium this effect can not compensate the fall in output caused by sabotage. The following assumption just formalizes this.

*A*3. For each agent i, and for $l_i = (l, .., l)$, $l \geq 0$, $l_{-i} = 0$,

$$-\frac{\partial s_i(R^l)}{\partial R_j}\frac{R_j^l}{s_i(R^l)} \leq \frac{\partial f(R^l)}{\partial R_j}\frac{R_j^l}{f(R^l)}, \tag{4.2}$$

where $R_i^l = T - (n-1)l$, $R_j^l = T - Kl$, for all $j \neq i$, and $R^l = (R_1^l, ..., R_n^l)$.

Let us interpret this assumption. Recall that the necessary condition when $M > n-1$ is written as

$$\frac{\partial s_i(R^0)}{\partial R_i}\frac{R_i^0}{s_i(R^0)} \leq \frac{\partial f(R^0)}{\partial R_i}\frac{R_i^0}{f(R^0)}\frac{(K+1)(n-1)}{K+1-n}, \text{ or}$$

$$\frac{1}{n-1}\frac{\partial s_i(R^0)}{\partial R_i}\frac{R_i^0}{s_i(R^0)} \leq \frac{\partial f(R^0)}{\partial R_i}\frac{R_i^0}{f(R^0)}\frac{(K+1)}{K+1-n}$$

By symmetry and anonymity and using $\sum_{k=1}^n \frac{\partial s_k(R^0)}{\partial R_i} = 0$, we obtain that

$$\frac{\partial f(R^0)}{\partial R_i}\frac{R_i^0}{f(R^0)} = \frac{\partial f(R^0)}{\partial R_j}\frac{R_j^0}{f(R^0)}, \text{ and}$$

$$\frac{1}{n-1}\frac{\partial s_i(R^0)}{\partial R_i}\frac{R_i^0}{s_i(R^0)} = -\frac{\partial s_i(R^0)}{\partial R_j}\frac{R_j^0}{s_i(R^0)}$$

Hence the necessary condition can be written in terms of the elasticity of the share and the production function with respect to R_j, namely,

$$-\frac{\partial s_i(R^0)}{\partial R_j}\frac{R_j^0}{s_i(R^0)} \leq \frac{\partial f(R^0)}{\partial R_j}\frac{R_j^0}{f(R^0)}\frac{(K+1)}{K+1-n}$$

Thus, $A3$ is similar to the necessary condition but stronger on two counts. On the one hand it is evaluated not only in the point of zero sabotage but in all points described above. On the other hand it is a little bit more restrictive since in $A3$ we do not have the factor $\frac{(K+1)}{K+1-n}$. Therefore when $A3$ holds the necessary condition holds as well.

It is easy to see that in Example 1, $A3$ evaluated in the point of no sabotage says that $\alpha \leq r$. However, when we evaluate inequality 4.2 in the other points required by $A3$, the degree of substitutability plays a role. Notice first that, in this case

$$\frac{\partial f(R^l)}{\partial R_j}\frac{R_j^l}{f(R^l)} = \frac{r(R_j^l)^\rho}{(R_i^l)^\rho + (n-1)(R_j^l)^\rho}.$$

Differentiating with respect to ρ we obtain

$$-r(R_j^l)^\rho(R_i^l)^\rho \frac{\ln R_i^l - \ln R_j^l}{\left((R_i^l)^\rho + (R_j^l)^\rho(n-1)\right)^2},$$

which is negative because in the points where the condition should be satisfied, $R_j^l \leq R_i^l$. Therefore the larger ρ the more restrictive is the condition.

We are now prepared to state and prove our next result.

Proposition 3. *Assume $A1$, $A2$, $A3$, and $K > n-1$. Then $l_{ij} = 0$ for all $i,j \in \{1,..,n\}$ is a Nash equilibrium.*

Proof. Given $l_{-i} = 0$, because of the previous Lemma, under assumption $A1$ we can restrict our attention to symmetric best responses. If we prove that all $j \neq i$, $\frac{\partial \pi_i(l_i,0)}{\partial l_{ij}} \leq 0$ for all $l_i = (l,...,l)$, then $l_i = 0$ is the best response for this agent. Let $R_j^l = R_j(l_i,0) = T - Kl$, $R_i^l = R_i(l_i,0) = T-(n-1)l$, and let

$R^l = (R_1^l, .., R_n^l)$. We know that

$$\frac{\partial \pi_i(l_i, 0)}{\partial l_{ij}} = f(R^l)(-\frac{\partial s_i(R^l)}{\partial R_j}K - \frac{\partial s_i(R^l)}{\partial R_i}) + s_i(R^l)(-\frac{\partial f(R^l)}{\partial R_j}K - \frac{\partial f(R^l)}{\partial R_i})$$

Since $K > n - 1$, and A2 holds $-\frac{\partial s_i(R^l)}{\partial R_j}K - \frac{\partial s_i(R^l)}{\partial R_i} \geq 0$. So in this case, the share of an agent increase with the sabotage activity. However, under $A3$, this effect does not compensate the decrease in the total production due to the sabotage. By $A3$, $f(R^l)(-\frac{\partial s_i(R^l)}{\partial R_j}K)+ s_i(R^l)(-\frac{\partial f(R^l)}{\partial R_j}K) \leq 0$, and since all the other terms that appear in the above expression are negative, $\frac{\partial \pi_i(l_i,0)}{\partial l_{ij}} \leq 0$, as we wanted to prove. ■

Our final example shows that starting with a situation where there is a Nash equilibrium without sabotage, an increase in the degree of substitutability causes not only that the sufficient condition fails but that the previous equilibrium ceases to exists.

Example 2. *Consider again Example 1 with* $n = 4$, $T = 10$, $K = 9$ *and* $r = \frac{1}{2}$. *In this case, the necessary condition is* $\alpha \leq \frac{5}{6}$. *But, the degree of substitutability plays a role in the existence of a Nash equilibrium without sabotage. In particular, if* $\rho = -1$, *the sufficient is* $\alpha \leq \frac{1}{2}$. *Then, if* $\alpha = \frac{1}{2}$ *zero sabotage is an equilibrium. However, for this degree of meritocracy, if* $\rho = 1$ *zero sabotage is not an equilibrium because* $\pi_i(0, 0) = 1.5811$, *and for* $l_i = (1.1, 1.1, 1.1)$, $\pi_i(l_i, 0) = 1.5969$. *So an increase in sabotage activities pays off.*

5. Conclusion

In this paper we have presented a model where agents can sabotage other agent's inputs in the framework of cooperative production and we have shown necessary and sufficient conditions to avoid sabotage in a Nash equilibrium. Possible extensions of our work include repeated interaction, agents with different productivities and the consideration of other forms of antisocial behavior like giving information that could destroy part of the input or stealing the input of other agents.[8] Let us comment briefly on the extensions that we think are more challenging.

1. **More General Sharing Rules:** We may consider more general sharing rules of the form $s_i(R_1,..,R_n) = s(R_i, \sum_{k=1}^n g(R_k))$ where g is a non decreasing and concave function. This would allow for sharing rules like $s_i(R_1,..,R_n) = \frac{(R_i)^\lambda}{\sum_{k=1}^n (R_k)^\lambda}$, $\lambda \in [0,1]$. Unfortunately, the sufficient conditions considered in this paper for preventing an equilibrium with sabotage do not work in this case. For instance,when $K < 1$, there might be equilibria with and without sabotage (see Beviá and Corchón (2003), Example 1) and, in some cases, even if $K < n-1$, only equilibria with sabotage exists (see Beviá and Corchón (2003), Example 2).

2. **Variable Working Time:** As we remarked in the Introduction we assume that the length of the working time is exogenously given. This assumption is motivated by simplicity and also it is well suited to study certain cases. However the case where each member of the cooperative individually decides the length of her working time is also relevant. How different is this case from the case analyzed in this paper? Clearly, the necessary condition for no sabotage still holds, because any working time has to be distributed optimally between productive and sabotage activities. But in this new framework no sabotage is not necessarily a socially optimal choice: For instance, egalitarian methods do not encourage effort because an agent effort yields benefits to all agents. In this situation it might

[8] Another possibility is to consider that agents can devote part of their time to preventing sabotage by others.

be that sharing rules that do not encourage sabotage also do not encourage a high level of effort. Therefore, social optimality may require to choose a sharing rule which encourages both effort and sabotage. Thus a model with a variable working time requires a considerable departure of the methods developed in this paper.

Summing up, in this paper we have studied a model of cooperative production and we have shown that an organization populated by rational agents might be self-destructive. In broad terms we might summarize our findings as follows. When the possibilities of destruction are small in relationship with the number of agents (i.e. $K \leq n-1$), all sharing rules satisfying two mild conditions yield a Nash equilibrium with no sabotage (Proposition 1). Under an additional condition this equilibrium is unique (Proposition 2). However, when the possibilities of destruction are not small (i.e. $K > n-1$), the degree of meritocracy that is compatible with the absence of sabotage depends negatively on the degrees of congestion and substitutability (Proposition 3). These findings can be used for the design of compensation systems and in organizing productive activities to avoid sabotage.

6. Appendix.

Proof of Proposition 2.

Suppose we have an equilibrium with positive sabotage, $(l_1, ..., l_n)$, and let $R^S = (R_1(l_i, l_{-i}), ..., R_n(l_i, l_{-i}))$. Then,

Step 1. There is at least one agent i such that $\sum_{j=1,j\neq i}^{n} l_{ji} \leq T$.

Suppose that for all agent i, $\sum_{j=1,j\neq i}^{n} l_{ji} > T$. Then, if we sum for all agents, $\sum_i \sum_{j=1,j\neq i}^{n} l_{ji} > nT$, but this is impossible since, because of the time constraint, for all j, $\sum_{i=1,i\neq j}^{n} l_{ji} \leq T$.

Step 2. There is at least one agent i such that $R_i(l_i, l_{-i}) > 0$.

Suppose that, for all agent j , $R_j(l_i, l_{-i}) = 0$. Then $\pi_j(l_i, l_{-i}) = 0$ for all j. By Step 1 we know that there is an agent i such that $\sum_{j=1,j\neq i}^{n} l_{ji} \leq T$. Since $K < 1$, and $R_i(l_i, l_{-i}) = 0$, the amount of time devoted to sabotage activities by this agent i should be strictly positive. But this can not be an equilibrium. If this agent reduces her sabotage activities, the total output will be positive and her input positive. Consequently, she will get a positive amount. Therefore, she will be better off.

Step 3. There are at least two agents, i and j, such that $R_i(l_i, l_{-i}) \neq R_j(l_i, l_{-i})$.

Suppose on the contrary that for all i, and j, $R_i(l_i, l_{-i}) = R_j(l_i, l_{-i})$. By the anonymity condition, $\pi_i(l_i, l_{-i}) = \frac{1}{n} f(R_1(l_i, l_{-i}), .., R_n(l_i, l_{-i}))$. Suppose agent i reduces her sabotage activity to zero, let $\hat{l}_i = 0$. Since $K < 1$, $R_i(\hat{l}_i, l_{-i}) > R_j(\hat{l}_i, l_{-i})$ for all $j \neq i$. Thus, $\pi_i(\hat{l}_i, l_{-i}) \geq \frac{1}{n} f(R_1(\hat{l}_i, l_{-i}), .., R_n(\hat{l}_i, l_{-i})) >$

$\frac{1}{n} f(R_1(l_i, l_{-i}), .., R_n(l_i, l_{-i}))$, which implies that agent i is better off.

Step 4. If $R_i(l_i, l_{-i}) \leq R_k(l_i, l_{-i})$, and $T - K \sum_{j\neq i} l_{ji} > 0$, then $l_{ik} = 0$.

Suppose on the contrary that $l_{ik} > 0$. Let us see that this can not be an equilibrium because the share of this agent is decreasing with l_{ik}. That is,

$$(-1)[\frac{\partial s_i(R^S)}{\partial R_i} + K\frac{\partial s_i(R^S)}{\partial R_k}] \leq 0.$$

Since $s_i(R^S) = s(R_i(l_i, l_{-i}), y(l_i, l_{-i}))$ with $y(l_i, l_{-i}) = \sum_j R_j(l_i, l_{-i})$, we have that:

$$\frac{\partial s_i(R^S)}{\partial R_k} = \frac{\partial s(R_i(l_i, l_{-i}), y(l_i, l_{-i}))}{\partial y}.$$

Since $R_i(l_i, l_{-i}) \leq R_k(l_i, l_{-i})$ and $\frac{\partial^2 s(x_i, y)}{\partial x_i \partial y} \leq 0$ we get the following:

$$\frac{\partial s(R_i(l_i, l_{-i}), y(l_i, l_{-i}))}{\partial y} \geq \frac{\partial s(R_k(l_i, l_{-i}), y(l_i, l_{-i}))}{\partial y} = \frac{\partial s_k(R^S)}{\partial R_i}.$$

Thus,

$$(-1)[\frac{\partial s_i(R^S)}{\partial R_i} + K\frac{\partial s_i(R^S)}{\partial R_k}] \leq (-1)[\frac{\partial s_i(R^S)}{\partial R_i} + K\frac{\partial s_k(R^S)}{\partial R_i}].$$

Since $\sum_{j=1}^n s_j(R_1, .., R_n) = 1$, then $\sum_{j=1}^n \frac{\partial s_j(R^S)}{\partial R_i} = 0$, therefore,

$$\frac{\partial s_i(R^S)}{\partial R_i} + \frac{\partial s_k(R^S)}{\partial R_i} = -\sum_{j \neq i, j \neq k} \frac{\partial s_j(R^S)}{\partial R_i} \geq 0.$$

which, since $K < 1$, implies that

$$(-1)[\frac{\partial s_i(R^S)}{\partial R_i} + K\frac{\partial s_k(R^S)}{\partial R_i}] < (-1)[\frac{\partial s_i(R^S)}{\partial R_i} + \frac{\partial s_k(R^S)}{\partial R_i}] \leq 0.$$

Step 5. Suppose, without loss of generality, that $R_1(l_i, l_{-i}) \leq .... \leq R_n(l_i, l_{-i})$, with at least a strict inequality because of Step 3. If $R_k(l_i, l_{-i}) \leq R_i(l_i, l_{-i})$, $T - K\sum_{j \neq i} l_{ji} > 0$, , and $i \neq n$, then $l_{ik} = 0$.

Notice first that if $T - \sum_{j \neq k} l_{kj} - K\sum_{j \neq k} l_{jk} < 0$, $l_{ik} > 0$ can not be an equilibrium, because agent i can decrease the time dedicated to sabotaging agent k without affecting the input of agent k but increasing her input, which implies that she will be better off. If $T - \sum_{j \neq k} l_{kj} - K\sum_{j \neq k} l_{jk} \geq 0$, let us see that

$l_{ik} > 0$ can not be an equilibrium because the share of this agent is decreasing with l_{ik}. That is,

$$(-1)[\frac{\partial s_i(R^S)}{\partial R_i} + K\frac{\partial s_i(R^S)}{\partial R_k}] \leq 0.$$

Since $s_i(R^S) = s(R_i(l_i, l_{-i}), y(l_i, l_{-i}))$ with $y(l_i, l_{-i}) = \sum_j R_j(l_i, l_{-i})$, we have that:

$$\frac{\partial s_i(R^S)}{\partial R_k} = \frac{\partial s(R_i(l_i, l_{-i}), y(l_i, l_{-i}))}{\partial y}.$$

Since $i \neq n$, $R_i(l_i, l_{-i}) \leq R_n(l_i, l_{-i})$ and $\frac{\partial^2 s(x_i,y)}{\partial x_i \partial y} \leq 0$ we get the following:

$$\frac{\partial s(R_i(l_i, l_{-i}), y(l_i, l_{-i}))}{\partial y} \geq \frac{\partial s(R_n(l_i, l_{-i}), y(l_i, l_{-i}))}{\partial y} = \frac{\partial s_n(R^S)}{\partial R_i}.$$

As it was shown in Step 4,

$$\frac{\partial s_i(R^S)}{\partial R_i} + \frac{\partial s_n(R^S)}{\partial R_i} = -\sum_{j \neq i, j \neq n} \frac{\partial s_j(R^S)}{\partial R_i} \geq 0.$$

which, since $K < 1$, implies that

$$(-1)[\frac{\partial s_i(R^S)}{\partial R_i} + K\frac{\partial s_n(R^S)}{\partial R_i}] < (-1)[\frac{\partial s_i(R^S)}{\partial R_i} + \frac{\partial s_n(R^S)}{\partial R_i}] \leq 0.$$

Step 7. For all agent i, $T - K\sum_{j \neq i} l_{ji} > 0$.
Suppose on the contrary that there are k agents such that $T - K\sum_{j \neq i} l_{ji} \leq 0$. Suppose that these agents are the first k agents. By Step 5, these agents do not suffer sabotage from agents $k+1$ to n. Thus $T - K\sum_{j=1, j \neq i}^{k} l_{ji} \leq 0$ for all $i = 1, ..., k$. Adding these inequalities for $i = 1, .., k$, we get: $kT - K\sum_{i=1}^{k}\sum_{j=1, j \neq i}^{k} l_{ji} \leq 0$. But, by the time constrain $\sum_{i=1, i \neq j}^{k} l_{ji} \leq T$, and since $K < 1$, $kT - K\sum_{i=1}^{k}\sum_{j=1, j \neq i}^{k} l_{ji} > 0$.
Step 8. There is not an equilibrium with positive sabotage. By the previous

Steps, we know that for all $i \neq n$, and for all j, $l_{ij} = 0$. Thus, if there is an equilibrium with positive sabotage, only agent n is using part of her time in sabotage activities. But since $K < 1$, and $R_n(l_i, l_{-i}) \geq R_i(l_i, l_{-i})$ for all i, the productive activities of agent n is larger than the productive activities of any of the other agents, which implies that we can not get an equilibrium with positive sabotage. ■

References

[1] Beviá, C., and Corchón, L. (2003). "Rational Sabotage in Cooperative Production", working paper UAB-IAE 541.02. http://pareto.uab.es/wp/2002/54102.pdf

[2] Fabella, R.V. (1988). "Natural Team Sharing and Team Productivity," *Economics Letters* 27, 105-110.

[3] Gibbons, R. (1998). "Incentives in Organizations," *Journal of Economic Perspectives,* 12 (4), 115-132.

[4] Holmstrom, B. (1982). "Moral hazard and Teams," *Bell Journal of Economics*, 13, 324-340.

[5] Holmstrom, B., and Milgrom, P. (1991). "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics and Organization,* 7, Special Issue, 24-52.

[6] Itoh, H. (1991). "Incentives to Help in Multi-Agent Situations," *Econometrica,* 59 (3), 611-636.

[7] Konrad, K. A. (2000). "Sabotage in Rent-Seeking Contests," *Journal of Law, Economics and Organizations*, 16 (1), 155-165.

[8] Lazear, E. P. (1989). "Pay Equality and Industrial Politics," *Journal of Political Economy*, 97 (3), 561-580.

[9] Macho-Staedler, I., and Pérez-Castrillo, D. (1993). "Moral Hazard with Several Agents," *International Journal of Industrial Organization*, 11, 73-100.

[10] Moulin, H. (1987). "Equal or proportional Division of Surplus and Other Methods," *International Journal of Game Theory*, 16, 161-186.

[11] Nandeibam, S. (2002). "Implementation in teams," Working paper, University of Birmingham.

[12] Prendergast, C. (1999). "The Provision of Incentives in Firms," *Journal of Economic Literature,* 37,7-63.

[13] Pfingsten, A. (1991). "Surplus Sharing Methods," *Mathematical Social Sciences*, 21, 287-301.

[14] Roemer, J., and Silvestre, J. (1993). "The Proportional Solution for Economies with Private and Public Ownership," *Journal of Economic Theory*, 59, 426-444.

[15] Sen, A. (1986). "Labour Allocation in a Cooperative Enterprise," *Review of Economic Studies*, 33, 361-371.