# Allocative Efficiency and an Incentive Scheme for Research[1]

Anindya Bhattacharya
Department of Economics and Related Studies,
The University of York, UK

Herbert Newhouse
Department of Economics
University of California, San Diego

This Version: April 2011

**Abstract**

In this paper we examine whether an incentive scheme for improving research can have adverse effect on research itself. This work is mainly motivated by the Research Assessment Exercise (RAE) and the Research Excellence Framework (REF) in UK. In a game theoretic framework we show that a scheme like RAE/REF can actually result in deterioration of the over-all research in a country though it may create a few isolated centres of excellence. The central assumption behind this result is that high ability researchers produce positive externalities to their colleagues. We assume these externalities have declining marginal benefit as the number of high ability researchers in a department increases. Because of this declining marginal benefit an incentive scheme like the RAE or REF may lead to over-concentration of the high ability researchers in a few departments and thus, the total research in the entire country may suffer.

JEL Classification No: C71, C72, I22, I28.

Keywords: research, RAE, REF, coalitions, strong nash equilibrium.

[1] We are indebted to Alessandra Canepa, Vincent Crawford, Karen Mumford, Arunava Sen and seminar audiences at Caen, Lancaster, Swansea and York for useful comments and discussions. Of course, the errors and shortcomings are ours.

## 1 Introduction

In this paper we examine whether incentive schemes for research can have an adverse effect on research itself. This work is mainly motivated by the Research Assessment Exercise (RAE) and the Research Excellence Framework (REF) in UK. In a game theoretic framework we show that a scheme like RAE/REF can actually result in deterioration of the over-all research in a country though it may create a few isolated centres of excellence. To capture the payment from such incentive schemes, a better department may have a tendency to replace persons with relatively lower ability by persons with relatively higher ability. A worse department may be unable to keep its persons with relatively higher ability and be forced to rely on persons with relatively lower ability.

The intuition behind this is as follows. We assume that a researcher gains from the external effects of her academic surroundings. Furthermore, we assume that for an academic professional this effect declines as her research ability increases. Therefore, overall research output may increase when higher ability researchers are distributed more equally among research institutions. However, an incentive scheme like RAE/REF awards an *entire department* some lump-sum payment on the basis of some *over-all* research output of the department and every member in the department has access to this reward. So, to capture this payment, a department will have a tendency to replace persons with relatively lower ability by persons with relatively higher ability. Thus, a strict hierarchy of departments would emerge—a few very good departments followed by a string of bad departments. For RAE 2001 such a phenomenon has been observed in the departments of UK (see Hare (2003)). Due to the declining returns of externalities in ability, overall research may increase if we move a high ability researcher from the top department in this hierarchy to a lower-ranked department.

This paper is simply a formal modelling of this idea to highlight this possible phenomenon in a precise manner. Our interest is in the peer effects between the researchers rather than the quality of the match between a particular researcher and her institution. For this reason we adopt the model of strategic coalition formation of Hart and Kurz (1983) rather than the more standard matching model (for example Bulow and

Levin (2006))[2]. We analyze the equilibria of a department-formation game and show that it is possible to obtain the somewhat perverse result that overall research may decrease for society if departments are rewarded for their individual research outputs.

Our purpose is simply to show that an incentive scheme such as the RAE/REF may have undesirable consequences on the allocation of researchers across departments. We make what we feel are simplifying but reasonable assumptions and show that such an incentive scheme for research lowers the overall research output for one particular example. The validity of our assumptions and whether or not our example is relevant are both empirical questions and are beyond the scope of this paper.

In Section 2 we provide a simple numerical example that illustrates the point made in this paper. Our model is explained in detail in Section 3. The results are collected in Section 4. Section 5 summarizes the plausible relevance of our modeling and analysis and finally concludes.

## 2 An Introductory Example

To illustrate the main point of our paper we start with a simplified example. The full model is formulated in Section 3.

Suppose there are twelve researchers in an economy, six with high ability and six with low ability. Research must be performed in a department and each department requires exactly four researchers. An individual's research output is based on her intrinsic ability and on the abilities of her colleagues. Specifically an individual's research output increases additively by a fixed amount if exactly one of her colleagues is of high ability. Her output increases by a larger amount if two of her colleagues are of high ability but there is no additional increase if all three of her colleagues are of high ability.

Arrange the twelve researchers into three departments. Define a perfectly stable department as one where each researcher is producing her maximum possible research

[2] Our framework could be adapted to include a number of researchers and a department that may form a coalition. However, for simplicity, here we model a department solely as a collection of its academic members.

output. An unstable department is one where at least one researcher is producing below her maximum possible research output. Assuming each researcher is paid a strictly increasing function of her own research output, no researcher at a perfectly stable department can strictly gain by forming a new department with other researchers. As an example, suppose the researchers are divided into the following departments: Configuration A – (HHHH), (HLLL), (HLLL). The first department given here is perfectly stable; each researcher in this department has at least two colleagues of high ability. The other two departments are not; each researcher at these departments has one colleague of high ability. From these two departments, two high ability researchers and two low ability researchers would all strictly gain if they formed a new department.

Recall that a strong Nash equilibrium is a coalitional equilibrium concept where no coalition of agents can jointly deviate so that each member of this coalition becomes strictly better off. For our example a strong Nash equilibrium will consist of a perfectly stable department, followed by a department with the remaining high ability researchers, followed by a department consisting entirely of low ability researchers. (The general case is proved in Proposition 2 below.) Two strong Nash equilibrium configurations are Configuration B – (HHHH), (HHLL), (LLLL) and Configuration C – (HHHL), (HHHL), (LLLL).

Now suppose an incentive scheme rewards members of a department based on the average level of research that occurs at that department. Suppose also that the incentive scheme payment is strictly increasing in the average research level of a department. Then Configuration C is no longer an equilibrium. The three high ability researchers from the first department and one of the high ability researchers from the second department could form a new department and all of them would strictly gain. The only strong Nash equilibrium given this incentive scheme is Configuration B.

However, note that the total research output is lower in Configuration B than it is in Configuration C. The fourth high ability researcher in the first department does not add to the externality whereas a third high ability researcher in the second department would add to the externality.

In the remainder of this paper we analyze a more general model capturing this basic idea. We show, first, that under some assumptions, such an incentive scheme will result in a strict hierarchy as the unique equilibrium outcome. Next we show that with such an incentive scheme, the total equilibrium research output may fall.

## 3 The Model

*The Agents:*

The finite set of $n$ players (each an academic) is denoted by $N$. Each $i$ in $N$ has an intrinsic ability for research $a_i$ lying in the interval [0, 1]. Without loss of generality, we order the players according to their ability for research, i.e., $i>j$ if and only if $a_i < a_j$.

We model the production of research as a two-stage game. In the first stage academics form departments as described later in detail. In the second stage each $i$ chooses a level of effort, $e_i \in [0, \bar{e}] \subseteq \mathbf{R}_+$ to produce research. The cost of effort for each $i$ is given by $c(e_i, a_i)$ with $\frac{\partial c}{\partial e_i} > 0$, $\frac{\partial^2 c}{\partial e_i^2} > 0$. If $a_i > a_j$ then, for each $e$,

$$\frac{\partial c}{\partial e}(e, a_i) < \frac{\partial c}{\partial e}(e, a_j).$$

*The Environment for Research:*

First, we represent the different aspects of research (volume, quality etc.) as a composite scalar variable. Next we assume that research can be conducted only in an institutional setting—say, in an academic department. A department $D$ is a non-empty subset of $N$. We assume that each feasible department must have exactly $k$ ($\leq n$) members[3]. As we have mentioned above, a player gets some positive externality in

[3] In the next subsection we explain the precise meaning of a feasible department. We can generalize this assumption a bit in the following way. Suppose the minimal size of a department must be $k$ (this is intuitive as a department consisting of only one member (say) is ridiculous!) and there is a congestion cost if the department size exceeds $k$. Then we can show that in equilibrium every department will have exactly $k$ members. However, little of importance is gained by this additional complication.

research from the presence of other members in a department[4]. We assume that this externality for a player $i$ in a department $D$ is given by:

$$q_i\,(D) = \frac{\sum_{j \neq i:\, j \in D} a_j}{|D| - 1}\,.$$

The total research output for person $i$ in department $D$ is given by:

$$x\big(e_i, a_i, q_i\,(D)\big)$$

where $e_i$ is person $i$'s level of effort, $a_i$ is her intrinsic ability and $q_i(D)$ is the level of the externality obtained by $i$ in department $D$. For all $i$, $\frac{\partial x}{\partial e_i} > 0$ with $\frac{\partial x(0)}{\partial e_i} \to \infty$ and $\frac{\partial x(\bar{e})}{\partial e_i} \to 0$. For all $i$, $\frac{\partial^2 x}{\partial e_i^2} \leq 0$. We sometimes express the output of $i$ by $x_i$ with no possibility of confusion. A person's research output is continuous and also strictly increasing in $q$, the level of externality, until a fixed $q_i^m$ <1. In [$q_i^m$,1], $x_i$ is constant with respect to $q$. Additionally we assume $\frac{\partial x}{\partial e_i} > 0$ is increasing in $q$ until $q_i^m$.

In [$q_i^m$,1], $\frac{\partial x}{\partial e_i}$ is constant with respect to $q$. We assume that if $i > j$ then $q_i^m \geq q_j^m$, that is, the less the intrinsic ability of a researcher, the further she is helped by externality from her colleagues. We denote the maximum value of $q_i^m$ (across all *i's)* by $\bar{q}$. This assumption expresses the idea that the effect of externality ceases at some point. This is a critical assumption for our main result.

Next we assume that if a player $i$ is more able than a player $j$ then the marginal research output of $i$ is more than that of $j$ at any given effort level given the same coworkers. We state this assumption as Condition A.

*Condition A: If* $a_i > a_j$ then for all $q$, $e$, $\frac{\partial x_i}{\partial e}\big(e, a_i, q\big) > \frac{\partial x_j}{\partial e}\big(e, a_j, q\big)$.

[4] This externality is empirically observed in U.S. universities in Kim, Morse and Zingales (2006) although they find the effect has diminished over time.

*The Department Formation Game:*

Instead of modelling the recruiting of faculty members as a matching process (see, e.g., Roth and Sotomayor (1990)) we model the process of the formation of a department *ab initio.* We use the model of strategic coalition formation introduced in Hart and Kurz (1983). This is a game in normal form where the outcome resulting from a strategy profile is a configuration of departments formed endogenously as a result of the strategic choice of the members. Given the well-known flux of the academics prior to the RAEs, such a modeling should be acceptable! We assume perfect information. This is justifiable as the research ability of a person is observed quite precisely by publications, participations in conferences etc.

The set of players is *N*. The strategy of a player is to announce the department she wants to be in. Therefore, formally, the strategy set for player *i*,

$$\sum_i = \{S \subseteq N \mid i \in S\}.$$

The outcome of a strategy profile $(S_i)_{i \in N}$ is a partition of *N*, *C*=($D_1$,..., $D_j$). Each member of this partition is a department. If a department is of cardinality *k* then it is feasible; otherwise it is infeasible. Suppose for *i* in *N*, *C*(*i*) is the unique element of *C* that contains *i*. Then,

$$C(i) = \{i\} \bigcup \{j \in N \mid S_i = S_j\}.$$

So, complete agreement among the potential members concerning who are to be included in the department is necessary for a feasible department to be formed[5]. If a department is infeasible (containing more members than *k* or less) then each player in such a department receives a pay-off of 0. If a department is feasible, then we distinguish two regimes. In the original regime, called O-regime, a person *i* in a department *D* receives a pay-off equal to her research output minus her cost of effort:

$$u_i^O\left(e_i \mid D\right) = x\left(e_i, a_i, q_i\left(D\right)\right) - c\left(e_i, a_i\right)$$

as described above. The total research output of a department is the sum of the research outputs of its members.

[5] Any partition that contains the maximum number of feasible departments is a Nash equilibrium for this stage of the game. We chose the strong Nash equilibrium solution concept to eliminate the equilibria we feel are vacuous.

Now suppose there is an incentive scheme so that the Government pays a lump-sum payment to the entire department on the basis of its average research output[6]. We model the pay-off of the players in such a regime (we call it R-regime) as follows. Suppose for a department *D* the average research output is $x_D^a$. Then *D* gets an additional payment $|D|\phi\left(x_D^a\right)>0$ and every *i* in *D* gets an additional lump-sum payment $\phi\left(x_D^a\right)$ where $\phi\left(\cdot\right)$ is a strictly increasing, concave function.[7] Then the pay-off to player *i* in department *D* is given by:

$$u_i^R\left(e_i \mid D\right)=x\left(e_i,a_i,q_i\left(D\right)\right)-c\left(e_i,a_i\right)+\phi\left(x_D^a\right)$$

This completes the description of the game.

The solution concept we use is a hybrid one in the spirit of backward induction. We assume that in the effort subgame each *i* in each department plays a Nash equilibrium in pure strategies given the effort choice of other members. Below we show that, fortunately, given our assumptions, such a Nash equilibrium is unique. Then, given the Nash equilibrium effort choices, we look at the Strong Nash Equilibria in pure strategies (SNE) (see, e.g., Bernheim et al. (1987)) of the reduced game of department formation under the two regimes. Recall that a strategy profile $(S_i)_{i\in N}$ is an SNE if there does not exist $D\subseteq N$ such that the players in *D* can jointly deviate (while those in *N\D* stick to the equilibrium strategies) and each player in *D* can strictly gain by such a deviation[8]. As we have already noted above, given the structure of our game, the power of an individual player is minimal. So, SNE, rather than a non-cooperative solution concept, is an appropriate solution concept for the reduced game of department formation.

[6] Although the exact ranking according to a specific incentive scheme like the RAE is much more complicated, the average research performance of a department is a good summary indicator of such rankings because, for example, RAE takes into account both the total output of the department as well as the percentage of the academic staff included in the RAE submissions.

[7] We can use other measures of central tendency like the median or some quantile of the distribution of research outputs of the department members as the basis for the Governmental lump-sum payment. Then also, the intuition of our result would be valid. However, of course, the precise conditions for the results would change.

[8] The notion of SNE is similar to the notion of stability in the matching literature. See, for instance, Kelso and Crawford (1982).

## 4 The Results

We proceed as follows. First (in Lemma 1) we show that for both the regimes, there exists a unique Nash equilibrium level of effort choice for each player in each department. Given this, we can meaningfully analyze the reduced game of department formation in the first stage as each player is the sure about what pay-off she will get at the play in the second stage. Next we show in Proposition 1 that under Condition A and assuming that $\phi$ is linear, the unique equilibrium configuration of departments that will be obtained in stage 1 under the R-regime is strictly hierarchical, i.e., the $k$ highest ability persons would be in one department, the set of next $k$ highest persons in another department, etc. Then we show in Proposition 2 that in general, more than one equilibrium configurations of departments may emerge in the O-regime. Finally, in Proposition 3 we provide an example where the total research output in one of the equilibrium configurations of departments for the O-regime is more than that in the unique equilibrium configuration of departments in the R-regime.

First we look at the level of effort chosen in equilibrium by a player within a department at the second stage of the game.

**Lemma 1** *There exists a unique Nash equilibrium in pure strategies for the effort choice subgame in both regimes.*

**Proof:** Note that for both regimes the pay-off function of each $i$ is continuous in the profile of effort choices and strictly concave in $e_i$. Since the strategy set is compact, a Nash equilibrium in pure strategies exists.

Next, for either regime, for any department, $D$, consider the $(|D|\times|D|)$ matrix $J$ such that the $ij$-th entry of $J$ is given by $\frac{\partial^2 u_i}{\partial e_i \partial e_j}$.

Given that $\frac{\partial^2 c}{\partial e_i^2} < 0$, *J* can be easily shown to be negative definite. To see this, for the R-regime, note that the *ii*-th entry of *J* is given by:

$$(1+\phi')\frac{\partial^2 x_i}{\partial e_i^2} - \frac{\partial^2 c}{\partial e_i^2} + \phi''\frac{\partial x_i}{\partial e_i}\frac{\partial x_i}{\partial e_i}$$

and the *ij*-th entry (with *i* different from *j*) is given by $\phi''\frac{\partial x_i}{\partial e_i}\frac{\partial x_j}{\partial e_j}$.

Therefore, *J* can be written as the sum of two matrices *J'* and $\phi''$*J''* where *J'* is a diagonal matrix whose *ii*-th entry is given by:

$$(1+\phi')\frac{\partial^2 x_i}{\partial e_i^2} - \frac{\partial^2 c}{\partial e_i^2}$$

whereas, the *ij*-th entry of *J''* is given by $\frac{\partial x_i}{\partial e_i}\frac{\partial x_j}{\partial e_j}$.

For any |*D*|-dimensional vector *a*, the product *a'*$\phi''$*J''a* is $\phi''(\sum_{i \in D} a_i \frac{\partial x_i}{\partial e_i})^2$ which, given the concavity of $\phi$ is non-positive. And given our assumptions on the functions *x* and *c*, the matrix *J'* is negative definite. Therefore, the matrix *J* is also negative definite.

Then by Rosen (1965), there exists a unique pure strategy Nash equilibrium in the effort subgame. □

The effort level $e_i$ chosen by player *i* in equilibrium of the effort-subgame in department *D* in the original (O) regime satisfies the following first order condition:

$$\frac{\partial x}{\partial e_i}\left(e_i, a_i, q_i\left(D\right)\right) = \frac{\partial c}{\partial e_i}\left(e_i, a_i\right) \qquad [1]$$

Similarly, the effort level $e_i$ chosen by player *i* in equilibrium of the effort-subgame in department *D* in the incentive (R) regime satisfies the following first order condition:

$$\left(1+\phi'(\bar{x}_D)\right)\frac{\partial x}{\partial e_i}\left(e_i, a_i, q_i(D)\right)=\frac{\partial c}{\partial e_i}\left(e_i, a_i\right) \qquad [2]$$

where $\bar{x}_D$ is the total research output produced in *D* in equilibrium.

Now we look at the problem of the existence of an SNE in the reduced game of department formation. We start with another lemma.

We show that total research output is higher in a better department. First we define a better department.

**Definition 1** *Department $D^B$ is a better department than $D^W$ if $a_i^{D^B} \geq a_i^{D^W} \quad \forall i \in \{1, 2, \ldots k\}$ with $a_i^{D^B} > a_i^{D^W}$ for at least one $i \in \{1, 2, \ldots k\}$ when $D^B$ and $D^W$ are both ordered from the best member to the worst member (here $a_i^D$ stands for the ability of the i-th player in department D).*

**Lemma 2** *Suppose $D^B$ is a better department than $D^W$. Then the total research output in $D^B$ in equilibrium is more than that in $D^W$.*
*Moreover, let $\phi$ be linear. Let $i_{D^B}$ be the i-th ranked player (according to ability) in $D^B$ and $i_{D^W}$ be the i-th ranked player in $D^W$. Then $e_{i_{D^B}}{*} \geq e_{i_{D^W}}{*}$ where $e_{i_D}{*}$ is the equilibrium effort choice of the i-th ranked player in department D.*

**Proof:** Suppose otherwise. Then $\exists D^W$ s.t. $x\left(D^W\right) \geq x\left(D^B\right)$. (Here, by *x(D)* we denote the total research output in department *D in equilibrium.*) For this condition to hold, at least one researcher in $D^W$ must exert more effort than the identically ranked researcher in $D^B$. That is, $\exists j \in D^W$ and $\exists i \in D^B$ such that the equilibrium level of effort by player *j* is more than that of player *i*, which we write simply as $e_j > e_i$. Recall that researcher *i*'s first order condition for the effort subgame in the R-regime (Equation 2):

$$\left(1+\phi'\left(\overline{x}_D\right)\right)\frac{\partial x}{\partial e_i}\left(e_i,a_i,q_i\left(D\right)\right)=\frac{\partial c}{\partial e_i}\left(e_i,a_i\right)$$

Now, $\left(1+\phi'\left(\overline{x}_{D^W}\right)\right)\le\left(1+\phi'\left(\overline{x}_{D^B}\right)\right)$ by concavity of $\phi$. Also, by Condition A, $\frac{\partial x}{\partial e_j}\left(e_j,a_j,q_j\left(D^W\right)\right)\le\frac{\partial x}{\partial e_i}\left(e_i,a_i,q_i\left(D^B\right)\right)$ since $e_j > e_i$, $a_i \ge a_j$ and $q_i\left(D^B\right)\ge q_j\left(D^W\right)$.

However, $\frac{\partial c}{\partial e_j}\left(e_j,a_j\right)>\frac{\partial c}{\partial e_i}\left(e_i,a_i\right)$ due to our assumptions about $c\left(\cdot,\cdot\right)$.

Then the first order condition cannot hold for both $i$ and $j$. Therefore total output must be higher in the better department. This argument, clearly, also works for the O-regime (there, simply, the condition $\left(1+\phi'\left(\overline{x}_{D^W}\right)\right)\le\left(1+\phi'\left(\overline{x}_{D^B}\right)\right)$ is left out).

Now, assume that $\phi$ is linear, that is, its derivative is constant. Then, replicating the argument above we find that $e_{i_{D^B}}*\ge e_{i_{D^W}}*$ where $e_{i_D}*$ is the equilibrium effort choice of the $i$-th ranked player in department $D$ as required in the lemma.[9] □

Lemma 2 implies that the equilibrium total research output in a better department (as defined above) is more than in any worse department.

Next we show that under Conditions A and the assumption of linearity of $\phi$ in the R-regime, the *unique SNE outcome would be such that a strict hierarchy of departments would form*. The $k$ highest ability persons would be in one department, the set of next $k$ highest persons in another department, etc. Formally:

**Proposition 1** *Suppose Condition A holds and let* $\phi$ *be linear. Then, in the R-regime, the unique SNE outcome* $\left(D_1,...,D_j\right)$ *is as follows. For every* $i,k\in N\ \&\ m,n\in\{1,...,j\}$ *such that* $\left[i\in D_m,k\in D_n\right]$, $\left[m>n\Rightarrow i>k\right]$.

[9] The linearity of $\phi$ is a simple sufficient condition, but not necessary for the second part of Lemma 2 to be valid.

**Proof:** Let $D_1$ be the department consisting of the *k* best researchers. Let *D*' be any other department with $i \in D_1, D'$. The department $D_1$ is a better department than *D*' as specified in Definition 1. Rank the researchers in $D_1$ and *D*' according to ability. Lemma 2 shows that $e_j^{D_1} \geq e_j^{D'}$ for each person of rank *j* in either department. By Lemma 2 we also get that $\sum_{\substack{j \neq i \\ j \in D_1}} x_j \geq \sum_{\substack{j \neq i \\ j \in D'}} x_j$.

Suppose $u_i^R\left(e_i^{D'}* \mid D'\right) > u_i^R\left(e_i^{D_1}* \mid D_1\right)$, that is, let player i's pay-off given her equilibrium choice of effort in department D' be more than her pay-off given her equilibrium choice of effort in department $D_1$.

Then $u_i^R\left(e_i^{D'}* \mid D_1\right) > u_i^R\left(e_i^{D_1}* \mid D_1\right)$.

But then $e_i^{D_1}*$ cannot be a Nash equilibrium choice for player *i* in the effort choice subgame in department $D_1$. □

*Remark 1:* Proposition 1 gives the unique equilibrium for the R-regime. Specifically this equilibrium has the *k* best researchers in the first department, the remaining *k* best researchers in the second department and so on[10]. Next we will demonstrate the possible existence of other equilibria in the O-regime.

**Proposition 2** *There exists at least one SNE for the department formation game in the O-regime.*

**Proof:** Our proof is constructive.

First we say that a department *D* is perfectly stable (given a regime) if for every *i* in *D*, $u_i^O\left(e_i^{D}* \mid D\right) \geq u_i^O\left(e_i^{D'}* \mid D'\right)$ for every other department *D'*. Given the SNE solution concept no player would deviate out of a perfectly stable department. Now we describe the construction.

[10] In a paper on this theme La Manna (2008) explains that this type of hierarchy will result because the better departments get more funding from the RAE which leads them to hire better researchers. He then uses reliability theory to determine when such a hierarchy is desirable.

Step 1: Form a perfectly stable department $D_1$. Then, from $N \setminus D_1$, form another perfectly stable department. Continue this process as long as possible. This process would terminate owing to the finiteness of *N*. Let this collection be ($D_1$,..., $D_j$). (Note that the set of such departments may be empty.)

Step 2: Form $D_{j+1}$ by taking *k* ``best'' players (in terms of individual intrinsic ability) from the remaining $N \setminus \{D_1 \bigcup .. \bigcup D_j\}$ and so on until no more feasible departments can be formed.

Note that a player from department $D_{j+1}$ may gain by forming a new department with players from ($D_1$,..., $D_j$). However, by Lemma 2, none of the players in ($D_1$,..., $D_j$) would gain strictly by forming such a department. Furthermore, a player from department $D_{j+1}$ cannot gain by forming a new department with players from $\left( D_{j+1}, ..., D_{\lceil \frac{n}{k} \rceil} \right)$.

Therefore taking ($D_1$,..., $D_j$) as given, no player from department $D_{j+1}$ will deviate. Similar logic demonstrates that no player will leave the remaining departments formed in Step 2.

The resulting set of departments is an SNE outcome. □

*Remark 2*: Note that the strict hierarchy found as the unique SNE in the R-regime is also an SNE in the O-regime.

Proposition 3 below gives our desired result. It shows that it is possible to have strictly higher total research output in an equilibrium outcome in the O-regime compared to the unique equilibrium outcome in the R-regime.

**Proposition 3** *Given the above assumptions it is possible to have strictly higher total research output in an equilibrium outcome in the O-regime compared to the unique equilibrium outcome in the R-regime.*

**Proof:** We demonstrate this with an example.

Suppose $n = 4$ and $k = 2$.

$a_1 = 0.7,\ q_1^m = 0.2$

$a_2 = 0.6,\ q_2^m = 0.3$

$a_3 = 0.4,\ q_3^m = 0.6$

$a_4 = 0.2,\ q_4^m = 0.7$

$c\left(e_i, a_i\right) = e_i^2 - a_i e_i$

The following results hold for any sensible output function.

Consider two structures of departments:

Structure A: $D_1 = (1, 2)$, $D_2 = (3, 4)$.

Structure B: $D_1 = (2, 3)$, $D_2 = (1, 4)$.

Note that by Proposition 1, Structure A is the unique equilibrium for the R-regime and by Proposition 2, is also an equilibrium for the O-regime. Structure B is an equilibrium for the O-regime; both departments are perfectly stable since each player in each department receives her maximum externality.

Consider the two structures under the O-regime.

Player 1 and player 2 each receive her maximum externality. Therefore each of them will produce the same amount in either structure.

Recall that by Equation 1, player 3's first order condition for equilibrium choice of effort for structure A, $e_3^{A*}$, requires that:

$$\frac{\partial x}{\partial e_3}\left(e_3^A *, a_3, 0.2\right) = \frac{\partial c}{\partial e_3}\left(e_3^A *, a_3\right).$$

Similarly, player 3's first order condition for equilibrium choice of effort for structure B, $e_3^{B*}$, requires that:

$$\frac{\partial x}{\partial e_3}(e_3^{B*}, a_3, 0.6) = \frac{\partial c}{\partial e_3}(e_3^{B*}, a_3).$$

Also note that:

$$\frac{\partial x}{\partial e_3}\left(e_3^B*, a_3, 0.6\right) > \frac{\partial x}{\partial e_3}\left(e_3^A*, a_3, 0.2\right)$$

$\Rightarrow \quad e_3^B* > e_3^A*$ by our assumption on the function $c(\cdot,\cdot)$.

Similarly, $e_4^B* > e_4^A*$.

Therefore the total output under the O-regime at Structure B is greater than that for Structure A.

Then, if $\phi'(\cdot)$ is small enough, then by the continuity of the pay-off functions of the players, the total output under the O-regime at Structure B is greater than that for the unique equilibrium under the R-regime.

*Remark 3*: Of course, this result holds for many other examples; we just provide one fairly simple case.

## 5 Conclusion

The above results are not meant as a condemnation of any specific incentive scheme such as the RAE or the REF. Rather, we merely demonstrate that it is possible that such incentive schemes may lower total research output given individually optimizing researchers. Since our results are dependent on the values of the parameters, any specific incentive scheme for research may or may not lower the overall level of research in an economy. One area of future research is how probable it is that such a perverse scenario exists.

## References:


1. Bernheim D., B. Peleg and M. Whinston (1987): "Coalition-Proof Nash Equilibria I: Concepts", *Journal of Economic Theory,* **42,** 1-12.

2. Bulow, J. and J. Levin (2006): "Matching and Price Competition", *American Economic Review*, **96**, 652-668.

3. Epple, D. and R. E. Romano (1998): "Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects", *American Economic Review*, **62**, 33-62.

4. Hare, P. (2003): "The UK's Research Assessment Exercise: Impact on Institutions, Departments, Individuals", *Higher Education Management and Policy*, **15**, 43-62.

5. Hart, S. and M. Kurz (1983): "Endogenous Formation of Coalitions", *Econometrica,* **51**, 1047-1064.

6. Kelso, A. and V. Crawford (1982): "Job Matching, Coalition Formation, and Gross Substitutes". *Econometrica*, **50**, 1483-1504

7. Kim, E. H., A. Morse and L. Zingales (2006): "Are Elite Universities Losing Their Competitive Edge?", NBER Working Paper 12245.

8. La Manna, M. (2008): "Assessing the Assessment. Or the RAE and the Optimal Organization of University Research". *Scottish Journal of Political Economy*, **55**, 637-653.

9. Rosen, J. B. (1965): "Existence and Uniqueness of Equilibrium Points for Concave N-Person Games", *Econometrica,* **33**, 520-534.

10. Roth, A. and M. Sotomayor (1990): *Two-sided matching: a study in game-theoretic modeling and analysis.* Econometric Society Monographs, vol. 18. Cambridge University Press, Cambridge.

11. Winston, G. C. and D. J. Zimmerman (2003): "Peer Effects in Higher Education", in *College Decisions: How Students Actually Make Them and How they Could*, C. Hoxby, ed., University of Chicago Press.