Science and Public Policy, 34(10), December 2007, pages 000–000
DOI: 10.3152/030234207X265394; http://www.ingentaconnect.com/content/beech/spp

Focal randomisation: an optimal mechanism for the evaluation of R&D projects

Elise S Brezis

In most countries, governments intervene in the process of R&D by financing a substantial part of it. The mechanism employed for choosing the projects to be financed is a committee composed of experts who evaluate projects in their field of specialisation, and decide which ones should be funded. This method is conservative. Proposals for new ideas are too often rejected, and inventions are commonly thrown out of the set of potential projects. In this paper, I propose a mechanism that will allow less conformity: focal randomisation. This states that projects that are unanimously ranked at the top by all reviewers will be adopted. Projects perceived as valueless by all are rejected, while projects that are ranked differently are randomised. I compare the average return under the present and proposed mechanisms. I examine under which conditions this new method is preferable, and its consequences on economic growth.

IN MOST COUNTRIES, governments intervene in the process of R&D by financing a substantial part of it. The reason for this intervention is that research and development undertaken by one firm has positive spillover effects on the entire economy. Since firms do not take these spillover effects into account, they invest in R&D less than the optimal amount. It is difficult to estimate these externalities, but it could double the real rate of return. Moreover, basic research is a non-excludable investment and firms have no interest in undertaking it in a competitive market. Therefore, government financing is necessary. In some countries, it can amount to more than 60% of total R&D funding.

Such an intervention, however, creates a problem: how does the government choose which projects to finance? It could be that the projects chosen are not those with the highest potential growth for the economy and therefore not optimal for the country.

In this paper, I discuss the mechanism of evaluating and choosing the projects to be financed in a given field, but do not treat the problem of deciding which field to finance. The mechanism employed in most countries is a committee composed of research fellows who evaluate projects in their field of specialisation and decide which ones should be funded. Alternatively, the decision is left to the committee chair, taking into consideration the referees' reports. Both mechanisms are based on what is termed 'peer review'.

The problems with peer review have been analyzed at length and, in the next section, we give an outline of the different problems. However, most of these negative effects do not affect the rate of growth of the economy as does one specific problem: the conservative bias. Indeed, peer review presents a bias against innovative applications, as emphasised by many and especially Martin (1997: 72):

> A common informal view is that it is easier to obtain funds for conventional projects. Those who are eager to get funding are not likely to propose radical or unorthodox projects. Since you don't know who the referees are going to be, it is best to assume that they are middle-of-the-road. Therefore, a middle-of-the-road application is safer.

This paper proposes a mechanism that reduces the negative effects of the conformity bias. It can be used for evaluating R&D projects and articles. However, the consequences of making mistakes in funding

---

Elise S Brezis is Director of the Azrieli Center for Economic Policy Department of Economics Bar-Ilan University, 52900 Ramat Gan, Israel; Email: brezie@mail.biu.ac.il; Tel: +972 3 531 8946; Fax: +972 3 738 4034.

Professor Elise S Brezis is currently the director of the Azrieli Center for Economic Policy at Bar-Ilan University. Previously, she was head of the Statistics Division at the Research Department, Bank of Israel and, from 1999 to 2003, she was the President of the Israeli Association for the Study of European Integration. She holds a PhD in Economics from MIT (USA). Her research interests cover a wide range of topics related to economic growth, from R&D to higher education and human capital. She has published articles in the *American Economic Review* and *European Economic Review*, the *Oxford Encyclopedia* and for New Palgrave. She has edited a book on *Elites and Economic Growth* (Elsevier). Her current interests are evaluation of R&D projects that affects economic growth, and evaluation of entry exams at universities, on which she has published a series of articles.

research are more severe than with publishing articles, since in the latter a good piece of work might not be known until later. However, if great research were not funded, it would not come through.

This proposed mechanism is based on the typology of Arrow, which divides new technologies into inventions and innovations: innovations are new applications of an already known technology, while inventions are completely new technologies for which it is impossible to foresee their economic consequences. This paper will show that, for projects in the category of innovations, it is not difficult for specialists in the field to evaluate them, and therefore the mechanism designed for evaluating innovations is adequate. However, projects entering the category of inventions (a new technology with unforeseen applications) cannot be evaluated properly and may therefore be rejected. Thus, the method employed leads to the rejection of inventions, and there is a need for a new mechanism to value them.

The errors in estimating the value of these inventive projects can be very substantial. Some examples, from the past highlight these errors of estimation. For instance, 50 years ago, the Chief Executive Officer (CEO) of IBM claimed: “I think that the world market for computers is for no more than five computers”. Eighty years ago, the commander of the Allied Forces during World War I said: “Planes are a nice toy but with no military value”. One of the worst estimations on an invention was written by the CEO of Western Union in 1876 when deciding which project to finance: “The telephone has too many problems to be taken seriously as a telecommunication tool. No value for our firm”.

The mechanism used nowadays for evaluating projects is conservative, and new ideas might be less accepted than they should be. Inventions are commonly thrown out of the set of potential projects and, as a consequence, the government chooses investments with lower return.

In this paper, I propose a mechanism that will allow less conformity: the ‘focal randomisation mechanism’; it will lead to acceptance of some projects with high potential, without accepting projects that certainly have no future. This mechanism stipulates that, when referees have a good understanding of the value of certain projects, their report will be final; but on projects that appear to be inventive, and which referees cannot evaluate, we propose that projects will be chosen randomly. This paper examines under which conditions this mechanism is preferable.

I present three simple models that differ in their assumption about the information set of the referees. In the first section, the model assumes almost perfect information, while, in the model of the third section, referees have almost no efficient way of evaluating projects. It is clear that, in reality, the third model is the one prevailing. Our policy recommendations, presented in the conclusion are based on this last model.

The literature

The literature on evaluation can be divided into two main categories: research on the effects of *ex-ante* evaluation and research on *ex-post* evaluation. The *ex-ante* evaluation is related to choosing research to be funded or papers to be published. The *ex-post* evaluation literature focuses on how to evaluate the impact of the research funded.

It should be noted that most of the literature on evaluation is in fact analyzing evaluation *ex-post*, that is, how we can determine the results and effects of a particular piece of research.[1] The literature on *ex-post* evaluation is mainly divided into two main channels: scientific and experimental models vs management-oriented systems models. Lately Jaffee (2002) has proposed a new design for *ex-post* evaluation: regression-discontinuity. His proposal is based on the fact that all projects can be divided into two groups: the projects that have been funded, and those that were not. Consequently, we can check if there is a significant difference between these groups. This will allow us to examine whether funding has an effect on research.

This paper is about *ex-ante* evaluation, that is, how to value proposals and which one to fund. Peer review today is the most common mechanism for evaluative testing of R&D projects or research. Most of the main funding agencies believe that expert review is the most effective mechanism for evaluating the quality, leadership and relevance of research (especially basic research) performed and funded by them. Ultimately, decisions regarding the selection and funding of research programs must be made by agency managers informed by expert review. So, the peer review is used in almost every country to distribute public funds for research.

In the USA, for instance, the NAS (1999: chapter 4) asserts:

> Federal agencies should use expert review to assess the quality of research they support, the

> relevance of that research to their mission, and the leadership of the research. Expert review must strive for balance between having the most knowledgeable and the most independent individuals serve as members. Each agency should develop clear, explicit guidance with regard to structuring and employing expert review processes. The most effective way to evaluate research programs is by expert review. The most commonly used form of expert review of quality is peer review.

Despite the wide use of this way of picking projects to be funded, there is some criticism of this mechanism. Peer review leads to many biases (see Smith, 1994). The first is called the confirmatory bias, which has been analyzed by Mahoney (1977). Scholars have a tendency to reject research that will show results against their own theoretical perspective. There is a clear bias for scholars to accept research that supports or confirms their beliefs. Moreover, there is some proof of a lack of impartiality, vested interests or rivalry (see Martin, 1997).

Peer review tends also to present an institution bias: reviewers favour submissions from prestigious institutions (see Godlee *et al*, 1998). Prestige of the person or institution affiliation affects the reviewer recommendation. Peters and Ceci (1982) have shown that names and affiliations affect acceptance: they have resubmitted papers that had already been published, after changing the names and affiliations of the authors. Their results confirm a bias related to reputations of colleagues and institutions.

There is also a gender bias. A Swedish study found clear evidence of discrimination against women in awarding research grants (see Wenneras and Wold, 1997). There is also a “positive” bias. It seems that referee tend to prefer positive than negative results. Already Bacon (1621) wrote: “the human intellect .. is more moved and excited by affirmative than by negatives.”

The bias, I intend to examine in this paper is the ‘conservative bias’. This is against innovative ideas and inventions. Horrobin (1990) has suggested that peer review is a conservative process. It encourages research in a known field, and makes multidisciplinary work difficult. It looks for safe research so that innovative, risky and unconventional ideas will not be funded.

Peer review has a clear bias towards research that confirms the reviewers’ beliefs; there is also an institutional bias favouring prestigious institutions, a gender bias against women and a conservative bias against innovative ideas

This paper proposes a mechanism based on peer review that will cure the problem of conservative bias. It will also permit other wrongdoings in peer review to be weeded out. Lately, there is proof of misconduct with peer review: the most important are failure to check proper data and plagiarism. The US National Science Foundation conducted an inquiry into plagiarism. They found cases of researchers who were asked to peer review a proposal for research turning it down, and then proposing it themselves for another grant; the methodology has been copied word for word! In the UK, detecting fraud has become the aim of the Committee on Publication Ethics (COPE).

I will show in this paper that the focal randomisation mechanism (FRM) can reduce the bias against inventions, and also the effects of wrongdoing in peer review.

A framework for FRM

First model: homogeneity of reviewers

Our first model assumes that all reviewers are similar and they all take the time that is needed to evaluate well the different projects.

Assumptions

i. We have k projects, from which only h can be funded.

ii. Projects are divided into two groups: innovations and inventions. Innovations are technologies based on an already-known technology. A referee that is specialised in this field is capable, if he spends some time, of analyzing the different projects and evaluating them correctly. Inventions are projects presenting a new technique and with unforeseen knowledge as to its future value. We assume in this first model that referees are capable of making the distinction between innovations and inventions, even if they cannot evaluate the invention projects. In this model we also assume that all referees make the same decision about inventions and innovations, since they are all alike. There are k_o innovation projects, and k_v invention projects ($k_o+k_v=k$).

Model Let us define:

V_{oi} the true value of the ith innovation project, $i=1\ldots k_o$

V_{vi} the true value of the ith invention project, $i=1\ldots k_v$

U_{oi} the value given by the referees of the ith innovation project

U_{vi} the value given by the referees of the i invention project

We order the projects in an increasing value such that

$$V_{o1} < V_{oi} \ldots < V_{ok_o}$$
$$V_{v1} < V_{vi} \ldots < V_{vk_v}$$

The value of an innovation project, V_{oi}, is based on two parameters. The first is its originality compared to the technology already known, that we define as D_i (for distance from the known technology), where $D_i \in [0, D]$, and the longer the distance from the known technology, the higher the value.^{2}

The second element is the cleverness of the project, B_i (for brightness); B_i € $[0, B]$. The more clever a project, given its D_i, the higher the value. Therefore we have:

$$V_{oi} = \alpha D_i + \gamma \mathrm{Bi} \tag{1}$$

We assume in this first model that the referees identify, without error, the value of the innovation. We have, therefore, that:

$$U_{oi} = \alpha D_i + \gamma B_i \tag{2}$$

An invention has a third element which is the inventive part of the technology, λ_i. Therefore:

$$V_{vi} = \alpha D_i + \gamma B_i + \beta \lambda_i \tag{3}$$

Referees cannot recognise the true value of an invention, λ_i, and they all give the value λ^*.

$$U_{vi} = \alpha\, D_i + \gamma\, B_i + \beta \lambda^* \tag{4}$$

Moreover, we assume that, since referees have difficulty in recognising the future value of inventions, they tend to underestimate it, and give to λ^* a lower value than the average of λ_i.

Mechanism for choosing optimal projects To present the mechanism of focal randomisation in a clear and vivid way, we will use a numerical example.

We assume that four projects have to be chosen out of ten (k=10, and h=4). There are seven innovations, and three inventions (this is in the information set of the referees). The value given by the referees to these inventions is low. Therefore, in the present mechanism, none of the inventions will be chosen.

We set $\alpha = 10$; $\gamma = 1$; and the D_is and B_is are:

[4,5,7,8,12,3,9], [0,30,70,65,50,160,110]

Therefore V_{oi}s and U_{oi}s are:

[40,80,140,145,170,190,200]

For the inventions, we set $\alpha = 10$; $\gamma = 1$; $\beta = 1$; $\lambda^* = 20$;and the D_is, B_is and λ_is are:

[6,5,1]; [0,0,0]; [30,80,300]

Therefore V_{vi}s and U_{vi}s are respectively:

[90,130,310]; [80,70,30]

Given these data, we show in the first row of Table 1, the optimal choice from the point of view of the country. The invention (the tenth project) and innovations number 5, 6, 7 should be funded.

Under the present system, and under our assumptions that innovations are well ranked, we get that the invention (#10) is not going to be funded, and therefore the value for the country is lower.

Our focal randomisation mechanism in this simple model states:

> *All invention projects should be pooled, and one of them should be picked up in a random way.*

From this example, we see that focal randomisation allows a greater return to government financing because, under the present process, all inventions would have been rejected.

It is clear that countries in which the government is very risk averse, this mechanism should not be adopted. If, for some reasons, the country prefers a lower growth in the near future to a much higher growth in the distant future, it should not invest in inventions. This would be the case if the country runs a big current account deficit and cannot easily borrow abroad. However, if the country can borrow overseas, and has no credit rationing, then it should invest in inventions to increase the growth rate in the long run. For developed countries, governments have to finance the projects that display high risk, since the private market will invest in low-risk projects with high returns. So, if the government does not fund the 'inventions', it loses some of its goal.

We note also that, if there were perfect correlation between D_is and λ_is, randomisation would not be needed and invention with the higher D_is would be chosen. It would still be necessary to separate the process of choice for innovations and inventions. This model has assumed that referees have perfect information about the value of innovations. Next, we present a more realistic model.

Table 1. First model: new and old methods

	Optimal choice		Present system		Randomisation	
1	#10	310	#7	200	1/3(#8+#9+#10)	177
2	#7	200	#6	190	#7	200
3	#6	190	#5	170	#6	190
4	#5	170	#4	145	#5	170
Total		870		705		737

Second model: heterogeneity of reviewers

In this section, we waive the assumption that referees have perfect knowledge on the value of an innovation, as well as the assumption that referees are all alike. Sometimes, some of the reviewers do not invest enough time to understand perfectly the import of the project. It is widely known that: “We are concerned that the standard of the reports that we receive from our peer reviewers is not always very high. Many of the men and women whom we ask to review for us are busy people. Perhaps they don’t have the time or motivation to do the job as well as they should” (Martyn, 1992: 322).

In consequence, we assume that time spent on reviewing affects the quality of the report. We maintain that referees can make the distinction between inventions and innovations.

Assumptions

i. We have k projects, from which only h can be funded.
ii. Projects are divided in two groups: innovations and inventions. As opposed to the first model, a referee is not always capable of evaluating innovations correctly. The more time he spends investigating the project, the better proxy he gets to the project value.

Model As in the first model, we have:

V_{oi} the true value of the ith innovation project
U_{oij} The value given by referee j to the ith project

The definition of V_{oi} is identical to the first model:

$$V_{oi} = \alpha D_i + \gamma B_i \qquad (1)$$

In the previous model, the referees evaluate D_i and B_i without error. We now assume that the cleverness part, B_i, can be evaluated with no error: all referees can say if a project is clever or not. However, the distance (the novelty) from the technology already known, D_i, is not easily evaluated as in the previous model. We define T_j as the time that referee j takes to investigate an innovation and assume that, if the time invested is higher than the innovation distance, that is, $D_i \leq T_j$, then the referee can correctly estimate the true value of the innovation. However, if $D_i > T_j$, then he does not understand the true value. We assume that the more time a referee spends analyzing the project, the closer he gets to the true value D_i; and the greater the difference between D_i and T_j, the larger the error in valuation. The specific form chosen, for sake of simplicity is:

$$\begin{aligned} U_{oij} &= \alpha D_i + \gamma B_i \text{ for } D_i \leq T_j \\ &= \alpha T_j + \gamma B_i \text{ for } D_i > T_j \end{aligned} \qquad (5)$$

The difference, Δij, between the true value, and the estimation is therefore:

$$\begin{aligned} \Delta ij &= 0 \text{ for } D_i \leq T_j \\ &= \alpha(D_i - T_j) \text{ for } D_i > T_j \end{aligned} \qquad (6)$$

The error of estimation is zero if the time invested by the referee is greater than D_i, and is a positive function of the difference between the innovative value of the project and the time given by the referee, $D_i – T_j$.

How is the referee to choose the amount of time he spends on each project? On one hand, he wants to take the least time possible, $k \times T_j$, since he could use this time for alternative tasks. It is assumed that every referee has his own subjective value to time, s_j. On the other hand, the referee does not want to make too large an error about the true value of the project, since he is concerned that the best projects will be chosen. Therefore, the loss function he wants to minimise is:

$$L = \min\,(\Delta j)^2 + s_j\,(\mathrm{k} \times T_j)^2 \qquad (7)$$

where $\Delta j = \sum \Delta_{ij}$ that is, the sum of the errors he makes.

The optimal T_j for each referee is the argmin of function L. It is found by taking the first order condition (FOC) of equation L. Thus, referees with high subjective value of time, will prefer to spend less time T_j on each project, and for all projects such that $D_i > T_j$ their error will be relatively greater. The optimal T_j for each referee is denoted by T_j*.

Concerning inventions, we assume the same function as in the previous section:

$$V_{vi} = \alpha D_i + \gamma B_i + \beta \lambda_i \qquad (3)$$
$$U_{vi} = \alpha D_i + \gamma B_i + \beta \lambda^* \qquad (4)$$

Mechanism for choosing optimal projects Under these assumptions, the optimal way to rank projects is to accept the projects that all referees have ranked at the top, and to reject all the projects that all referees have put at the bottom. For projects on which referees do not agree as to ranking, randomisation is optimal. We term this mechanism: ‘focal randomisation’, since randomisation is performed only on a subset of projects.

To compare the mechanism presently used with the one proposed, we take the same numerical example as in the previous case, although five projects will be funded. So we assume k=10, and h=5. We have seven innovations, and three inventions, as in the previous case.

We set α =10, γ =1; and the D_is and B_is are:

[4,5,7,8,12,3,9], [0,30,70,65,50,160,110]

For the inventions, we set α =10; γ =1; β=1; λ^*=20; and the D_is, B_is and λ_is are:

[6,5,1]; [0,0,0]; [30,80,300]

Therefore the V_{oi}s and V_{vi} are respectively as in the previous case:

$$[40, 80, 140, 145, 170, 190, 200]\ [90, 130, 310]$$

About the valuation of the peer review, let us assume that we have three referees with s_j such that the optimal time spent for reviewing the projects is 4, 5 and 9 respectively. Therefore from equation (5), U_{oij} j=1,2,3 are :

$U_{oi\,1}$= [40, 70, 110, 105, 90, 190, 150]; the ranking is: 6, 7, 3, 4.
$U_{oi\,2}$= [40, 80, 120, 115, 100,190, 160]; the ranking is: 6, 7, 3, 4.
$U_{oi\,3}$= [40, 80, 140, 145, 140,190, 200]; the ranking is: 7, 6, 4, 5.

and on average we get:

$$U_{oi} = [\ 40, 77, 123, 122, 110, 190, 170] \tag{8}$$

In Table 2, we present the options under the different mechanisms, focusing only on innovations. In the first row of Table 2, we present the optimal choice from the point of view of the country, and projects 4, 5, 6 and 7 should be funded.

Under the present system, and given the valuation presented in equation (8), the choice of the projects is as shown in Table 2. The projects funded are: 3, 4, 6, 7.

Our focal randomisation mechanism in this simple model states:

> All projects that are ranked in a consistent way at the top by all reviewers *should* be adopted. Projects that are ranked differently *should* be randomised.

In our example, #7, #6 and #4 are adopted by all reviewers. On #3 and #5, we will randomise. The results are presented in the last column of Table 2.

We see that the possibility of randomisation for the innovation projects that fall into the middle category increases the total value of the projects. The inventions, as in the previous case, are determined by randomisation and, therefore, comparing the new method to the one in use today, we have the results shown in Table 3.

In this model, we have assumed that referees can make a distinction between inventions and innovations. Focal randomisation is the preferable mechanism, since it gives a better result than the regular peer-review mechanism that is used nowadays. However, the assumption that referees can make a distinction between innovations and inventions can sometimes be inappropriate; it is waived in the next model.

Table 2. Second model: method in use today

	Optimal choice		Present system		Randomisation	
1	#7	200	#6	190	#7	200
2	#6	190	#7	200	#6	190
3	#5	170	#3	140	#4	145
4	#4	145	#4	145	1/2 (#3+#5)	155
Total innovation projects		705		675		690

Third model: the general case

In this section, we keep all the assumptions of the second model, except the one that claims that referees can make the distinction between inventions and innovations. Moreover, referees differ in their perception about inventions. Some referees, when analyzing an invention project, may believe that it is a good idea and its probability of success is high, but others will disagree.

Assumptions

i. We have k projects, from which only h can be funded.
ii. Projects are no longer divided into two groups from the point of view of the referee. Referees are different in their subjective value of time, as well as their degree of imagination and possibility of being inventive.

Model As in the first and second models, the true value of innovations and inventions are respectively:

$$V_{oi} = \alpha\, D_i + \gamma\, B_i \tag{1}$$
$$V_{vi} = \alpha\, D_i + \gamma\, B_i + \beta\lambda_i \tag{3}$$

In fact, we can write that the value of a project, independently of being an invention or an innovation is always equation (3), when for an innovation the value of λ_i is 0. Therefore we get:

$$V_i = \alpha\, D_i + \gamma\, B_i + \beta\lambda_i \tag{9}$$

We order all the projects (inventions and innovations) in increasing value such that

Table 3. Second model: new methods and old

	Optimal choice		Present system		Randomisation	
1	#10	310	#6	190	#7	200
2	#7	200	#7	200	#6	190
3	#6	190	#3	140	#4	145
4	#5	170	#4	145	1/2 (#3+#5)	155
	#4	145	#5	170	1/3 (#8+#9+#10)	177
Total projects		1015		845		867

Table 4. Third model ranking

Rank	(2) D_i	(3) B_i	(4) λ_i	(5) V_i	(6) U_{i1} (T=4, λ=20)	(7) U_{i2} (T=5, λ=30)	(8) U_{i3} (T=9, λ=150)	(9) Average
1	4	0		40	40	40	40	40
2	5	30		80	70	80	80	77
3	6	0	30	90	60	80	90	77
4	5	0	80	130	60	80	130	90
5	7	70		140	110	120	140	123
6	8	65		145	105	115	145	122
7	12	50		170	90	100	140	110
8	3	160		190	190	190	190	190
9	9	110		200	150	160	200	170
10	1	0	300	310	30	80	160	90

$$V_1 < V_i \ldots < V_k.$$

The referees try to estimate these values. We denote U_{ij} the value given by referee j to the ith project. It is, as in the previous model, a function of the time spent analyzing the project and the cleverness of it. It is also a function of the referee's opinion on how innovative the project is. As stated earlier, nobody can really perceive the value of an invention, but some referees are more inventive than others and have better intuition as to what the future portends. We call λ_j the intuition of j, which is distributed normally on the whole range $[0, \infty]$. The referees, therefore, make two types of errors, one on D_i and one on λ_i.

On the distance, B_i, the error made by referee j is, as in the second model:

$$\begin{aligned} \Delta_{oij} &= 0 \text{ for } D_i \leq T_j && (6)\\ &= \alpha\,(D_i - T_j) \text{ for } D_i > T_j \end{aligned}$$

When considering the inventive element, λ_i, we make three assumptions. The first is that the more creative the referee, the better he/she estimates the invention element. The second is that, if the referee is more creative than the project proposed, he/she makes no error in the value. The third is that the error is an increasing function of the difference between the true value and his/her creative possibilities.Therefore:

$$\begin{aligned} \Delta_{vij} &= 0 \text{ for } \lambda_i \leq \lambda_j && (10)\\ &= \beta(\lambda_i - \lambda_j) \text{ for } \lambda_i > \lambda_j \end{aligned}$$

We get that the valuation given by a referee is:

$$\begin{aligned} \mathrm{U}_{ij} &= \alpha\, D_i + \gamma\, B_i + \beta\lambda_i \text{ for } D_i \leq T_j \text{ and } \lambda_i \leq \lambda_j && (11)\\ &= \alpha T_j + \gamma B_i + \beta\lambda_i \text{ for } D_i > T_j \text{ and } \lambda_i \leq \lambda_j \\ &= D_i + \gamma B_i + \beta\lambda_j \text{ for } D_i \leq T_j \text{ and } \lambda_i > \lambda_j \\ &= \alpha T_j + \gamma\, B_i + \beta\lambda_j \text{ for } D_i > T_j \text{ and } \lambda_i > \lambda_j \end{aligned}$$

Mechanism for choosing optimal projects We show that the optimal ranking mechanism is to accept projects that all referees have ranked at the top (on which the variance between referees is low), and to reject the projects that all referees have put on the bottom (on which the variance between referees is low). For those on which referees do not agree as to the ranking (the variance is high), randomisation is optimal. We compare in a numerical example the mechanism presently used with the one proposed.

The same example is used as in the two previous models, but innovations and inventions are presented together. As in the first model, k=10,and h=4. We have seven innovations and three inventions, but the referees are unaware of it. We set α =10; γ =1; β=1; and the D_is, B_is and λ_is are as in the previous examples and shown in Table 4, rows 2, 3 and 4 respectively:

Let us assume that there are three referees with T_j and λ_j that are respectively: (4, 5, 9) and (20, 30, 150). Therefore, from equation (11), U_{ij} j=1, 2, 3 are as presented in columns 6–8. The average of these three referee reviews is presented in column 9. The projects chosen under the different systems are shown in Table 5.

As can be seen, focal randomisation leads to higher average return. It allows projects with high returns to be accepted. This does not imply that all inventions financed will be successful. On average, however, inventions have a higher return than innovations. This is the key element in the importance of inventions being accepted.

Conclusion

This paper has examined the consequences of the conformity effect on *ex-ante* evaluation. It has presented

Table 5. Third model: projects chosen

	Optimal choice		Present system		Randomisation	
1	#10	310	#9	200	#9	200
2	#9	200	#8	190	#8	190
3	#8	190	#6	145	#6	145
4	#7	170	#5	140	1/3[#5+#7+#10]	207
Total		870		675		742

Focal randomisation leads to higher average return and allows projects with high returns to be accepted: on average, inventions have a higher return than innovations; this is the key element in the importance of inventions being accepted

some simple models, quite closed to reality, in which randomisation produces the best solution. Since referees tend to understand projects better that are close to existing technology, they may reject inventive projects that will yield high return in the future. Peer review leads to conformity, while a random choice will permit the funding of inventive projects.

I would like to underline that focal randomisation should not be used on the whole set of proposed projects, because some of them are valueless, which referees perceive immediately. Other projects are very good, and should be chosen without randomisation. On the rest of the set of projects, randomisation seems to be optimal. Therefore, it is a focal randomisation. Initially, the projects are grouped in three categories, and only on the target group of the 'unclear' value is randomisation performed.

Rational scholars sometimes feel uneasy with a randomisation mechanism. Why should bright scholars throw dice, when it is so clear to them that they make the right choice! Therefore, they do not like to waive the right to choose.

First, it is important to underline that scholars keep the power on what we term focalisation, that is, the scholars decide which projects enter the group entitled to randomisation. When all scholars decide unanimously that a specific project is not good, it is immediately thrown out; when all decide that a project is good, it is immediately chosen. However, there are many projects on which there are disagreements. Focal randomisation will not throw out these projects but will perform randomisation on this focal group.

These models have taken into consideration only that the mistakes of the referees are a result lack of information. It could also be that referees choose projects in which they are not completely disinterested. They could act not in the public interest exclusively, but might have self interest, and might, for some subjective reason, dislike a project. Their report would, therefore, not take only the public interest into consideration. This 'public choice' perspective would strengthen the importance of introducing randomisation, into which no elements of sympathy, approval or power enter. Randomisation on the projects over which referees disagree not only increases diversity, but is a way of avoiding the tendency to accept projects of 'club' insiders.

Notes

1. A broad coverage of *ex-post* evaluation can be found in Fahrenkrog *et al*, 2002.
2. For simplicity, we assume that D is 15, and B is 200.

References

Fahrenkrog, G, W Polt, J Rojo, A Tubke and K Zinocker eds. 2002. RTD Evaluation Toolbox —Assessing the Socio-economic Impact of RTD Policies. Strata Project HPV1 CT 1999-00005. Seville, Spain: IPTS Technical Report Series.

Godlee, F, C R Gale and C N Martyn 1998. The effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomised controlled trial. *Journal of the American Medical Association*, **280**(3), 237–242.

Horrobin, D F 1990. The philosophical basis of peer review and the suppression of innovation. *Journal of American Medical Association*, **263**(10), 1438–1441

Jaffe A 2002. Building programme: evaluation into the design of public research-support programmes. *Oxford Review of Economic Policy*, **18**(1), 22–34.

Mahoney, M 1977. Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, **1**(2), 161–175.

Martin, B 1997. Peer review as scholarly conformity. In *Suppression Stories*, B. Martin, pp. 69–83. Wollongong: Fund for Intellectual Dissent.

Martyn, C 1999. Peer review: some quotations from Socrates. In *Peer Review in Health Sciences*, eds. F Godlee and T Jefferson. London: BMJ Publishing group.

NAS, National Academy of Sciences 1999. Research and the government performance and results acts. Committee on science engineering and public policy evaluating federal research programs.

Peters, D and S Ceci 1982. Peer review practices of psychology journals: the fate of submitted articles, submitted again. *Behavioral Brain Science*, **5**(2), 187–195.

Smith, R 1994. Promoting research into peer review. *British Medical Journal*, **309**(6948), 143–144.

Wenneras, C and A Wold 1997. Nepotism and sexism in peer review. *Nature*, **387**(5), 341–343.