Institutional Checks and Balances for a Social Contract of Science.

Public Research as a Principal-Agent Problem.\(^N\)

[ Preliminary draft: please do not cite, yet. ]

Remo Fernández Carro

Instituto Juan March

Universidad Carlos III de Madrid

March 2003

Abstract: Some recent literature in Sociology and Political Science as well as in Economics studies science policy as a principal-agent problem. This approach examines the whole institutional system of science as a solution for a one-sided riddle, how can the government police moral hazard and adverse selection of scientists and how can it build trust on their behaviour and results. This article proposes that most of those institutional solutions also solve for agents the problem of principals’ commitment. Actual institutions, procedures and regulated activities in public science guarantee the overall contract against misbehaviour from both parts, given the unusual information asymmetries between each scientist and his or her patron, but also the asymmetries in power resources. The need for government’s commitment and for institutional guaranties to science agents is much more evident when it lacks –in dictatorships, for instance– but it remains in democracies and partly explains science institutional set.

Keywords: Sociology of Science, Science Policy, Research Policy, Science and Government, Principal-Agent, Incomplete Contracts, Adverse Selection, Moral Hazard, Reward System of Science, Science Institutions, Institutionalism.

\(^N\) Research has been done first at the Instituto Juan March (Centro de Estudios Avanzados en Ciencias Sociales) for my Doctoral Thesis –supervised by José María Maravall and read at the Universidad Autónoma de Madrid in May 2002–, and finished at the Universidad Carlos III Madrid. I am grateful to my supervisor, and to Sonia Alonso Sáez de Oger and Luis Ortiz Gervasi for a number of valuable suggestions and criticisms.
Some recent works in Political Science and in Sociology, mainly those by Guston (2000, 1999, 1996), Van der Meulen (1998) and Braun (1993), suggest that it is worth to model the relationship between science and politics as a principal-agent problem.\(^1\) A previous literature from Economics suggested that a contract between a scientist and any patron implies problems of both adverse selection and moral hazard—and thus of principal-agent—\(^2\) mainly due to strong information asymmetries. However, though the cited theoretical models explain institutions of science described by the Social Sciences, they focuses only in partial aspects of the problem or only in some of the institutions. Moreover, they focus chiefly in the difficulties for the patron to police agent’s misbehaviour (see, for instance, Guston 1999), although the association is established between equal partners rather than a servant and a master.\(^3\) Symmetrical problems can be described for the agents to control principal’s wrongdoing. The cited authors disregard, on one hand, the role of science institutions to protect scientist from the relationship inherent instability, and its role in creating agent’s trust on principal’s guarantees. As a consequence they ignore the way other social and political institutions stabilize the relationship: the social compact of science is not about policies but about politics. On the other hand, models proposed in the literature are based mostly on prospective judgements on the credibility of promises (Turner 1990:190) although the reward system—and principal-agent theory—works on retrospective assessment of past performance: this imply a different relationship and some other institutions. Often, works on science studies do not profit from the research made in Economics about principal-agent problems, and they do neither benefit from Neo-Institutionalist works on how political institutions solve moral hazard and adverse selection problems of governments.\(^4\)

---

\(^1\) Rip (1994), and Rip and van der Meulen (1995) use the idea as a metaphor. Guston (1996) highlights the analysis by Turner (1990), as an implicit principal-agent one. Braun (1993) focuses rather in intermediate institutions in government—and treats the research councils in science policy only as a case.


\(^3\) Brooks (1997) underlines this—the expression is his—when he glosses the creation of the American social compact of science in the Bush-Kilgore debate.

\(^4\) Symmetrically, works from Economics do not pay attention to empirical results from Sociology or Political Science about science structure and dynamics—except for the forty-years-old theories by Robert Mer-
In this paper, I propose that key institutions of the *republic of science* are established to solve pervasive problems that arise in the relationship between a scientist and any patron, and among the science community in this relationship: *hidden action* and *hidden information* as well as *collective-action* problems. Institutions like *peer review*, *academic boards and panels*, *science journals and papers*, *research councils* and their inner structure, *science funds*, and *reputation* or *credit* are well explained as mutual guarantees for both scientists and governments in the institutional structure of science. Moreover, an *aggiornamento* of Merton’s *reward system of science* can make it the incentive scheme of the whole contract of science: I will explore this idea further in this paper. Science policy tools can also be explained under this light.

The model proposed here focuses in science that is intended to be published and that is non-proprietary –instead of secret or patented science. It tries to explain the whole institutional system of science after the few basic principles of principal-agent theory, and accordingly with observations on how science works. Other *ad hoc* explanations are avoided. It extends the basic assumptions of the cited literature to take into account guarantees for the agent, and models it as repeated games with retrospective pay rules. The model, indeed, is built on the concepts of *utility* and *interest* that inform the *strong programme* of the sociology of science.

---

footnote: I discuss this later on. On institutionalist approaches to political institutions see, for instance Moe (1990 and 1984).

footnote: We can label it *pure or basic*, although it includes also *applied* science. Dasgupta and David (1987) discuss this criterion to set apart science and technology. Incentives are different when principals and agents can share, for example, patent dividends.

footnote: Like that of *boundary work*, which is not determined by *necessary or transcendent* characteristics of science (Guston 1999:87).
Some assumptions.

“The scientist, like the bureaucrat, is a challenge to the legislator while nominally his subordinate. His recondite knowledge gives him the upper hand. (...) The politician cannot reasonably be expected to possess the knowledge necessary for adequate supervision of the scientist’s work, and yet he feels a responsibility for doing so. Where the results of their work cannot be checked in detail, and where indeed the executive branch discourages it, there is a corresponding increase in the fervor to check on their qualities, on their attitudes, and above all on their loyalty”.

As the theory goes deep in rational choice explanations and follow the logic of its basic assumptions, it must describe most of the behaviour of principals and agents as a result of rational maximization of individual utility. Science institutions, heuristically, should serve this logic of maximization, prevent collective action problems and mitigate transaction costs. Actors try to set and operate them. As a consequence, most of the conflicts we could observe in the community –as well as the discourses on conflicts– would deal with the operation of those institutions and the organizations that carry them out. Eventually, conflicts would arise about the failures of the institutions to be set.

Shils’ quotation illustrates the main problem of the relationship, information asymmetries: even the incertitude of results, modelled by Arrow, tends to disappear with time. The patron of scientists cannot supervise directly neither the effort the agent pays, nor her skills, and not even the quality of the results. The usual resort to a third party’s external evaluation is nearly useless when only few others can evaluate a scientist work. The problems of information asymmetries and the distrust repeats over and over.

To simplify my model, I will focus on the maximization of material rewards: even if personal ends deal with intellectual satisfaction, material goods can be taken as instrumental for these ends. I contend that emotional recompenses for living a life according to the professional ethos, as Merton proposed, is the driving force of science.

---

7 Edward Shils (1962:610) was worried about scientists and politicians in the Cold War.
8 I also assume that institutions can be used as means to exert power (Moe 1990) and that a specific set of institutions can fail to achieve the goals the actors who established them wanted for (see, for instance, North, 1990).
Rather, I will follow Barnes (1977) and other authors who suggest that interests and utility drive science, through an indirect reward system based in reputation or credit.9

Generic Principal-Agent Theory studies the kind of association in which a customer relies on a doer to fulfil a task he cannot do. We call the first principal and the second agent. In the simplest version of the model, the agent chooses an action that affects both her and principal’s welfare; the principal selects a payoff rule based on the evaluation of agent’s performance. However, information asymmetries prevent the principal to know both the action, a, its nature or its magnitude (the level of effort of this action), and the circumstances, θ, that affect the productivity of the action. Circumstances, such as the state of the world or the agent’s productivity, are only observed by the agent.

The theory focuses on the payoff rule that, in each specific case, brings optimal agent’s action given her hidden information. A good payoff rule can filter less productive agents before the contract is established. Or it can help to reveal agent monitoring information once the relationship begins. The implicit assumption is that the agent can lie, can counterfeit, can fake, can falsify or distort information, can hide, cloak, disguise, or conceal her effort, and can swindle the principal out. Science results, that cannot be directly weighed up by the principal, offer more opportunities for agent’s misbehaviour—and most science studies authors are worried about this. This is the reason for a principal-agent ex post payoff system, a way to create patron’s trust on agent.

However, principal can also misbehave. Theory often assumes that he will honour the contract or the contract payoff rule. This is correct when a contract is established in a well founded institutional system that can enforce it—and these institutions can be

---

9 Hagstrom (1988) was the first to propose that science is a gift-giving exchange system based on reputation, similar to economic institutions like the Melanesian kula. The criticism on Merton’s reward system of science based on public recognizing of priority and on a set of norms, Merton’s ethos, began with Barnes and Dolby (1970) or some articles of Michael Mulkay (see 1991: 39-50, 51-61 and 62-78, and 1979). Mulkay points that “(…) conformity to the supposed norms and counter-norms of science is largely irrelevant to the institutional processes whereby professional reward are distributed. Researchers are simply rewarded for communicating information which their colleagues deem to be useful in the pursuit of their own studies. There are no institutional mechanisms for rewarding conformity to either set of social norms; nor it is possible to show that the provision of acceptable information presupposes the implementation of either set (…)” (1991:67). Barnes summarizes this indirect reward system in his essay About Sociology (1985); see also Latour and Woolgar (1986) point of view about a cycle of credibility.
taken for granted. But this is not always the case: without enforcing mechanisms the contract cannot work. Moreover, the value of science results for the principal is commonly noncontractible; and the patron’s interests or expectations can change with time. The situation is similar to a prisoner’s dilemma game (see Figure 1). The agent (1) decides the effort she puts on an action; the principal (2) decides then whether to pay or not. If there is not a previous commitment, there is no reason for the principal to pay; and, obviously, there is no optimal cut-off effort level for the agent to put. The principal is better off if he does not pay, and the agent anticipates this result and chooses not to put any effort. Most of the risk of the relationship goes to the agent.

![Figure 1](image)

Agent’s ordered preferences are: being paid without putting any effort in, being paid after the effort has been done, not being paid but not to spend any effort, and, finally, not to be paid though the effort was done:

\[(\text{no e, p}) > (\text{e, p}) > (\text{no e, no p}) > (\text{e, no p})\]

The principal prefers not to pay when effort has been done, prefers to pay the effort done, not to pay when effort is not put in, and pay when the agent puts no effort in:

\[(\text{e, no p}) > (\text{e, p}) > (\text{no e, no p}) > (\text{no e, p})\]

More often, contracts rely on a previous credible commitment of one or all of the parts. Institutions can be the result of this commitment, an element of the relationship that warrant actors’ behaviour and protect both from their counterpart’s misbehaviour. Incertitude of science results allows scientist’s patron too haggling the scientist’s fee. In the absence of such an institutional warrants framework, there is a role for previous commitment and for newly shaped warranty institutions. A payoff rule is the result of a negotiation between equals rather than a one-sided imposition. Some key institutions of science, reputation for instance, are thus established to monitor both scientists’ and their patron’s behaviour.

---

10 Alghion and Tirole (1994: 1,189), studying proprietary science contracts, assume that the nature of a science result is ill defined ex ante; the value of the innovation is either a private benefit or an unspecific monetary benefit.

11 This is not exactly a prisoner’s dilemma game because player 2 knows player 1’s decision in advance.

12 Ferejohn (1999) observes that tendency on theorists to believe that the payoff rule is put forward by the principal. He suggests, conversely, that in some cases the agent propose an already tailored contract to the principal.
A previous credible commitment ($c$) of the principal ($2$) makes incredible any agent’s ($1$) menace to choose no $e$. Then, they are better off if they collaborate.

The relationship, nonetheless, is repeated in time. Research is a long term activity and most of its features, such as scientific careers or research projects, are years long ones. Science institutions are also long term and retain this quality of the relationship.

**Government as the scientist’s principal in a repeated game model**

When scientists’ patron is the government, an actor much more powerful than the whole community of science, there is a vast difference on power resources. It is not easy to find enforcement mechanisms for politicians to keep promises to individual researchers. In democracies, periodical elections are the mechanism to control incumbents;\(^{14}\) Ferejohn (1986) describes voter-incumbent relationship as a repeated principal-agent one based on past performance of candidates. Although it could be seen a time inconsistency source for principal’s interests, because governments change more often raising transaction costs, it is a good mean to stabilize the contract through society’s surveillance of politician’s commitments. It imposes a cost to the politician for changing opinion. Consequently, a whole model of the social contract of science in democracy makes society both the principal of scientist’s principal—and this way the real scientist’s patron—, and a third party that receives some benefit of research and gives the

---

\(^{13}\) In this toy model I assume that all incumbents want science, and that they all want the same kind of science. However, the case could be that they do not support scientific research or that they want a particular field or some, applied industrial research instead of agricultural or medical science, for example.

\(^{14}\) One should expect that, if democracies are better than polyarchies in enforcing private contracts and in checking incumbents’ behavior (North 1990, Dreze and Sen 1989, Sen 1992 and 1999, Maravall 1987), then democracies are more productive of science. Empirically, this is the case (Fernández-Carro 2001).
government some information on scientists performance. In this section, I will focus on the government-scientists association and I will use an agency model close to Ferejohn’s to explain it. I will assume, first, that the principal can directly observe the result of the game.

In every round of the repeated game, the agent observes a certain state of the world, $\theta$, and accordingly chooses an action, $a$, or a level of effort. The principal cannot observe $\theta$ nor $a$; in the world of research, he is not expected to possess the knowledge necessary to understand neither agent’s information nor actions although most of them are public, or disclosed soon. This state of the world includes information about the world of science –opportunities to make a certain discovery within a discipline, for instance– and private agent’s information –agent’s qualifications or talent; it also includes Arrow’s uncertainty, although in time its mean is zero. Nature plays first but this is not directly observable by the principal. The incumbent observes then the result of the action, $a\theta$, and decides whether to keep the scientist working and give her payoff, or not. So the prize is for the agent to remain at work in the next round. If results are below a certain cutoff level the agent is turned out and the payoff is 0. Although a higher cutoff level implies higher agent’s effort, there is a danger of abandonment if it is too high. Cutoff level depends directly on the utility for the agent to continue at work and thus on her wage.

The consequence is that an unfavourable value of $\theta$, given any cutoff level, and the scientist will choose to abandon and will not expend any effort. With a state of the

15 The entire political system can be taken as a series of nested principal-agent games. Moe (1984:765) explains: “Democratic politics is easily viewed in principal-agent terms. Citizens are principals, politicians are their agents. Politicians are principals, bureaucrats are their agents. Bureaucratic superiors are principals, bureaucratic subordinates are their agents. The whole of politics is therefore structure by a chain of principal-agent relationships, from citizen to politician to bureaucratic superior to bureaucratic subordinate and on down the hierarchy of government to the lowest-level bureaucrats who actually deliver services directly to citizens”. Moe finishes inviting us to apply the model to each hierarchical stage of the government in which, “Aside from the ultimate principal and the ultimate agent, each actor in the hierarchy occupies a dual role in which he serves both as principal and as agent” (766).

16 Ferejohn does not focus in the problem of principal’s credible commitment, but the set is about the same: the principal commits because this strategy is subgame perfect and he does not profit else. However, agent’s doubts about commitments can be taken as embedded in her evaluation of the state of the world, $\theta$, as I will discuss later.

About this kind of models, see also Morrow (1994: ch. 9).
world mainly determined by individual qualifications or personal talent, a high cutoff level produces a screening effect on low qualification or less talented people. But when the state of the world is mostly related to prospects out of the agent’s control, a high cutoff point can prevent good candidates to remain; and the same happens when payoffs are low, and when opportunities outside are better.

Principal is better off in this model if he commits. However, the agent may mistrust about his ability to keep commitments—due, for instance, to corruption of principal’s officials or to their incapacity to evaluate the work. Lack of principal’s commitment can be modelled in the game as an external determinant of $\theta$: an agent’s evaluation of principal’s propensity to misbehave. It implies a lower cutoff level—allowing adverse selection of scientists— or a need of higher wages. Principal’s inability to create trust increases transaction costs and decreases principal’s welfare. 17

Finally, competition among scientists or the existence of actors willing to be researchers increases the utility of the principal, whereas scientists’ collusion decreases it. 18 Competition between scientists, and institutions enforcing it, is a prevalent feature of science as it is well documented since Merton’s early works on the reward system and struggle for priority (1957); it is one of the institutional solutions of those problems. Competition between principals, on the contrary, limits their utility by increasing the bargaining power of agents, 19 and can also threaten the benefit of some of the principals in favour of others if these collude with agents.

The preceding models are incomplete, yet. First, the principal is unable to evaluate even the published results because scientists have all the chances to hide—behind their esoteric language, for instance. He cannot reasonably be expected to possess the knowledge necessary for adequate supervision of the scientist’s results, in Shils’ words.

---

17 Ceteris paribus, polyarchies will do better than dictatorships. Governments in polyarchies are more capable to produce credible commitments because—and as far as— reduces officials’ opportunities for corruption through, on the one hand, government transparency and free social information that limits information asymmetries between incumbents and citizens, and, on the other hand, through periodical elections of some of the higher officials (I advance this argument in note 14 at page 7); and governments are the ultimate warranty of contracts through its monopoly on the legal use of force.

18 I follow Ferejohn (see 1986:18-22).

19 David (see 1998: 19-) elaborates the problems of that competition in early science; however, even at that time, patrons are often a handle in the world of science though researchers are innumerable and less powerful.
He cannot evaluate neither scientists’ qualifications or talent, up to a certain extent, nor their public actions: the cancelled Superconducting Super Collider, SSC, for example, would have it advanced American particle physics or would it have been too expensive a mean for doing this research?

Secondly, society’s utility is well served by scientists’ individual competition but science is a collective enterprise. While rewards must be personal to encourage good work, scientists need a lot of other’s. On the one hand, individual contributions to a specific result –experimental data, a theory– are often hard to distinguish. On the other, incentives foster individual misbehaviour and their recondite knowledge gives them, again, the upper hand, even against other colleges (Merton 1957). There is a strong problem for the agents’ community to limit collusive behaviour and individual misconduct –although spread misbehaviour could limit principal’s trust and destroy the mutually beneficial association. Every single researcher has strong incentives both to free-ride –when others fulfil–, and to prevent others’ misbehaviour. The problem is how.

Both, scientist and incumbent, are better off if they collaborate. An agreement can be reached about mutual control institutions, and both have incentives to offer a good contract to the other part. But, the main problem for carving each other trust is different from plain informational asymmetries: it is about mutually reliable evaluation of results within principal-agent framework. A specific result in science is not contractible ex ante; and its value or merit can change with time or with paradigms’ evolution. Principal must rely on evaluation by researcher’s colleagues; moreover, agents will not accept a different one. Institutions evaluating results and careers are, thus, the central institutions of science: institutions like reputation and peer review, and those that use this or similar procedures –boards, panels, committees and the like. The whole institutional framework will serve, then, all the actors’ mentioned purposes and it will avoid all the observed problems: mutually reliable evaluation of science results and careers.

---

Polanyi (1969) did a seminal contribution on this topic.

Sometimes appropriable results of research can be contracted ex ante; more often, agents incentive is the sharing of the results (embodied in patents); see the cited Aghion and Tirole (1994), and Dasgupta and David (1987). Novelty and originality of knowledge, however, can be contracted ex ante and certainly are.
within the institutional framework of principal-agent relationships with institutional warranties of mutually credible commitments, a system of checks and balances.

Next sections will develop how the institutions described by the sociology of science serve the interests of science’s patrons and employees.

Assumptions on scientific knowledge

Economics often consider products of science as commodities —of a very special kind, indeed. Knowledge is taken as a merchandise not only in science and technology issues but on most studies devoted to information and knowledge in the discipline. It seems useful as an analytical strategy: usually Economics does not need a different definition.

Less often, however, knowledge has been taken as a service. As far as ‘knowledge’ imply someone who knows about something, and as far as new scientific knowledge cannot be directly used by each one —without becoming an expert—, we hire scientists to use science for us. After Thomas Kuhn’s The Structure of Scientific Revolutions and his concept of paradigm, the community and the results of science cannot be taken separately.22 A practical implication is that scientific research produces both knowledge and those who know, people having the specific skills and knowledge to understand new results, and to apply them eventually. Research at the university, for instance, produces graduates able to use new knowledge, as well as research in a firm produces know-how to take advantage of new discoveries, even when it fails to achieve its own.23

The main consequence is that society depends on scientists to evaluate scientists and their work; moreover, depends on specific-field-experts to evaluate specific-field-experts and their results. After evaluation, the service scientists give is certification.24

---

22 Kuhn (1962). For previous work on this topic, see Fleck (1979); for the impact of Kuhn in social sciences, see Barnes (1982).

23 Current assumptions in Economics (see Dasgupta and David, 1987, and Dasgupta and Maskin, 1987) are that neither society nor particulars do profit for repeated discoveries, and that information has the attributes of a public good. The underlying assumption in research competition is “the winner takes all”. However, the relationship knowledge-specialist implies that the second-to-arrive derives some utility from work done in the way of tacit knowledge and know-how. It allows to participate in another round of invention and discovery —and thus society profit for having skilled people. Conversely, this public good cannot be used by everyone —only for those having the specialized and tacit knowledge to understand it.

Science institutions certify both knowledge and experts. Certifying institutions are prevalent in science and have mechanisms both to support certified knowledge delivering, and to limit delivery of unreliable or contested results. Science, again, is about mutually reliable evaluation and certification of results and careers.

Further, science results are preliminary. It takes some time for a discipline to accept a new result and to make use of it. It took Physics several years to acknowledge Einstein’s 1905 article on relativity, for instance. Criteria to provisionally accept a result depends on each discipline, and each paradigm within it. It is related to both theoretical and empirical concerns, to both originality in the border of the disciplinary matrix and to normal science results in the inner disciplinary core. Publication or citation does not mean, for a time, but a recognition for a minimum quality of the work done, or the merit of the attempt. Consequently, evaluation is always about preliminary results.

Following these assumptions I will propose a general description on the science exchange system that I will develop further in this article. Science consists in a swap of an specialized service for individual rewards, mediated by the reward system of science. The service consists in the production, understanding, certification, diffusion and use of original knowledge. Rewards –material and honorary– are indirect and are brought to individual researchers according to the recognition or credit their discipline ascribe to them; it depends, in turn, on individual contributions to this particular discipline.

Principal’s Science

In the patron’s point of view, the system must be established to limit moral hazard and adverse selection, and to maximize production of the highest quality at the lowest cost, indeed. Principal’s aims are to foster competition between agents and to avoid collusion; he needs to reveal enough steering information about agents— but not too much. The limitation is participation constraints when this highly skilled agent can opt for a different job. Agents’ utility should be equal at least to their utility in alternative contracts. Before, principal’s tools to drive science are designing an incentive system

---

25 Akerlof (1984:15) calls the certifying establishment to all the institutions of the education system (and thus some of the science system). I will discuss it later.

26 In the language of Constructivism, science is inherently a political activity — taking politics in broad terms.

through his payoff rule, and collecting information on agents’ effort and type –skills and talent. Both options are very costly in this case, as we have just seen: reputation, the very base of the system, is hardly measurable as an outcome.

Another tool is monitoring, using the payoff rule to induce behaviour driven by hidden information: one way is setting for the agent a low fixed wage –the fixed component of pay– and linking a variable payoff –the insurance pay– to an observable outcome. It use to be the case in science: a minimum entry salary, and incentive payoffs based on publications later. However, too low the fixed wage –or no fixed component at all– and the agent supports too much risk; conversely, too high the fixed wage and the risk is for the principal. In the republic of science, the payoff rule can be linked to publication: published papers are apparent and countable outputs.\(^{28}\) Journal articles become thus one of the main institutions of science and a very ritualised feature: they get a symbolic –social– value and a emotive –individual– one (Merton 1973). They take the place of the regular reports for principal’s information gathering and allow output measurability to set the payoff rule.\(^{29}\) Moreover, regular publication reveals part of agent’s private information and induces behaviour driven by agent’s private information: to maximize their output, agents will select research issues more suitable to their hidden skills and talent, and more profitable in a shorter time, given each scientific field constraints.\(^{30}\)

After monitoring, information collecting is much more easy for the principal.

However, publication does not fully guarantee quality, research is not paid by weight and research careers cannot be assessed in rough numbers. Some features are added for the institutional system of science to evaluate both qualitative and quantitative results. First, publication is previously assessed by author’s colleagues: peer review is a version of the widespread scientific institution of boards, panels, juries, tribunals, com-

\(^{28}\) A low entry salary, even below the mean for equally skilled graduates, limits adverse selection by filtering. Low talented agents who cannot trust on recovering later their personal investment will leave the field. However, problems of participation constraints remain; and agents’ discount factor should be very low, too.

\(^{29}\) Principals ask for a visible outcome as a way to share risks (see Eisenhardt 1989). Scientific papers are the best outcome the patron can have.

\(^{30}\) This is the quasi-market mechanism described both by Polanyi (1969), and also by Mulkay (1991:51-61).
mittees, and the like. In this case, referees are agents of the periodical’s editor whose interests in keeping its reputation on peer review management are aligned with those of scientists’ main patron. It helps to limit his transaction costs in evaluating output quality.

Secondly, scientific journals set minimum requirements for publishing aside of peer review. Journals also have a reputation for attainment, i.e., for quality. Thus, each published article can be taken as a measure of raw production, and publishing in some scientific periodicals can be counted as a measure of excellence. Coherently, scientists choose to read articles by journal or author’s reputation; and articles –and authors– gain some reputation based on the periodical where they are printed. Some organizations are developed then to collect statistical information about scientific publishing. Every single journal, article, author, organization, country, etc., can be valuated by the number of articles published, the quality or impact of the periodical, the number of others’ citations, and other combined measures.

Thirdly, periodicals are a scarce resource for researchers and it produces competition. “Publish or perish” is a well known proverb among researchers. As far as the principal sets publication in the very core of the payoff rule, he profits from one of the ‘natural’ institutions of science: competition and competition for priority (Merton 1957). It is fostered to arise principal’s utility.

However, competition for publication is but one aspect of competition for rewards in science. The principal can promote competition manipulating payoff rules: lowering fixed payoffs, as we have seen, and curbing variable payoffs. But science variable payoffs are rarely direct, not even those based on publications. The model is that of a long-term career, what is coherent with the long-term specialised training of scien-

---

31 I will develop this further. Peer review is an evaluation procedure in which panel members or referees are unknown for the author, and often vice versa.

32 Technical requirements have become also ritualised. The Institute for Scientific Information, ISI, lists the minimum requirements for a periodical to be considered scientific: regularity, peer reviewed approval of originals, international editorial conventions (“informative journal titles, fully descriptive article titles and abstracts, complete bibliographic information for all cited references, and full address information for every author”), and “English language article titles, abstracts, and keywords” (see ISI 2003).


34 The already cited ISI is the best known of those institutions, see note 32.
tists. Rewards consist in promotions along a clearly ruled chain of positions. Promotions are linked to personal reputation and merits, these are assessed by boards of colleagues, and based on publications—and on publications quantitative and qualitative measures. Both within an organization internal market or in wider external labour markets between different organizations, this is the ordinary mode of employment in the republic of science. Thus, principal’s manipulation of the variable payoff consists in limiting the number of higher positions or in establishing added conditions. Again, the main problem for this manipulation is participation constraint: utility of climbing to the upper positions compared with utility of going to alternative activities.

Competition for resources—other than printed surface in a journal—is also rife in science. It has increased with the growing costs or research equipment, particularly of experimental science: Big Science is but an aspect of a general trend. The government can use this competition to foster general competition in science. Most of the national systems of science distinguish rewards from funding, but this competition for resources based in project evaluation can be taken as a characteristic feature of modern systems of science. It is a mechanism parallel of that of reward assignment, and it works with the same tools and institutions: panels and peer review. The mechanism induces in turn two other institutions of science, the research fund and the research project, that I will develop later. Although it is partly driven by evaluation of work done and attainment reputation, the contest for funding of individual research projects and institutions alike are the place for Turner’s “prospective judgement on the credibility of promises” and the focal point of the practices of attestation he describes (1990:190-93). In the cycle of credibility of research it is an instrument to keep working and publishing and, thus, it makes the system of rewards more indirect and much more determined by agents’ competition.

---

35 Tirole (1994:6-) observes that long term careers with low-powered incentives are optimal for organizations producing hard to measure or non contractible goods (“experience goods”, for instance). In other page he suggest also “a mixture of direct and indirect rewards for information collection” because “information is often a difficult object to describe ex ante in an incentive scheme” (op.cit.:25).

36 Rip article explores how this difficult task is achieved (1994:10): “While Ziman is right, in principle, in saying that the attempt to judge the quality of future research is a ‘higher form of nonsense’, in practice a workable repertoire of how to handle proposals has emerged, as well as a certain competence of how to write fundable proposals”.

In this general framework, major principal’s concern would be about avoiding agent’s misbehaviour in the critical points: panels, boards, committees and all other gatekeepers that use peer review and similar procedures. It is easier, indeed, than driving each individual scientist or single element of the whole republic of science. Moreover, those places can be customized to produce information about the system: the bureaucracy to collect curricula, old projects and applications, databases on referees, bibliographies, bibliometrics and scientometrics, information on funding, and other steering relevant information is small and easier to control. The whole apparatus becomes easier to drive through the indirect system of rewards. This bureaucracy is the basis for any public system of science.

Scientists as agents, and the Republic of Science

Autonomy, at least relative, is an often recognised requirement of science, and it endorses science overall credit. It also justify the old myth of an independent republic of science: scientists themselves decide on research matters and participate on science policy decisions. Moreover, science organizations enjoy an independent status within government, and are often autonomous agencies. However, although relatively separated of society and politics, they have an influence on political decisions, they advise incumbents and public, and they help to check—and to made accountable—political actors on technical and non-technical issues.

But relative independence is only an instrument for researchers’ confidence. It is but one of the guaranties of principal’s fulfilment. The whole relationship is mediated: mutually agreed procedures and institutions protect agents’ decisions on research issues from principal’s direct command. As far as her utility maximization depends on private information based decisions, limits to free agent’s decisions decreases her utility—or, conversely, increases costs. The scientist should be free to choose research field and subjects, to choose collaborators, and to decide what, when, how and where to publish

37 Popularized by Michael Polanyi, it was first used by Charles Babbage in his On the Decline of Science in England (1830). See Ronayne (1984:76).

38 Relative independence collides sometimes with common public management and bureaucracies. Braun (1993:155) underlines this contradiction between an “administrative rationality” and a “scientific mode of operation” or “scientific mode of action”.

results—and consequently what, when, and where to teach. Limits to these freedoms, if any, should be subject to clearly established norms and procedures, within the limits of the discipline, and oversaw by colleges: all those are common provisions under incomplete contracts, which is the case. Being rewards individual, results should be individually attributed: we can take scientists signature, very ritualised also, as a key side of the institution of the paper. Finally, evaluation should always be left to scientist’s colleges, to those who understand and know the value of any specific contribution.

This collective evaluation by peers extends to the already mentioned critical points. Each scientist want to be evaluated by peers at panels, boards or tribunals when results publication, a research position or project funding are at stake. Conversely, each scientists want to be eligible in equal terms to those institutions, and those of the certifying establishments, which influence future opportunities and conditions of evaluation. Rules to create boards and to select its members are carefully negotiated and customized in each institution. Tight rules compensate for scientists’ extreme individual discretion. The goal is to limit misbehaviour, free riding and collusion—and this is also in the interest of the principal. Conflicts around those institutions frequently arise in the daily life of science (although they attract scholars’ attention less often than controversies or the struggle for priority), and causes most of the explicit and implicit rules in the republic of science. It all sums a good proportion of science legislation, but is so common and widespread that is often taken for granted.

Some of those general guarantees are written in laws above actors. These are particular solutions to the general problem of principal’s commitment: freedoms of speech or ideology, for instance, are written in Constitutions as well as more scientific and academic specific guarantees. Panels or committees and their rules are often under the ordinary justice or under specific conflict-solving institutions.

Laws, however, are not enough to explain government’s commitment and scientists’ attainment. Legal processes are much too slow for science’s day by day, and this rises transaction costs. Mutual attainment is usually solved through institutional arrangements. On the one hand, agents can offer an already tailored mutually satisfactory contract. In his work on American constitutionalism, Ferejohn (1999) explains how agents—the government, this time—had incentives to offer their principal—society—a
whole set of institutions to limit their action and to reveal hidden information. Scientists, for instance, regularly disclose information to their main patron, society, in the form of popularisation or other ways of knowledge sharing. Institutions that prevent their misbehaviour, depicted above, are part of this offer as far as they comply with science mode of operation.

On the other hand, solutions are in the institutional set configuration: a kind of checks and balances with separation of powers system. Government commits itself by separating, first, research execution from political decision and, then, technical management from political or scientific concerns. Science bureaucracies, specially in polyarchies, consist in chains of mediating autonomous institutions in three or four steps: independent laboratories or centres, autonomous research councils and finally the government. Within laboratories, scientists keep their relative autonomy to decide; in modern research systems this is fostered by external funding by project, which protects them from immediate superiors misbehaviour.

When we combine some of the described institutions in a relatively autonomous mediating bureaucracy we have a research council. Research councils are often embodied in large distinctive organizations that carry some or all of the tasks explained before: propose boards or committee members, back those boards credibility, collects information about the national (or sectorial) research system and every single element of it, etcetera. Councils also solve, partly, intertemporal inconsistency of governments. However, its main task is to distribute research funds through regular public contests of research projects: research councils are, thus, the chief institutions of modern public science systems. Arie Rip (1994:3) portrays them, following an OECD report, “as hesitating between a ‘parliament of scientists and a government bureaucracy’”, a government for the republic of science. I propose that research councils are both to mitigate the

---

39 Although in Ferejohn’s case agents have more opportunities to do it: here they can only suggest. See note 12 in page 6 on Ferejohn’s remarks to the prejudices of agency theory use.

40 The duty of scientific knowledge popularisation is firmly rooted in science formal and informal norms and in its professional ideology.

41 Tirole (1994:17-27) explains systems of checks and balances and separation of powers within governments and bureaucracy as a way to solve problems of intertemporal allocation of control rights (collusion, for instance, see op.cit. section 5), problems of mutual oversight (op.cit. sect. 7), and problems of information gathering or alternative information collecting (op.cit. sect. 8).

Rip’s thick description is useful here. Since research councils are agents both of the government and, in some way, of the scientists –those whom councils help to limit other colleges’ misbehaviour (op.cit.: 12)–, they have to win resources to remain trustworthy. The need to show themselves reliable is not so acute here that in the general government-scientists problem because this bureaucracy’s interests can be mostly perpendicular to its principals’. Its main inner task, create evaluative repertoires, is also not-conflicting with theirs. Here Rip underlines the role of reputation –a similar reputation to that widespread currency of the republic of science– as the means to manage trust.43 Research councils also try to win credit in their own credibility cycle, closely linked to those of scientists (op.cit.: 11-13), and the same happens with research funds. Another way is building the internal separation of powers, clearly separating technical or political advise by research evaluation offices and advisory councils from decision making, and this is aside of direct political command.

Research councils or funds, although having their own dynamics, operate coherently with the major principal-agent relationship of public science, the relationship between society and government with scientists, and its problems. I have tried to show that all other science institutions behave the same, and that other explanations for their need, shape and behaviour are secondary within an agency theory framework.

Conclusions and empirical implications

Science institutions serve the purpose of stabilizing actors behaviour, not only scientists’ indefinite actions by allowing enforcement, but also government and incumbents ones by backing credible commitments. For principals, institutions should foster

42 Guston again (1999, 2000) gives it the role of boundary setting as a science boundary organisation (he lengthily discusses around those concepts). However, it is hard to sustain that this boundary work has any influence in principal-agent problems (the theoretical framework Guston uses) or in actors’ interests and utilities –but secondarily.

43 I do not develop here the role of reputation in the world of science, although it epitomizes to me the institution in this world. Scientist’s reputation could be a feature of science state of nature, but helps principal’s control of research through fire alarms systems—as well as research council’s reputation (see McCubbins and Schwartz 1984). But the principal has also to win a reputation for attainment to have agents trust—and the same applies to research councils, again.
researchers’ “natural” competition for priority and resources, and should link rewards to measurable outputs—that made the contract results-based rather than behaviour-based, although outputs evaluation depends on agents yet. Also, institutions should help to reveal agents’ private information in a way that rends it abundant but not overflowing. Reputation and reputation seeking, for instance, facilitate control as it is linked with rewards, too. For scientists, in turn, institutions are long term guarantees of “technical rights” –closely related to their privately held information– and of fair evaluation and rewards assignment in formal institutions. Principal’s reputation is again one of those guarantees. Mutual surveillance at the system’s critical points under mutually agreed rules is built by both agents and principal in their mutual interest. But it is not enough under incomplete contracts.

One solution is that array of fire walls between incumbents and scientists shielding the republic of science from political power. It is not justified neither by the strong information asymmetries and high uncertainty of the case, nor by ad hoc explanations like as boundary building. It is easier, and more coherent with the agency theory general framework, to see it as a solution for principal’s commitment. It is a variety of separation of powers and checks and balances arrangements.

Empirically, it is difficult to observe in well developed public research systems how actors’ interests build them: some of the described institutions are simply taken for granted and neglected. However, major reforms give an opportunity to observe interests machinery at work, for instance, as well as the ex novo creation of new research systems. Historical accounts of the early science bureaucracies are chronicles of negotiations around social contracts, written in the language of principal-agent problems—as does Shils’ quotation show. Bush’s vehement Endless Frontier bump into Truman’s demands for accountability after Steelman’s report and Kilgore’s opposition. Smith (1992) explains it as a feature of American democratic culture, the feeling that public discussion is well over the scientific experts’ judgment. But before, the British Privy

---

44 In his explanation of the American science bureaucracy reform around 1950, Bruce Smith summarizes the trade off between government and society, and scientists demands (1994:42): “Most observers assumed that there [within the post-war pact] would be no serious conflict between civilian and military needs, education and research, or basic and applied science. The needs of society and of science would be served in such a way as to protect both the independence of science and its accountability to the general public”.


Council Committee for Scientific and Industrial Research—soon Department of State—and its Advisory Council were created autonomous, “in normal times, free from any serious pressure of administrative duties and immune from any suspicion of being biased against the application of research results by administrative considerations” (see Ronayne 1984:15). The mentioned Advisory Council, made of prominent scientists and industrials, had the mission of “(...) ensuring that scientists rather than bureaucrats determined the direction in which the funds would flow” (op.cit.:13). Science had to be, at the same time, free from bureaucrats influence and from social and political pressure, and accountable to political incumbents and to people. Similar discussions took place in other European countries at the inter-war period, and similar institutions were developed.

Transitions to democracy are another place to observe institutional building, indeed. Science organizations suffer major stresses to adapt to new accountability demands and, even more, to scientists demands of government’s commitments. If this general framework is correct, we should observe the building of checks and balances systems within public science general organizations or, at least, demands to do this in the scientific communities. Furthermore, we should observe autonomous institutions and autonomism devices more often in polyarchies than in dictatorships, and it should imply more research productivity of democracies (see note 14, page 7).

In science’s daily life of already formed public systems of science, those institutions should be the arena of conflicts. The game is the control of crucial points. Minor reforms would imply negotiations on actors’ roles around those institutions. The development of research planning systems has been taking place, in fact, through these key institutions of research councils and research funds by building new organizational guarantees—such as independent offices for research evaluation— or by deepening the research councils internal separation of powers. Major changes in social trust on science or in political interests, and thus on the emphasis on different areas or fields, have im-


46 See in Fernández-Carro (2002) a discussion about the different regimes influence on research and on science institutions.
plied changes in science organizations. However, its very existence—or the need of the institutions these embody—has never been questioned. 47

References


47 Changes that justify science systems ones and different theoretical periodizations —Guston (2000:140-45) attempts a summary of classifications of periods or eras based on changes of public trust or of science’s assigned mission.


selectionofmaterialforcoverage/199701.html (February 2003).


