ECONOMISTS:

TRUTH-SEEKERS OR RENT-SEEKERS?

Jesús P. Zamora Bonilla

Published in *Fact and Fiction in Economics*, edited by Uskali Mäki,

Cambridge University Press, 2002, pp. 356-374.

ABSTRACT: This paper proposes to extend economic analysis to the process of production of scientific knowledge in general, and of economic knowledge in particular. The model which is outlined here is based upon two fundamental assumptions: that the members of a scientific discipline can agree on a 'methodological constitution' (a system of methodological norms which make it compulsory to accept some data, hypotheses or theories when they have passed some predetermined tests), and that, when disagreement is allowed, the capacity to form coalitions among scientists tends to promote the epistemic efficiency of research. It is also argued that economic science presents some obstacles to the successful application of these mechanisms.

KEYWORDS: Economics of science; constitutional economics; public choice; rational expectations; scientific method; norms; coalitions; reflexivity.

1. THE CONSERVATIVE REVOLUTION MEETS REFLEXIVITY[1]

Although a vast majority of economic theories assume that human agents are rational utility maximisers, some economic schools are more willing than others to carry this vision of man to its ultimate consequences. In particular, two of the leading schools of the so called 'conservative revolution' of the seventies are famous for having made an extensive use of the hypothesis of the *homo oeconomicus*. These approaches were the Public Choice school and the Rational Expectations school. Both theories launched demolishing attacks to a couple of basic ideas underlying interventionist economic policies: the idea that economists could discover the working of the economic system, an the idea that this knowledge could and should be used by politicians to 'handle' or 'fine tune' the system so as to maximise social welfare. Stated differently:

1) both the economists and the politicians were assumed to behave in an altruistic fashion, not influenced by personal interests;
2) the fundamental equations describing the functioning of the economic system would neither be affected by its discovery, nor by the Government's intervention.

The Public Choice school directed its attacks mainly to the first of these assumptions, particularly against the standard view of the politicians' motivations. This school's fundamental thesis is that we cannot assume that economic agents are basically motivated by the maximisation of their own income, and, at the same time, that public servants are simply benevolent incarnations of the general interest. The State should rather be seen, according to Buchanan and his followers, as a revenue-maximising 'Leviathan', integrated by politicians and bureaucrats who will use their discretionary powers only in their own interest, at least to the extent that the limits established in the Constitution, or other constraining mechanisms, do not put stricter limits to that discretion. The power to tax, as well as the power to spend the revenue of these taxes, and the power of *not* to tax some privileged activities, could also be used by other rent-seeking agents to distort the otherwise efficient working of the markets, in order to benefit from artificial monopoly rents. The powers of the politicians and bureaucrats should be, hence, severely limited at the constitutional level, as a means of protecting the freedom and the welfare of the citizens.[2]

But, if we were to generalise seriously the *homo oeconomicus* hypothesis, so as to include under its scope *all* economic agents, it should obviously be applied to *economists* as well. If we can not trust politicians and bureaucrats when they talk on behalf of 'the public interest', can really we trust economists when they talk on behalf of 'the truth'? My argument simply replicates the story that ancient Greeks knew as 'the liar's paradox', or 'the paradox of the Cretan', attributed to the philosopher Epimenides: 'I'm a Cretan, and Cretans always lie'. Public Choice theorists seem to be saying us something like 'I'm a disinterested truth-seeker *homo oeconomicus*, and *homini oeconomici* are never trustworthy when they pretend to serve some interests which are not those of themselves'. The obvious moral is that, if individual greed is dangerous when it is not subjected to the discipline of the market or of the constitution, then economists' behaviour should be constrained either by a market-like mechanism or by a constitution-like mechanism, if it is going to produce reliable knowledge at all.

Economists tend to assume that scientific disciplines, including economics, actually function like a 'free market', in which competition entails that bad (*i. e.*, false) ideas and theories are replaced more or less automatically by good (*i. e.*, true) ones.[3] The lack of a broad consensus about very elementary topics in economics, and, more particularly, about the consequences of each kind of economic policy, shows that the working of such an 'invisible hand' is not very quick and efficient, at least within economic science. But, even if a deeper and broader consensus around a theory were observed, consensus would not be by itself a proof of the theory's correctness (after all, a general agreement might also be due to a 'monopoly' in the 'market for ideas', *i. e.*, to a lack of real competition). This means that it is not enough to *assume* that free competition of ideas would automatically lead to the truth; what we would like to have is a *proof* that this is the case, or at least we should have a plausible model showing that this can reasonably be the case.

The Rational Expectations school, on its turn, insisted more than other macroeconomic theories in the rationality of economic agents: according to this school, people make unbiased forecasts of the consequences of economic policies. The Rational Expectations hypothesis does not entail, as some of its critics have believed, that agents have a 'correct' model of the economy within their heads, only that they can learn to eliminate systematic errors in the ways they form expectations.[4] The 'correction mechanism' is supposed to be of an evolutionary kind, although, as far as I know, it has not been developed as much as, for example, the Darwinian mechanism proposed by Alchian and Friedman to defend the hypothesis that those firms which do not learn to maximise profits will not 'survive'.[5] The possibility of devising a specific evolutionary mechanism leading to unbiased expectations is, in any case, doubtful. In general, 'natural selection' does not guarantee that the equilibrium which is actually reached is an efficient one. In the case of expectations formation, if among the information processing systems which have been *actually developed* there were one leading to 'better' predictions than the rest, independently of which system the rest of agents had, then this system would be 'selected'; but this argument does not entail that the selected system is necessarily a 'rational' one, in the sense of being the best one among all the *conceivable* systems. Even if a mechanism producing 'perfect' predictions were discovered by some agents, it could be possible that an 'imperfect' mechanism were not superseded by the first one if the second happened to be an evolutionarily stable strategy.[6] After all, how could Rational Expectation macroeconomics explain, for example, the fact that the ('inefficient') Keynesian orthodoxy existed at all during two or three decades?

In fact, rather than explaining the formation of expectations by an evolutionary mechanism working on the cognitive systems of *individual* economic agents, it is reasonable to assume that, in a complex economy, the ability to forecast macroeconomic variables would be possessed by some group of *experts*, who would sell their services both to private agents and to public institutions. Instead of searching for an equilibrium in the evolution of cognitive systems, it would be more coherent with (new classical) economic theory the idea that actual macroeconomic theories, models and predictions are the outcome resulting from an equilibrium *in the 'market' for macroeconomic expertise*. More realistic macroeconomic models might contain, hence, an explicit description of that 'market', in the same sense that they can include a description of the money market or the labour market. Stated a bit differently: *economic knowledge should be seen, in part, as an endogenous variable of economic models*.[7]

In the following sections I will briefly describe and discuss some ideas that could be used in the construction of an economic model of the production of scientific knowledge. Section 2 presents an application of constitutional economics to the study of scientific *norms*, an approach which is illustrated in section 3 by means of two examples. In section 4 I will comment on the relevance of collective decisions in the mechanism producing a consensus about scientific *statements*. Lastly, section 5 discusses to what extent the mechanisms formerly described are applicable to economics as a scientific discipline.

2. SCIENTIFIC METHOD: A CONSTITUTIONAL APPROACH.

An economic model of the production of scientific knowledge must begin by making some assumptions about the utility function of scientists. I will assume that they mainly derive utility from getting their own theories or models accepted by their colleagues, and, in some disciplines, by the general public. Obviously, they also enjoy the production and possession of 'knowledge', for if they simply desired popularity and wealth, they would surely do better by devoting their efforts to sports, music, or other forms of entertainment. We can also suppose that scientists prefer having more income and wealth to less, and, at least in the case of economists, that they also get satisfaction from the political power of putting their own theories into practice.[8]

The last two elements in the scientists' preferences (income and political power) can nevertheless be seen as dependent on the first two ('recognition' among colleagues and among the public): the more popular your theories are, the more probable it is that you get political power and a high rent. The basic problem for a scientist is, hence, how to reach the first two goals, and this is perhaps a Herculean work: if convincing lay people that your theories are right may be a more or less difficult task, convincing your own colleagues *would be impossible* if they simply wanted (as you also want) to maximise the popularity of their own theories: accepting other colleague's theory would be like scoring a goal to their own goalkeeper! We do not need to assume that all scientists (nor even most of them) are absolutely reluctant to enhance their rivals' popularity; but I am trying to show that *even if scientists were so cynically motivated,* they could nevertheless reach an agreement on the norms of scientific method.

If each researcher within a scientific community intends basically to maximise the popularity of his own theories (both among his colleagues and among the general public), it follows that scientists only derive utility *in a direct way* from decisions taken by others. That is, each scientist does not receive his reward directly from *his* adoption of certain theories or statements, but from the theory choices made by *other* people. The only way of actively increasing his degree of recognition will be, hence, by exerting some influence on these choices. If we limit the choices under consideration to those made by scientists (not by the general public), this entails that science must be seen as a kind of *exchange*: in order to receive recognition, one has to offer something which fosters the recognition of others. A common line of thought among sociologists and economists of science is that scientists offer this recognition in exchange of the *information* provided by colleagues (especially by the person who has been the first one in disclosing that piece of information),[9] but this common idea has some problems. It particularly takes for granted that scientists value (correct) 'information', either *per se*, or because it serves them in their own pursuit of recognition. In the first case (right information is desired by its own), it would be unclear why researchers should worry about the 'true cognitive value' of a piece of information if, as radical sociologists of science have pointed out, scientists' fundamental concern is recognition *instead of* 'knowledge', and, more importantly, if the very idea of 'cognitive value' is nothing but a metaphysical chimera.[10] In the second case (right information is valued because its usefulness), accepting a piece of (right) information would only be useful for a researcher in his pursuit of recognition if *(i)* the other scientists also accepted that information, *(ii)* it tended to 'confirm' the propositions defended by the first scientist, and *(iii)* his colleagues were willing to explicitly recognise point *(ii)*; what needs to be proved is, then, why scientists accept many times information that 'disconfirms' their own theories, and why they devote effort to publicly recognise the 'authorship' of the information they are using, instead of simply using it without indicating its source.

In order to answer these questions, I propose to consider that what scientists exchange is not 'recognition for information', but the *mutual acceptance of constraints*: *they offer the compromise of subjecting their own theory choices to some rules, in exchange of their colleagues' observance of the same constraints*. These rules would function like the *methodological constitution* of a scientific discipline or of a scientific community, and they can be different for different groups of researchers, and even for the same group in different periods. This constitution is what makes 'the game of persuasion' possible at all, for a scientist could only try to persuade a colleague of accepting a theory if the latter's theory choices obey some determinate pattern, a pattern allowing the former to know what kind of arguments or strategies to use in order to persuade the latter.

The most essential aspect of the constitutional approach is its methodological individualism: it does not assume that the outcome of the social *interaction* among scientists is a social *entity* which we could call 'collective knowledge', and which would have the same analytical properties that knowledge, beliefs or information have in the case of individual scientists. Instead, the constitutional approach suggests to take a piece of information as 'public knowledge' among a group of scientists just if *every* member of that group publicly acknowledges that this information is acceptable according to a set of methodological rules that have also been *unanimously* accepted by the members of the group. If there is disagreement among a group about the norms that must be applied or about the outcome of their application, then our approach will simply identify a group's 'public state of knowledge about a certain issue' with the indication of what proposition (if any) each member of the group has publicly accepted concerning that issue.[11] Another peculiarity of the constitutional approach is that we need not necessarily be interested in finding out an *absolutely optimal* system of norms; our modest goal is, instead, to discuss whether a group of rational agents could reach a unanimous agreement about a *particular* norm or set of norms once they have agreed on looking for a norm of a certain *kind*.[12]

In order to participate in a collective decision about the acceptance of a rule, each member of a scientific group will basically take into account the probability that his own theories became accepted if a certain rule were adopted; note that many rules will usually be established *before* most theories, hypotheses and models in the scientific field are devised, as if they were adopted 'behind a veil of ignorance', and this will tend to make the choice of a rule more impartial; hence, the more uncertain the private benefits of alternative methodological norms are for a scientists, the stronger will be the influence of mere epistemic motivations in the choice of norms.

From the point of view of their role in the process of getting a theory accepted, these 'constitutional rules' can be classified as follows:

1) *Norms for theory comparison.* These rules tell under what circumstances can a theory, hypothesis or model be considered better than another, or under what circumstances will the epistemic value of a theory have increased or decreased. For

example, the norm indicating that having correctly predicted an unknown event increases the epistemic value of a model is a norm of this kind.[13]

2) *Norms of inference*. These indicate that, *if you have accepted some statements, you must also accept (or reject) another specific statement*. In practice, however, some statement-connecting norms will only make some combinations of statements more or less *untenable*, especially when the norms of theory comparison are not easy to apply to those cases, and hence, when it is not easy to establish which theory is better under the circumstances indicated by the inference norm. In any case, a researcher can benefit from his colleagues' adoption of some statement-connecting rules in an obvious way: if they obey a rule which tells (or from which it follows) that 'if you accept E, you must accept T_i', then, researcher i can increase the degree of acceptance of his own theory if he persuades many colleagues of accepting E in the first place.

3) *Action-related norms.* These tell scientists that, *if some people have performed certain actions in certain circumstances, and if they accept certain propositions, then they also must (or must not) accept another proposition, or they must (or must not) perform certain actions*. Rules of this kind entail that some statements must be accepted, not because other statements have been already accepted, but because of some additional reasons. These other reasons can be of a variety of types: they can be *results* of observations or experiments made by the researcher, or *reports* of observations or experiments made by others, or the *acceptance* of some statements by others (as in the case of many propositions one learns through textbooks, which in many cases are accepted as mere assertions of authority). Note that the individuals who have performed the actions stated in the antecedent of a rule of this kind need not be the same individual who is pondering whether to apply the rule or not. Note also that action-related norms not only refer to the obligation of accepting or not accepting a statement,

but also to the appropriateness or inappropriateness of performing *other actions*; for example, some of these rules can indicate how experiments must be conducted. Action-related norms play an essential role in the game of persuasion, since rules of inference only allow to persuade a colleague of accepting a theory if he already accepts some propositions that 'trigger' those inference norms.

4) *Enforcement norms*. These establish the penalty that has to be imposed on a scientist who has disobeyed some rule. Two important points must be indicated here. The first one is that, since we have supposed that the fundamental source of utility for a researcher is the degree of popularity of the theories proposed by him, the only kind of penalties that can be introduced in our model are those establishing that (some of) these theories must *not* be accepted. The second point is that enforcement norms are not only addressed to the infractors of the other kinds of rules, but to all the members of the scientific community, who, according to these norms, *have the obligation of not accepting* the theories proposed by the infractors; as a consequence, there will be some norms establishing a penalty for those scientists who have failed to apply a sanction when they had to do it.[14]

What can be said about these four kinds of norms from an economic point of view (as opposed, for example, to a sociological one)? In the first place, the *process of establishing* a norm (or a system of norms) can be studied as a negotiation in which a mutually favourable agreement is sought. For example, fig. 1 shows a situation where two scientists (or, perhaps, two different groups of equal size and power within a scientific community) have different preferences about the minimum degree of success that a theory or principle must have in order to be considered 'acceptable', that is, a norm of the second type. I will briefly discuss, in the next section, a formal model allowing to derive the utility functions used in the figure. For the moment, let us suppose that the 'strict' preferred to establish r^s as the minimum acceptable level of success, while the 'lenient' preferred r^l. *RU* represents the reservation utility: that which would be received if no agreement were reached and each scientist joined other community or found a different job. The 'strict' would neither accept a rule lower than a, nor lower than r^l (since r^l is better for both groups than any other point to the left); equally, the 'lenient' would not accept any point to the right of r^s or b. So, any rule between a and r^l would be a Nash equilibrium, and hence, a possible 'contract'.[15]

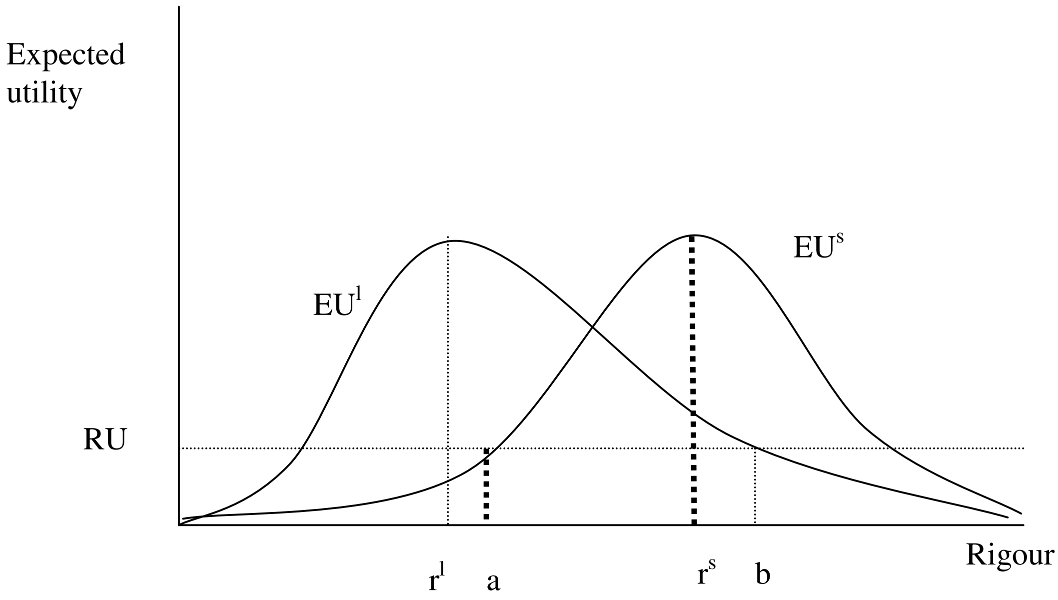

Figure 1

In the second place, we can also study the *efficiency* of a norm. In this context, the concept of efficiency has at least two different senses. In the first place, we can explore if the existing norms tend to produce situations where no scientist (or, by the way, other groups of citizens) can get a higher level of utility without someone else getting a lower one; *i. e.*, we can study if the norms are 'Pareto efficient' (for example, a norm above point *b* or below point r^s in fig. 1 would be inefficient in this sense). In the second place, we can investigate whether the existing norms of a scientific community tend to produce an optimum body of 'knowledge', at least according to the epistemic values of the members of the community (for example, a norm above point *b* would be epistemically better than *b* itself, though the incentives faced by the researchers preclude to reach an agreement so strict –if we grant scientists the liberty of choosing the norm). Another example of this concept of 'epistemic efficiency' would be the following: in the case of action-related norms, we can study the possibility of establishing ‘sincerity-promoting rules’ (or ‘strategy-proof mechanisms’), *e. g.*, ways of doing and reporting experiments, such that, if the agreement is obeyed by the others, each experimenter will find out that his best strategy is to report sincerely the result of his experiment even if this result goes against his own theory. This later kind of norms are of the highest importance, since they allow the construction of *intersubjective knowledge* out of the subjective experiences and private interests of each researcher. I will also succinctly present in the next section a mechanism having this property. The concept of epistemic efficiency will appear again in section 4.

Lastly, we can also study the *stability* properties of these norms, and especially the question concerning their self-enforceability. A minimum condition for a system of norms being self-enforceable is that it generates a coordination game, *i. e.*, one in which the best strategy for each player is to do what the others do. Of course, a scientific community can also reach an equilibrium in which not every member adopts exactly the same rules as the others: if different groups adopt different compromises, each individual scientist will consider the pros and the cons of joining one group or another, and his decision will depend, among other things, on the size of each group. Another possible application of our approach would be to consider the rules existing in a given community, not as the result of a consensus among its current members, but as something which has been ‘inherited’ by them from past generations, and that each generation can adapt to their own circumstances. The rigidity or flexibility of the norms may have a strong influence on their evolution, on the evolution of the scientific communities they regulate, and on the evolution of the knowledge produced by these communities. Hence, the study of scientific norms from an evolutionary perspective would also be of the greatest interest.

3. TWO EXAMPLES.

In this section I will offer two examples of 'constitutional methodological norms' that *could* be established among a scientific community. The first one refers to the choice of an inference rule, *i. e.*, the choice of an epistemic quality level or 'rigour' such that, if a theory happens to be the only one in surpassing it, it should be accepted by all the community members; I will call this level 'the acceptance standard'. Let U be the utility one scientists gets if his own theory is unanimously accepted by his colleagues, and let us assume that the utility he receives if his theory is rejected is zero; let $F(x)$ be the prior probability he has of devising a theory whose quality level is of x at most; and finally, assume that this probability is the same for the n members of the community. It is possible to proof that the expected utility of establishing x as the acceptance standard will be:

(1) $EU(x) = (1- F(x))F(x)^{n-1}U(N) + [1 - (1-F(x))F(x)^{n-1}]U(0) = (F(x)^{n-1} - F(x)^{n})U(N).$

The first order condition for the maximisation of this expected utility is:

(2) $EU'(x) = U(N)[(n-1)F(x)^{n-2}f(x) - nF(x)^{n-1}f(x)] = 0,$

and this entails that $F(x) = (n-1)/n$, or, what is the same, $1 - F(x) = 1/n$. That is, the members of a scientific community, were they motivated basically by public recognition, would desire that an acceptance standard were established such that the prior probability each member has of devising an acceptable theory is just the inverse of the number of community members.

If the inference norm does not command to suspend judgement if more than one theory happens to surpass the acceptance standard, but allows to accept *any one* of those theories, it can be proved that the preferred standard would be lower. On the other hand, if scientists are motivated also by 'knowledge', and they prefer *coeteris paribus* a stricter norm than a lower one (*i. e.*, if U depends positively on x in equations (1) and (2)), then it can be proved that the preferred standard would be higher. So, it is possible that not all scientists prefer exactly the same standard, and a process of negotiation like that described in figure 1 can take place; the standard which is finally chosen will have no epistemic virtue *per se*, besides the fact that, if it happens to be around the one established by equation (2), it will be a relatively strong one. Its main virtue is, instead, of economic nature: *it allows to establish an incentive system which makes it attractive enough for enough people the effort of scientific research*.

The second example I am going to present is a mechanism to allocate resources for the replication of experiments, one guaranteeing two things: a) that each replicator reveals the datum actually observed by him, instead of the datum predicted by his own theory (if he has one[16]), and b) that the resources are given exactly to the most skilful experimenters. I will suppose that the result of an experiment or observation can be summarised in a single magnitude, and that it is possible to define the 'distance' between any two different values of that quantity. I will also assume that each researcher makes the experiment or observation a number of times, using the resources he has been endowed with, and derives a certain statistical distribution of the possible values of the quantity, as well as a distribution of the possible values of the mean of all announced results. Lastly, I will suppose that the two distributions are symmetric.[17] The proposed mechanism is the following one:

a) each researcher decides whether to perform the experiment or not, and, if he decides to perform it, he has also to decide whether to reveal a result or not; all the researchers who decide to reveal a result do it at the same time;

b) if the difference between the mean of all revealed data and the datum revealed by a researcher is bigger than d, then this scientist gets a 'penalty' which decreases his utility level in V; if the difference is lower, then he receives a public recognition which increases his utility level in U; if a researcher does not reveal any result, his utility level does not change;

c) the financial and material resources the community can devote to making the experiment are allocated in the following way: every researcher (or team) announces the minimum margin of error d he is able to accept (based on his previous experience as an experimentalist), and the resources are distributed beginning by the researcher who has announced the minimum d, until they are exhausted; the margin to which all researchers are submitted corresponds to the one announced by the last researcher who has been endowed with resources;

d) all the members of the scientific community must accept as the value of the measured magnitude the mean of the revealed data, with a margin of error that depends on their variance.

It is possible to prove that the most rational strategy for each experimenter under this mechanism is to reveal sincerely both the minimum margin he is able to commit himself, and the true outcome of his replication of the experiment.[18] The first of these properties allows the scientific community to allocate the resources among the people who is most able of performing the experiment well, whereas the second one allows to trust the experiment's revealed results.

The application of some constitutional norms such as the two commented on above can warrant that the knowledge produced by a scientific community is relatively 'objective', bot in the sense of being impartial and of having been subjected to some strong quality tests. I am not asserting that exactly these mechanisms exist in the 'cognitively successful' scientific communities, nor that these 'constitutional' rules should be established within an 'assembly'. But I would like to suggest that those disciplines which have managed to establish some norms with analogous properties, are those which are also most successful as suppliers of reliable knowledge.

4. COLLECTIVE DECISIONS AND THE DYNAMICS OF SCIENTIFIC KNOWLEDGE.

The 'constitutional' rules of scientific method described in the previous section tend, if they are well defined and enforced, to promote agreement about data, hypotheses, laws or any other kind of scientific statements, even if the dominant (though not exclusive) desire of scientists were recognition. Nevertheless, we can not assume that actual methodological rules (particularly when we think about the norms of a concrete discipline or subdiscipline) are as well established as other types of legal rules are, since they are usually tacit, rather than encoded, and also because they are usually difficult to apply unambiguously to singular cases. This means that, even if the members of a scientific community have mutually agreed in constraining their choices of statements to some common rules, there is still the possibility that not all of them accept the same data, models or theories, although their methodological compromises may tend to reduce this variation, as compared to what would have been the case if those shared rules had not existed at all.

If methodological constraints are not enough for generating a full consensus around one solution to a scientific problem (or, more radically, if no unanimous agreement has been reached about methodological norms), we can legitimately ask what factors determine the actual degree of acceptance of each alternative solution among a scientific community. Elsewhere I have proposed a model which identifies three factors making a scientist more willing to accept a given statement:[19] the epistemic value the statement has for him (what will mainly depend on the interpretation he makes of the shared rules and of their application to that problem), the degree in which the statement is favourable to the theories proposed by him (again according to those rules), and the degree of acceptance of the statement within the scientist's community. The inclusion of this last factor needs some justification. In the first place, for many statements a scientist needs to use during her work, the 'private' information he will have about them will be very poor, and he will simply be unable of performing his work without trusting what his colleagues assert; so, the more agreement they will show on a certain statement, the less probable it will be that all of them have misleading information about it.[20] In the second place, communication among scientists tends to be easier the more 'knowledge' they share; hence, if communication is essential while trying to persuade a colleague, then you will need to accept a high proportion of the statements he accepts. In the third place, if you want others to accept the hypothesis you are proposing, then you will need to show that the methodological norms they accept entail that your hypothesis has an epistemic value high enough, *given the other statements they accept*; so, the more accepted is a statement among your community, the stronger will be the incentive you have to devise a hypothesis which is *supported* by that statement, and hence, the more interesting will be for you to accept it.[21]

Since the details of this model of theory choice can be examined in the papers referred to in note 19, I will only indicate here the more interesting consequences. In the first place, if the application of the constitutional constraints does not make compulsory to accept a particular solution to a given scientific problem, there can be more than one stable equilibria (*i. e.*, more than one vector $x = (x_1, \ldots, x_n)$, such that x_i is the degree of acceptance of the i-th solution among the scientific community, and that x is a steady state in the dynamics generated by the mutual interdependence of the scientists' decisions, *i. e.*, one where the desired decisions of all scientists are mutually consistent –a Nash equilibrium– and which is an attractor for other vectors which are not Nash equilibria). In the second place, under some circumstances, some of these steady states will be Paretian-inefficient, *i. e.*, it is possible that x is a stable equilibrium but there is another distribution of degrees of acceptance under which no scientist would have a lower level of utility, but al least one of them would have a higher level. In the third place, it may also happen that every scientist believes solution i is epistemically better than solution j, but there is a stable equilibrium in which $x_i < x_j$ (we can refer to this situation as one of 'epistemic inefficiency'). Lastly, the existence of multiple equilibria can make of the evolution of scientific knowledge a strongly *path-dependent* process, since the actual degree of acceptance of a scientific hypothesis will not only depend on epistemic factors (*i. e.*, each researcher's beliefs about its scientific merits, given the information he has) and on social factors (*i. e.*, the 'conformity effects' described in the preceding paragraph, as well as each researcher's desire of recognition), but also on the *order* in which the relevant information has been produced and disseminated, and on the *order* in which the competing hypotheses have been devised and publicly proposed.

These four results are really bad news for those who would want to use economic theory as an instrument to show the 'efficiency' or 'rationality' of scientific research, since they confirm some of the main points defended by radically relativist sociologists of science, or so it might seem at first sight. Nevertheless, I think that the problem only emerges due to the fact that we have modelled scientists simply as passive automata myopically adjusting to the situation they confront. If we consider, instead, that researchers have the possibility of being conscious of the 'problems' caused by their own decisions, as well as the capacity of 'negotiating' a solution to those problems, we can show that *collective action* among scientist can help to make the production of scientific knowledge a much more rational enterprise. Obviously, the constitutional choices referred to in sections 2 and 3 were also collective decisions; the main differences between that case and the present one are the following: in the first place, the constitutional choice would be a unanimous decision of all the members of a scientific community, whereas now I will study collective decisions taken by groups smaller than the full community; in the second place, the objects of the constitutional choice were methodological *norms*, and now we will consider decisions about to accept or to reject a scientific *statement*.

Suppose, in the simplest possible case, that the members of a scientific community have to choose between accepting a hypothesis or rejecting it,[22] and assume that there are two possible stable equilibria, x and y, corresponding to two different degrees of acceptance of the hypothesis.[23] If each researcher just decided *individually* whether to accept or to reject the hypothesis, then any of the two equilibria could be reached (for example, x), depending on the situation from which they started. But if scientists realised that there is another stable equilibrium (what could be possible thanks to fluent communication among researchers), and if some of them realised that they would reach a higher utility level in y than in x (among those who reject the hypothesis in x but accept it in y), this group of scientists could form a 'coalition' and change their former *individual* decisions of rejecting the hypothesis for a *collective* decision in the opposite sense. The last decision is 'collective' in the sense that it is only interesting to adopt it for each member of the group if the other members do the same; *i. e.*, the members of the group face a coordination problem in taking the decision. It can be proved that the collective compromise of accepting the hypothesis is self-enforceable if and only if the size of the coalition is bigger than the absolute difference between x and z (being z the unstable equilibrium lying between x and y; see note 23). If the coalition is formed and its decision is self-enforceable, then other scientists who where rejecting the hypothesis in x would find more interesting to accept it, what would induce more scientists to do the same, and so on, until equilibrium y is reached. This result also entails that there will be a possible coalition having enough power to 'jump' from equilibrium x to equilibrium y if and only if there is no possible coalition with the power of doing the reverse.

The possibility of 'jumping-at-once' from one stable equilibrium to another entails that, at least under the circumstances considered by the existing models, only one stable equilibrium remains, what eliminates the problems of underdetermination and path-dependence. Paretian inefficiencies are also eliminated, since those researchers who were better in equilibrium y than in x could form a coalition to 'jump' to y, if the actual equilibrium were x. The only remaining problem is epistemic inefficiency, although this is only due to our assumption that researchers can take into account, in order to accept or reject a hypothesis, not only their private opinions about it and the public pronouncements of their colleagues, but also the degree of support the theories proposed by them receive from that hypothesis; in the absence of this motivation, epistemically inefficient equilibria would be eliminated as well (more formally, if every scientist thinks that hypothesis *H* is better in the epistemic sense than *H'*, it can not be the case that the, in the equilibrium which is stable under collective decisions, *H'* has a higher degree of acceptance than *H*). Nevertheless, the formation of coalitions can have some *costs*: time must be devoted to talking and discussing with colleagues, and a risk is assumed by each member, for it is not sure that their collective decision will in the end be successful (for example, the size of the coalition might be too small). So, the possibility of a fluent communication among scientists is essential in allowing that their collective decisions may solve the problems generated by conformity effects.[24]

5. IS ECONOMIC KNOWLEDGE EFFICIENTLY PRODUCED?

Here we have considered science as an institution devoted to the production of 'knowledge', but integrated by people who are not exclusively (and perhaps not fundamentally) motivated by the 'pursuit of knowledge', but for the attainment of other personal goals. In a sense, society faces here a problem of the 'principal-agent' type: how can the people whose money is employed to fund scientific research (the 'principals') make scientists (the 'agents') behave as if they were truth-seekers, instead of mere rent-seekers? Some authors have argued that something like the 'invisible hand of the market' would also exist in the case of science, making more interesting individually for each researcher to look after valid results than to present erroneous or fraudulent ones, and to defend those theories which he personally thinks are better justified than those which better help him to reach his own private goals. Perhaps some economic models showing this possibility can be devised, but I suspect (for reasons stated in section 2) that only an agreement about scientific *norms –i. e.*, an exchange of *constraints*, instead of a mere exchange of 'actions'– can make the scientists' decisions be consistent with the pursuit of knowledge. In fact, even in the case of the market the 'invisible hand' only works because there is a legal order which defines and enforces property rights. On the other hand, if we choose to model scientists as individual utility maximisers, and ignore the possibility of negotiation and collective decision making, then all seems to indicate that equilibrium analysis will lead us to conclude that the production of scientific knowledge is not as rational and as efficient as the 'invisible hand' metaphor might suggest.

In the case of science, 'constitutional' rules would define what can, can not or must be taken as a solution to a problem, or as a better solution to another one, and hence, those norms *create* the game that scientists will later play, using either individual or collective strategies. From the point of view which is defended in the present paper, the more freedom scientists have to establish a system of methodological norms (and the more severe the chosen norms are), and the more freedom scientists have to communicate, negotiate or establish coalitions with colleagues, the more probable it will be that the 'scientific knowledge' produced by them be of a high quality *according to the epistemic criteria of scientists themselves*. I do not think there is, in general, any other criterion of epistemic value which could be more acceptable than consistency with the epistemic part of scientists' preferences, for the specialists in a scientific discipline have, after all, a deeper sense of what can be taken as 'knowledge' within their own field. The main problem of an institutional organisation of science *from the cognitive point of view* is, then, how to make scientists *strive* for reaching theories which have the highest possible epistemic value in their own opinion, and how to make them *sincerely reveal* their true opinions about the theories actually proposed.

The question is, hence, how does all this affect our discussion on economic knowledge? Is economics, as a scientific discipline, constituted in such a way that the mechanisms analysed in the previous chapters apply to it? Instead of offering a list of possible 'cognitive shortcomings' of economic science, I propose to take the fundamental elements of our collective-action model of scientific research and see whether they can be attributed to economics. In the first place, it is hard to deny that economics, particularly mainstream economics, is an activity regulated by norms, *severe* methodological norms. Papers published in 'top' journals must have passed strong quality tests, perhaps no less strict than in the 'harder' sciences, and in general there is little doubt among the profession about who are the most 'eminent' practitioners, those who have produced the 'best' theoretical or (less frequently) empirical work. But it is doubtful that the *aim* of those tests, and of the criteria used to identify who is a good economist, is specifically *to determine the compulsory acceptance of the content of published models or theories*. The main function of the methodological norms considered in sections 2 and 3 was simply to make it possible that a scientist becomes the winner in a race for the solution to a problem, and *being* the winner entails that all (or almost all) your colleagues *explicitly accept you're right*. In economics, instead, accepting that a model or theory is terribly good (even much better than its rivals) does not force you to recognise that the model or theory is (even approximately) *right*. In the majority of cases, what makes of an economic model a 'good' one (in terms of the quality rankings of the profession) is, instead, the mastery of analytical skills revealed by its creator. Economics is in this sense more similar to the fine arts or to sports than to other scientific disciplines: we will found out in it severe norms, and a fierce competition as well, but the losers in that struggle will not have the obligation of *adopting* the winners' work.[25]

In the second place, if methodological rules are not enough to promote scientific consensus in economics, will collective decision making avoid, at least, that epistemic inefficiencies take place within the discipline? Two factors can make it difficult that this mechanism is functioning in economics as efficiently as it may do in natural sciences. The first factor is that the reasons to accept an economic theory, model, hypothesis or datum may not only be the ones referred to in section 4: besides epistemic assessment, support of one's preferred theories, and conformity to others' decisions, an economic statement can also be accepted by an economists because it tends to promote in some way his economic, political or social interests. For example, usually the effects of economic policies on the general welfare are very uncertain for individuals, but *some* of their consequences for *some* agents are very clear: for example, the wealthy may not be sure about the *general* and *long term* effects of a big cut in the marginal rates of the income tax, but they will be very sure about some *particular* effects of this policy, and they may tend to accept and disseminate a theory which proposes a measure of this kind. Even if a situation happened in which an epistemically inefficient equilibrium were actualised and a possible coalition existed which were interested in passing to an efficient equilibrium (*i. e.*, if they were interested *for epistemic reasons* in 'jumping-at-once' to the acceptance of, say, an heterodox theory), this 'jump' would probably have *real* effects on economic policy, or on the working of the economic system, and the agents who thought that they were going to be harmed by this change might provide some incentives to economists in order to prevent them to 'jump'. Hence, since the consequences of economic theories almost always refer to interest-laden problems, it seems that a good strategy would be not to trust too much in those theories until a detailed argument is offered showing that the people who have produced and disseminated that ‘knowledge’ have made it so under a system of incentives guaranteeing the neutrality and objectivity of their conclusions.

The second factor making more difficult in economics that coalition formation tends to eliminate epistemic inefficiencies, is the fact that the ‘economic knowledge’ which is more relevant to take into account is not only that of economists, but the economic beliefs, intuitions or expectations of *economic agents* (*i. e.*, consumers, firm managers, politicians, and so on).[26] It is at least conceptually possible that some economists discovered ‘the true laws of the economic system’ (if such a thing existed at all), but no one else believed them. Since the evolution of the economy will depend more on the ‘opinions’ each agent has than on the ‘truth’ possessed by a minority within an ebony tower, a truly ‘reflexive’ economic theory (as, for example, Rational Expectation Macroeconomics pretended to be) should not automatically assert that people has true, unbiased knowledge of the relevant economic variables and mechanisms; this could only be accepted after having shown why conformity effects and ideological biases *could not* lead to stable equilibria in which some radically false economic beliefs survived and had a high degree of popularity. If these epistemically inefficient equilibria existed in the economy at large (and not only within economics as a discipline), it would be much more difficult to create a big enough coalition which could ‘jump’ to an efficient equilibria, simply because it would probably involve millions of people, and not only a few dozens (as in the case of scientific coalitions).

In conclusion, the credibility of economic knowledge depends basically on the incentives faced by those people who produce it, disseminate it and make use of it. As this credibility is growingly contested from some quarters, I think economists should devote more effort to justify the objectivity of their theories and the reliability of their recommendations. It can be argued that ‘the burden of the proof’ should be on the accuser's side, but actually no accusation is being made here: only the conceptual possibility of epistemic inefficiencies is indicated. It should be interesting that economic theorists provided an account of how an *objective* economic knowledge could be reached, as well as an empirical determination of whether the actual practice of economics corresponds to that theoretical account. In order to do so, the elaboration of an 'economics of scientific knowledge' seems to be an essential previous stage. I believe that deep and radical changes in economic theory can be expected if this programme is taken seriously by the profession.

REFERENCES

Axelrod, R.: 1997, *The Complexity of Cooperation*, Princeton, Princeton University Press.

Banerjee, A. V.: 1992, "A Simple Model of Herd Behavior", *The Quarterly Journal of Economics*, 107, 797-817.

Brennan, G., and J. M. Buchanan: 1985, *The Reason of Rules. Constitutional Political Economy*, Cambridge, Cambridge University Press.

Brock, W. A., and S. N. Durlauf, 1999, 'A Formal Model of Theory Choice in Science', *Economic Theory*, 14, 113-30.

Colander, D.: 1989, 'The Invisible Hand of Truth', in Colander and Coats (1989), pp. 31-36.

Colander, D. and A. W. Coats: 1989, *The Spread of Economic Ideas*, Cambridge, Cambridge University Press.

Dasgupta, P. and P. A. David: 1994, 'Toward a New Economics of Science', *Research Policy*, 23, 487-521.

Elster, J.: 1989, *The Cement of Society: A Study of Social Order*, Cambridge, Cambridge University Press.

Garnett, R. F. (Jr) (ed.): 1999, *What do economists know?*, London, Routledge.

Goldman, A. I., and M. Shaked: 1991, "An Economic Model of Scientific Activity and Truth Acquisition', *Philosophical Studies*, 63, 31-55.

Hands, D. W.: 1997, '*Caveat Emptor*: Economics and Contemporary Philosophy of Science', *Philosophy of Science*, 64 (proceedings), S107-S116.

Hardwig, J.: 1991, 'The Role of Trust in Knowledge', *Journal of Philosophy*, 88, 693-700.

Hull, D. L.: 1988, *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*, Chicago, The University of Chicago Press.

Kitcher, P.: 1993, *The Advancement of Science: Science without Legend, Objectivity without Illusions*, New York, Oxford University Press.

Latour, B.: 1987, *Science in Action: How to Follow Scientists and Engineers through Society*, Philadelphia, Open University Press.

Latour, B., and S. Woolgar: 1979, *Laboratory Life: The Social Construction of Scientific Facts*, London, Sage.

Lucas, R. E. (Jr): 1987, 'Adaptive Behaviour and Economic Theory', in R. M. Hogarth and M. W. Reder (eds.), *Rational Choice: The Contrast between Economics and Psychology*, Chicago, University of Chicago Press, pp. 217-242.

Mäki, U.: 1999, 'Science as a Free Market. A Reflexivity Test in an Economics of Economics', *Perspectives on Science*, 7.4, 486-509.

Slembeck, T.: 2000, 'How to Make Scientists Agree. An Evolutionary Betting Mechanism', *Kyklos*, 53.4, 587-592.

Vega-Redondo, F.: 1996, *Evolution, Games and Economic Behaviour*, Oxford, Oxford University Press.

Vromen, J. J.: 1995, *Economic Evolution*, London, Routledge.

Wible, J. R.: 1998, *The Economics of Science. Methodology and Epistemology as if Economics Really Mattered*, London, Routledge.

Zamora Bonilla, J. P.: 1999a, 'Verisimilitude and the Scientific Strategy of Economic Theory', *Journal of Economic Methodology*, 6, 331-350.

Zamora Bonilla, J. P.: 1999b, 'The Elementary Economics of Scientific Consensus', *Theoria*, 14.3, 461-88.

NOTES

[1] Financial support from the Fundación Urrutia Elejalde, from Spanish Government's DGICYT research project PB98-0495-C08-01, and from the Department of Economics of the Universidad Carlos III, has encouraged the writing of this paper. Some preliminary versions of it were presented in the Permanent Seminar of Economic Methodology of Universidad Autónoma de Madrid (November 1999), in the Seminar of Economics and Philosophy of the Cátedra Sánchez-Mazas (San Sebastián, Spain, 1999), and in the Seminar 'Theoretization and Experimentation in Economics' (Rovaniemi, Finland, December 1999). My thanks to Juan Carlos García-Bermejo, Andoni Ibarra, Uskali Mäki and Timo Tammi for inviting me to take part in them. I have also received helpful comments from Francisco Álvarez, Shaun Hargreaves Heap, Frank Hindriks, David Teira, and especially Juan Urrutia.

[2] See, for example, Brennan and Buchanan (1985).

[3] See Colander (1989), and, in general, the papers contained in Colander and Coats (1989). See also Wible (1998) for a convincing criticism of the idea that 'science is a free market'.

[4] Lucas (1987).

[5] After all, in what a sense would not 'survive' an economic agent who were not able to predict the rate of inflation in an unbiased way?

For a survey of the literature on economics and evolution see Vromen (1995).

[6] See, for example, Vega-Redondo (1996).

[7] The same idea has been recently defended in Mäki (1999), where he proposes to consider the applicability of an economic theory to itself (*i. e.*, its capacity to explain its

own degree of popularity and its –assumed– epistemic superiority with respect to alternative theories) as an additional test each theory should pass.

[8] Of course, the mere fact that somebody pursues private benefits, amongst other things, does not make of him automatically a 'rent-seeker'; by a 'rent-seeking scientists' I would mean, rather, one who used (at least a part of) the resources devoted to science in a manner which is only profitable for him, and not for the other members of the society (as long as these are interested in funding science for producing reliable knowledge). So, the concept of 'rent-seeking' only refers to the *behaviour* of researchers, no to their motivations: the pursuit of a high income or of political power will not make of a scientist a 'rent-seeker' if he acts in a manner which is compatible with the pursuit of the highest amount of knowledge.

[9] See, for example, Hull (1988), esp. ch. 10, and Dasgupta and David (1994), sec. 4.

[10] See, for example, Latour and Woolgar (1979). Personally, I do not accept these radical assumptions, but I am trying to show that even if scientists were as depicted by those authors, they would be interested in establishing some 'objectivity promoting' methodological norms.

[11] By recognising this essential methodological individualism, the constitutional approach avoids the problems rightly detected by Wade Hands (1997) in other 'naturalistic *cum* economic' explanations of scientific knowledge (such as Goldman and Shaked (1991) or Kitcher (1993)): our approach does not 'want something to emerge (a special type of belief) that is qualitatively different from the beliefs of the indidivual agents' (Hands (1997), p. S113), and it does not intend, as well, to construct a notion of 'public knowledge' as a kind of 'aggregate' having the same properties of individual beliefs.

[12] It is almost sure that no human group can devise and study all *possible* kinds of norms they might adopt; but once some *type* of norm has been proposed within a group, it is much esier for its members to discuss what *especific features* they would desire the chosen norm had.

[13] I have defended elsewhere that some prevailing methodological norms correspond to those that would be chosen by a group of scientists satisfying two conditions: a) the epistemic element in their utility functions corresponds to what I have called 'empirical verisimilitude of a theory given the existing data', and b) they choose the rules withouth knowing what probabilities of success *their own* theories will have under those rules (*i. e.*, they choose 'under the veil of ignorance'). See Zamora Bonilla (1999a).

[14] Enforcement norms are, in this sense, 'metanorms'. Axelrod (1997), ch. 3, shows that introducing this kind of norms in computer 'prisonner's dilemma' tournaments drastically reduces the frequency of 'defections'. On the other hand, the justification of obbeying a metanorm is not clear from the point of view of instrumental rationality; see, esp., Elster (1989), ch. 3.

[15] Of course, the positions of the four relevant points can change, depending on the shapes of the expected utility functions and on the reservation utility levels (which, in adition, need not be the same for each group).

[16] Or if there is one datum that is more 'favourable' for him in some sense.

[17] The results depend essentially on the symmetry assumption; it would be an interesting analytical problem to show whether some mechanisms with the same properties exist if this assumption is relaxed.

[18] A detailed proof is available from the author.

[19] Zamora Bonilla (1999b). Brock and Durlauf (1999) also present a similar, though mathematically more sophisticated model.

[20] Cf., for example, Hardwig (1991).

[21] See, *e. g.*, Banerjee (1992) for the economic modelling of conformity to others' behaviour.

[22] I am using the term 'hypothesis' in the widest possible sense, as a proposition of any level of generality whose truth is not known with absolute certainty.

[23] In this case it can be proved that there will also be an *unstable* equilibrium *z*, between *x* and *y*. For all the proofs referreds to in this paragraph, see Zamora Bonilla (1999b), pp. 470-71, and pp. 477-80. For the problem of stability, cf. also Brock and Durlauf (1999), p. 121, though they ignore the possibility of collective decisions, which is basic in my argument.

[24] The change in the degree of popularity of a hypothesis due to the collective decision of a coalition can be related to Bruno Latour's arguments about the role of 'enrolling allies' to make a theory triumph (cf. Latour (1987)). What Latour would probably not accept is the efficiency promoting character of these collective decisions.

[25] It is particularly sad the fact that empirical evidence is used so little to actually *resolve* economic disputations. For a recent proposal of establishing a kind of empirical competition among economic models, as well as a recognition allocation mechanism based on its results, see Slembeck (2000). I have tried to explain the relative lack of empirical testing in economic theory in Zamora Bonilla (1999a).

[26] The relevance of this 'everyday' economic knowledge has been emphasised recently in the articles contained in Garnett (1999).