

Scientific Inference and the Pursuit of Fame: A Contractarian Approach*

Jesús P. Zamora Bonilla^{†‡}

UNED, and Fundación Urrutia Elejalde. Spain

Methodological norms are seen as rules defining a competitive game, and it is argued that rational recognition-seeking scientists can reach a collective agreement about which specific norms serve better their individual interests, especially if the choice is made ‘under a veil of ignorance’, i.e., before knowing what theory will be proposed by each scientist. Norms for theory assessment are distinguished from norms for theory choice (or inference rules), and it is argued that pursuit of recognition only affects this second type of rule. An inference rule similar to ‘eliminative induction’ is defended on the basis of such a possible agreement. According to this contractarian approach, both the explanation and the justification of scientific norms only need to refer to the preferences of individual scientists, without assuming the existence of ‘collective’ points of view.

1. Introduction. Despite a long tradition of understanding the scientific method as a matter of careful application of logical rules, it is now commonplace to view methodological norms as conventions. One of the philosophical approaches which contributed most decisively to the acceptance of this position was falsificationism. According to Popper, not only

*Received October 2001; revised December 2001.

[†]Send requests for reprints to the author; Depto. Lógica y Filosofía de la Ciencia, UNED, 28071 Madrid, Spain; e-mail: jpzb@fsf.uned.es

[‡]Financial support for this research was received from Fundación Urrutia Elejalde, from The Spanish Government’s DGICYT research project PB98-0495-C08-01, and from the Department of Economics of the Universidad Carlos III. Previous versions of this paper were presented in the Seminar of Economic Methodology of Universidad Autónoma de Madrid (November 1999), and in the Seminar ‘Theoretization and Experimentation in Economics’ (Rovaniemi, Finland, December 1999); my thanks to Juan Carlos García-Bermejo, Uskali Mäki and Timo Tammi for inviting me to take part in them. Helpful comments were made by Francisco Álvarez, José Ramón Álvarez-Rendueles, Salvador Barberá, Miguel Beltrán, José Luis Ferreira, Shaun Hargreaves Heap, Frank Hindriks, Javier Ruiz-Castillo, David Teira, and especially Juan Urrutia and two anonymous referees.

Philosophy of Science, 69 (June 2002) pp. 300–323. 0031-8248/2002/6902-0008\$10.00
Copyright 2002 by the Philosophy of Science Association. All rights reserved.

the general rules of method had to be justified on the grounds of their contribution to the aims of science, but their application to concrete cases always needed some decision making, not determined by methodological rules alone. In particular, the acceptance of 'basic statements' was a provisional decision, *potentially* subjected to further criticism, but which was conventionally settled in order to stop the *actual* process of testing (e. g., Popper 1959, sec. 11). Likewise, for Lakatos not only the acceptance of basic statements was conventional in this sense, the commitment to a given research programme was also a conventional decision: even under appalling contrary evidence, defenders of a 'degenerating' programme could rationally support it, if they were confident in a future solution of its present anomalies (e. g., Lakatos 1977, sec. d4). Unfortunately, neither Popper's nor Lakatos' views offered many insights about how to determine the rationality of those *concrete* decision making processes; instead, both authors always presented this issue as if the problem of when to stop testing a basic statement, when to abandon a research programme, and so on, were a problem left to the judgement of the flesh-and-bone scientist, rather than to the philosopher's elucidation.

It is not strange that sociologically or psychologically minded authors, following Kuhn (1962), took precisely the concrete processes of acceptance of statements or practices as the topic which deserved more attention for students of science. These authors have put forward an impressive number of case studies intended to show that scientific decisions are the outcome of 'negotiations' influenced by interests, prejudices, biases, and so on. Researchers' reference to general rules of scientific method seemed to play only an ideological or rhetorical role, not justifying the epistemic value of those statements on which a scientific consensus is eventually reached (see, for example, Knorr-Cetina 1981, and Latour 1987).

But it should not be assumed that the presence of non-epistemic interests automatically invalidates the internal rationality of the process which leads to the acceptance of those theories. As every student of economics knows, 'greedy' private interests may produce, at least under some institutional settings, efficient public results, and perhaps something similar takes place in the case of scientific research. In this vein, some epistemologists have recently employed economic-theoretical tools to show that some possible institutional mechanisms make epistemic goals attainable, not only 'in spite' of private interests and psychic biases, but even *thanks* to them. For example, under some regimes of merit attribution the probability of finding a true solution to a scientific problem is higher than when researchers directly pursue 'truth' instead of 'merit' (see Goldman and Shaked 1991, Kitcher 1993, and Goldman 1999).¹ One fundamental prob-

1. Some good surveys of the 'economics of science' are Stephan (1996), Hands (1997) and Sent (1999).

lem with these results is, however, that it may not be in the interests of scientists to establish such merit attribution regimes; a ‘philosopher-monarch’ (to use Kitcher’s expression) could force scientists to do so, if he were really interested in promoting truth acquisition, but, in the absence of such an external enforcement, perhaps researchers would prefer to work under a different system of rules, one less efficient in the production of objective knowledge, but more efficient in satisfying their own private goals. Or perhaps they prefer to work without any kind of rules whatsoever!

In this paper, I will study first what systems of norms (and, in particular, what ‘merit attribution regimes’) would be chosen by the members of a scientific discipline, were they given the chance of making this choice by themselves and assuming they were ‘recognition-seekers’; and second, what epistemic properties the resulting norms would be expected to have. In order to approach these questions, I will not use a ‘general equilibrium’ model, but a different economic tool, ‘*constitutional economics*’-the study of the collective choice of norms by rational agents. If scientists did not obey any kind of rule while competing for ‘recognition’, then no researcher would ever have a reason to accept a ‘rival’ theory, since the pursuit of recognition is a zero-sum game (you can only win if others lose) and every researcher would have permanently open the possibility of rejecting all their competitors’ theories on the basis that ‘they are not verified enough’ (that this possibility exists is something we have to concede to falsificationism). So, if scientific research *does* exist, and if researchers *are* recognition-seekers, then the game of science *has* to be played under some rules, in particular, rules commanding the acceptance of certain propositions if certain circumstances obtain. Stated differently, the scientists’ decisions about what statements to accept must be subjected to some recognisable patterns, for *if no such patterns existed at all, no ‘merit-seeking’ person would choose to enter the game of science*, since she would not have any way of knowing what to do in order to get her own theories accepted by her possible colleagues. Therefore, I propose to consider the ‘negotiation of scientific knowledge’ as a negotiation about the norms governing the process of scientific research. The specific aim of this paper is to show that recognition seeking scientists may have a preference for some specific set of norms, and hence, that these preferences will count among the factors determining actual scientific norms.

Only a very limited type of norms will be analysed here: those stating when a theory is so ‘good’ that it becomes compulsory to accept it (these are what I will call ‘inference norms’). In particular, I will *not* study here the norms defining what kinds of properties of scientific theories will count as ‘epistemic virtues’; this problem is an extremely important one, and I have devoted to it some other papers,² but I have opted not to discuss it

2. In Zamora Bonilla (2000) and (forthcoming a) I have tried to show that some prev-

here, both because of space limitations, and because that discussion will probably distract our attention from the analysis of the inference norms, for this analysis will be the same no matter what epistemic virtues are preferred by researchers. In order to perform our analysis, the following assumptions about the members of a scientific community will be made: first, that they have already agreed on what will count as an 'epistemic virtue', although these criteria may be difficult to apply to concrete cases; second, that they do not necessarily worry about the epistemic value of proposed theories, even if they are wise enough to make an estimation of that value; and third, that what they are really pursuing is the merit of having devised a theory their colleagues explicitly recognise as 'correct'. The problem I will attempt to solve in this paper, with the help of a very simple economic model, is *whether the members of such a scientific community could reach an agreement on how to decide who must be taken as the winner of the race for recognition.*

The main conclusion that can be drawn from our model is that we can decompose scientists' motivations into two different elements: epistemic preferences and the pursuit of recognition. The first element can be used to explain what counts for scientists as a cognitive virtue, and hence, to explain the *scale* against which the scientific value of proposed scientific hypotheses is to be measured. On the other hand, the 'pursuit of fame' will normally have *no* influence on the choice of this scale, but, instead, it will determine *which point* on the scale is going to divide 'acceptable' hypotheses from 'unacceptable' ones. In this sense, the operation of the pursuit of fame provides the argument to fill the gap in falsificationism referred to in the first paragraph: the unsolved question of how scientists actually determine *when* to conclude the testing of a scientific statement in order to make a decision about its acceptability. Falsificationists would probably deny that such a decision is necessary at all, but we can take as an empirical fact about scientific research that countless statements (abstract theories, general hypotheses, experimental laws, and so on) are *actually* accepted by scientists, and it seems reasonable to try to explain *why* researchers behave the way that they do.

In the next section I will present the fundamental ideas of this contractualist approach, and justify its relevance for the study of scientific norms. Section 3 introduces a classification of types of scientific norms in terms of their role in the 'persuasion game'; only a brief comment will be provided about each type of rule, since the rest of the paper will focus on

alent norms for the comparative assessment of scientific theories can be explained under the assumption that the main 'epistemic virtue' preferred by scientists is what I have called 'empirical verisimilitude'. These papers are, therefore, a necessary complement of the present one in offering a more complete theory about the constitution of scientific norms.

inference norms. Section 4 will present the paper's basic analytical instrument, which is the 'success function', i.e., a function indicating the prior probability each researcher has of producing a theory which surpasses a battery of tests of predetermined strength. In Section 5, the properties of some possible inference norms are studied. Section 6 shows how an agreement about a norm can be reached even if the members of a scientific discipline do not have identical preferences about the optimum rule. Lastly, Section 7 summarises the main conclusions of the paper.

2. The Constitutional Approach to Scientific Norms. Norms are a controversial topic in economic theory. Most economic models assume that the relevant norms are given and simply study how agents will behave under those norms; sometimes, the application of a new norm or the withdrawal of an old one is suggested, as a way to enhance economic efficiency or social justice, but it is almost always assumed that the State will have the power (and the will) to implement such a change. This assumption leaves a number of questions open; for example, how can we be sure that politicians will actually impose the most efficient norms, instead of those which favour their own interests? Where do the criteria of social justice come from? Why has the State the powers it has? And, in particular, since many relevant norms are 'social conventions', not imposed by the State, the question is simply why do these norms exist at all? Some authors have denied that these questions may receive an answer which is ultimately and solely based upon the assumption of egotistic maximisation of utility on the part of individuals (see Elster 1989). Other authors, instead, have defended the notion that norms can 'emerge' and 'evolve' as an unintended result of the interaction of many individual decisions (e.g., Hayek 1973, Sugden 1989), though their approaches have little in common besides, perhaps, this conception of norms as 'emergent'. Lastly, a third group of authors have argued that rational and autonomous people have the capacity to negotiate and change the norms under which they live, independently of the origin of the pre-existing norms; this is the 'contractarian' or 'constitutional' approach (see especially Brennan and Buchanan 1985, and Vanberg 1994). I will not discuss here the pros and cons of these competing approaches; it is enough to recognise that, in order to explain the emergence of norms, it may be necessary to use some explanatory mechanisms besides instrumental rationality and strategic reasoning; for the 'emergentist' approach, this additional explanatory element would be *history*, while for contractarians it is basically the *autonomy* of individuals, as well as the *moral stance* of recognising this autonomy in other people (see Buchanan 1991). I will try to apply this contractarian perspective to the study of scientific norms, though other approaches may also be illuminating.

The most important aspects of constitutional economics that should be taken into account are the following:

- a) Though every norm may in principle be subjected to constitutional analysis, attention has mainly been paid to ‘fundamental’ norms, in the sense of rules which define basic liberties, or, more frequently, in the sense of norms which regulate the process of choosing other, ‘lower-level’ rules.
- b) The evolution of a game played under definite rules can be more or less predictable, because people, once the rules are established, tend to be pulled more or less automatically by their interests, and because there is usually a limited set of strategies open to each individual. However, the adoption of a concrete system of rules (like the emergence of a certain institution) is, in comparison, much more difficult to predict, because people must devise those rules in the first place. Constitutional economics does not attempt to survey all ‘possible’ systems of norms in order to determine which one is ‘optimal’; it simply tries to analyse specific types of rules (for example, some voting mechanisms), so as to inquire about their properties.
- c) In order to assess social states or economic outcomes, constitutional economics does not employ any normative criterion other than the *free agreement* of individuals. Neither must a definition of social justice or economic welfare be imposed ‘from above’, nor is there any need of ‘aggregating’ individual values or preferences. ‘Right’ outcomes are simply those deriving from the application of ‘right’ norms, and these are just the norms that people have actually chosen, provided that the choice has been unanimous, well informed, and free.
- d) A fundamental distinction is made between the choice *of* norms and choice *under* norms; the latter amounts to ‘playing a game’, which can obviously be extremely competitive; the former, instead, consists in *deciding which game to play*. As long as norms are expected to be in force for a long time, and to apply to many different situations, it would be very risky for individuals to insist on the collective adoption of a rule which just happened to be favourable to them in their present situation, since it may easily cease to be so as soon as circumstances change. In this sense, the choice of norms is made ‘under a veil of ignorance’, and ‘impartial’ rules—in the sense of warranting a satisfactory gain to anyone in a wide set of conceivable circumstances—will tend to be chosen (cf. Rawls 1971). Impartiality is also warranted by the fact that the chosen norms must be unanimously accepted among the members of a

group, at least as long as people have the opportunity of joining another group if they do not like the chosen rules.

The relevance of these aspects of constitutional economics to the analysis of scientific norms is easily understood. In the first place, when scientists ‘negotiate’ the content of a communication, or the acceptance of a hypothesis, or the use of a piece of equipment, they have to take into account not only their own interests, or the interests of the people they are negotiating with, but also the *justifiability* of their decisions before their colleagues. For example, if my experiment has produced a result that contradicts a prediction of your theory, we can ‘negotiate’ how I am going to present that result, but our colleagues would not allow that I simply decide to conceal it—if they happened to discover the terms of our ‘negotiation’. So, even if it were in my interest to diminish the impact of my discovery on your theory (e.g., because of the favours you have promised to me in exchange), I could not do it unless I provided an *argument* indicating why the result is not so damaging, and this argument must show that my stated decisions follow some patterns which are accepted as *general principles of argumentation* among our research community. The norms studied in this paper are just this kind of ‘patterns of justification’, particularly those that can be used in turn to justify the acceptability of ‘lower-level’ patterns.

In the second place, I will not try to present an ‘optimal’ set of scientific rules, i.e., one defining the best possible game for scientists to play. My aim is only to analyse a limited number of norms which happen to be similar to those that scientists seem to employ. So, I plainly admit that other authors may discover that different norms are more consistent with actual scientific practice and more satisfactory for each researcher than the norms studied here; but if this happens, I think such a discovery should count as a success of the contractarian approach.

In the third place, this approach does not attempt to assess the efficiency of scientific research through the comparison of scientists’ actual performance with some absolute epistemic standard engendered in the philosopher’s cabinet, one trying to reflect a God’s eye point of view (to use an expression suggested by an anonymous referee); neither does it try to provide a mathematical procedure for ‘aggregating’ scientists’ preferences, or for ‘weighting’ their cognitive and their private interests in order to find an ‘objective’, ‘supra-individual’ criterion of theory evaluation. As Hands (1997) has pointed out, behind most economic analysis of scientific research there is the ambition of offering an ‘invisible hand’ argument showing that researchers’ ‘selfish’ behaviour tends to produce cognitive results which are sound from an ‘objective’ point of view (by the way, this ambition explains the use of general equilibrium tools). The main problem,

as Hands rightly indicates, is that current techniques of economic analysis do not guarantee that one can coherently use this ‘objective point of view’, understood as an aggregation of individual perspectives. On the other hand, the constitutional approach is radically committed to methodological individualism; the ‘public’ state of scientific knowledge about a certain issue is simply identified with the specification of what proposition (if any) each member of the scientific community has accepted concerning that issue. For the constitutional approach, therefore, unanimity is the only acceptable ‘aggregation mechanism’, though it works just in those cases where no aggregation mechanism is needed; in the remaining cases, this approach does not look for any ‘collective point of view’.

In the fourth place, as long as methodological norms are subjected to negotiation by the members of a scientific community, and as long as the norms have wide applicability and are adopted as unanimously as possible, they will tend to be *neutral*. One reason is that it will be difficult for scientists to accurately predict the effect of each possible norm on their expectations of getting a high level of recognition; hence, researchers will have to use some additional criteria in order to decide what norms to accept, such as their epistemic preferences or their estimations of the probability of getting a solution if certain norms are established. Another reason is that, although the pursuit of recognition can be understood as a ‘zero-sum game’ (with some space open to cooperation, nevertheless), this will not be so in the negotiation about the norms of that game, partly because it is unlikely that there are dramatic differences in scientists’ estimation of the general probabilities of success or in the cognitive interests of the members of a scientific community.³

3. The Persuasion Game. According to many sociologists, scientists’ main motivation is not ‘the pursuit of truth’ but ‘the pursuit of recognition’. I will operationalise this concept as *the number of colleagues who explicitly accept as correct, the models, hypotheses, or theories presented by you*. As a referee of a former version of this paper pointed out, this definition is very close to David Hull’s (1988, 306), who says that “scientists *must* seek first and foremost to have their work accepted by their peers, not by government officials, science reporters, or the general public” (italics mine). Besides the obvious similarity between both definitions, some differences are worth mentioning: firstly, it seems that, for Hull, the pursuit of peer

3. This may be due to the fact that those epistemic preferences are *learned* during the process of becoming a member of the discipline. To honour the evolutionary approach, this entails that *history* (or ‘path-dependency’) might matter a lot in explaining the prevalence of some scientific norms, but I will not pursue this line of thought further at this point.

recognition is not a *basic* desire of scientists, but a kind of behaviour which is somehow ‘imposed’ on them by the scientific community (hence the ‘must’). Secondly, and more importantly, according to Hull “the most important recognition that scientists can receive for their work is its *use* by those scientists working in the same area, including of course their closest competitors” (pp. 309–10; italics mine). Though admitting that this can be a correct description of the way recognition is actually gained, I think that, in order to justify the conclusion that competition among recognition-seeking scientists can be an epistemically efficient mechanism, it would be necessary to *explain* why scientists would prefer to use those statements they think are true, if they also assume that their colleagues are not ‘truth-seekers’. In particular, we can point to three missing explanations in Hull’s account: the first refers to the question with which I started: how to decide *when* a hypothesis is so well confirmed that it is worth using? If no common rule is established about this, then each scientist will be free to assert that any rival theory (or any data contradicting his own theory) is not still ‘confirmed enough’, and he will tend to use only those data which support his own theory. The second problem is that Hull seems to presuppose that it will always be clear when a result supports a theory; my aim is, instead, to show how scientists can agree on certain rules which *define* that relation of support. Lastly, and more importantly, even if a colleague ‘internally’ accepted that your theory is right and usable, he might refrain from expressing publicly his belief, just in order not to increase your recognition level; it is necessary to explain, therefore, how the strategy of ‘giving recognition’ can become a rational decision *in spite* of this possibility.

Other economic explanations of the process of scientific research are also elaborated on the basis of assumptions similar to those of Hull’s. I refer in particular to the theories developed by Goldman and Shaked, and Kitcher. For example, according to the model of Goldman and Shaked (1991), it is possible that recognition-seeking scientists perform relatively well in epistemic terms, if they receive recognition proportionally to the change in their colleagues’ subjective probabilities produced by the revelation of the results of the experiments made by the former. This model takes for granted that those colleagues will have a reason to *believe* the revealed results (and to change their subjective probabilities accordingly), that they will also have a reason to *manifest* their true subjective probabilities, and that the experiments’ outcomes have been *sincerely* revealed. On the other hand, the models in Kitcher (1993 ch. 8) also take it for granted that scientists will be sincere in revealing both the results of their experiments or observations and their opinions about the correctness of each hypothesis; this allows him to use rather uncritically such concepts as ‘wrong result’, ‘right result’, ‘correct finding’, ‘getting solution’, ‘suc-

cess', and so on. Kitcher explicitly recognises the existence of this problem, though he decides to concentrate on a different set of questions.⁴ The aim of my own approach is, on the contrary, to study which norms defining what must be counted as 'honest' behaviour can be established by scientists themselves, if they take into account their own real interests (which may certainly include epistemic criteria, but also include 'recognition'). Stated in other terms, the model presented in this paper tries to offer an idealised and simplified economic explanation of *the process of deciding what constitutes a piece of knowledge whose acceptance is compulsory*; as far as I know, this process remains unexamined in the economics-of-science literature, and I think that it deserves attention quite independently of whether or not one accepts the simplified model I will present here.

The discussion of the preceding paragraphs points towards the necessity of explaining why a scientific discipline has the fundamental methodological norms it has, instead of others. According to the contractarian approach, the 'constitution' of a scientific discipline should be understood as *an exchange of constraints* on the acceptance of scientific statements: every member accepts the submission of his own choices of statements (data, hypotheses, models, laws, theories, and so on) to certain patterns of scrutiny, in exchange for his colleagues' acceptance of the same patterns. What the members of a scientific community have to negotiate is, *which* specific constraints they are going to accept. The rules which seem to be necessary in order to constitute such a game can be classified as follows:

- a) *Norms of inference*: These norms state that if you have accepted certain propositions, and if a different proposition is in some specific relation to them, then you must accept this second proposition. A reasonable assumption is that, if these norms are chosen 'under a veil of ignorance', they will *normally* be consistent with the elementary rules of logic and mathematics, in order to be truth preserving.
- b) *Norms about comparison of statements*: These rules, which I will also call 'norms of theory assessment', are similar to those of type *a*, but, instead of forcing the participants to accept the conclusion that a certain proposition must be accepted, they only force them to accept that one statement is better than another. So, these norms allow the establishment of an ordering of statements by their 'quality' (though this ordering is probably not complete). These ordering rules may be used to defend the preference for a particular

4. Kitcher (1993), p. 304, note 2: "As will become apparent, I shall be concerned more with assessing competence than evaluating honesty, but this should not be taken to imply that issues about sincerity are unimportant."

theory in those cases where norms of type *a* do not uniquely determine the choice of one theory among a set of alternatives. These norms will also be used to define a certain ‘quality level’ such that only those theories which happen to surpass it can be accepted. It is not necessary to determine this level quantitatively (although I will do so in the following sections in order to simplify my arguments), for it can also be determined by means of examples (i.e., if a tentative solution to a problem is as supported by the evidence as this solution to this other problem was, then the former must be accepted). Under the veil of ignorance, norms consistent with the epistemic preferences of scientists will tend to be chosen, for it is unlikely that one will be able to forecast the influence of a norm of this kind on the probability of ‘winning’ in a scientific contest.⁵ Nevertheless, each researcher may have a preference for a rule which favours those theories which have a virtue she is particularly apt to produce (for example, if one is very able in expressing theories axiomatically, she would tend to ‘vote’ for a rule which favours axiomatic theories over informal ones).

- c) *Norms related to actions*: Rules of type *a* only allow you to persuade a colleague to accept your theory if he happens to have previously accepted some pertinent statements; so you would need to persuade him to accept these statements. This can be done by showing that their acceptance follows from other rules of type *a* and the fact that he has accepted still other (more basic) statements, and so forth. But this process can not proceed infinitely, so, if the game of persuasion is to be possible at all, there must exist some different procedure for forcing a scientist to accept some statements. For positivists, this procedure was the ‘direct’ observation of empirical facts, but our approach gives scientists the freedom to decide what kinds of *actions*, performed under what circumstances, are to produce results whose acceptance is compulsory. These norms would regulate, for example, the procedures for performing experiments or observations, including not only their ‘technical’ aspects, but also their ‘institutional’ ones (who must be taken as a qualified experimenter, how the results must be published, and so on). It is reasonable to think that, under the veil of ignorance, scientists would prefer action related norms with the property of encouraging the performance of experiments which yield unbiased results, as well as the sincere revelation of these results.⁶

5. For a discussion of what these cognitive preferences could be, and what norms for theory comparison would derive from them, see the papers quoted in note 2.

6. A brief description of a norm of this kind is offered in Zamora Bonilla (forthcoming b).

- d) *Norms of enforcement*: These establish the penalty that has to be imposed on a scientist who has disobeyed some rule. They 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 to ‘punish’* 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 were required to do so.⁷

Once the rules of a scientific discipline are more or less explicitly defined and accepted, research processes can take place within it. My aim in this paper is not to study the dynamics of those processes, but simply to illuminate some mechanisms which may serve to select the rules under which the game of persuasion is played.⁸ In particular, I will only try to study the negotiation of a norm of type *a*: one indicating when a theory has reached an epistemic value so high that it deserves to be accepted. It should be noted that usually no system of norms is able, by itself, to warrant a universal agreement on every topic within a scientific community. As I said in the last section, norms are used to justify decisions, but this does not entail that, in each case, only one possible decision is justifiable. What is important is that norms preclude the justification of some decisions in some cases.

4. The Success Function. The norms of inference I will study in the next section refer to the choice of a level of epistemic value a hypothesis has to surpass in order to become acceptable as ‘the’ correct solution to a given problem. A preliminary assumption is, hence, that a certain ordering of the possible levels of epistemic value must have been agreed upon by the members of a discipline or subdiscipline, and this will have been done on the basis of the ‘norms for theory comparison’ referred to in the last section. Just for concreteness, we can understand this level as a weighted sum of the number of conceptual and empirical tests that the hypothesis has passed and the number of tests it has failed to pass, but other interpretations can reasonably be defended (my interpretation is obviously close to Laudan’s notion of ‘problem solving capability’, cf. Laudan 1977). However, the difficulty exists that both the *definition* of the epistemic values

7. Norms of enforcement are, therefore, ‘metanorms’ in Axelrod’s sense (see Axelrod (1997), ch. 3). This author shows that introducing this kind of norm in computer ‘prisoner’s dilemma’ tournaments drastically reduces the frequency of ‘defections’. On the other hand, ‘metanorms’ are even more difficult to justify than ‘norms’ from the bare assumption of instrumental rationality (see Elster (1989), ch. 3).

8. For the dynamics of *theory* choice, as opposed to the choice of *rules* for theory choice, see Brock and Durlauf (1999) and Zamora Bonilla (1999).

and the *determination* of the value of a concrete hypothesis can be matters of disagreement: not every researcher will necessarily attach the same strength or relevance to each kind of test, and it can even be an open question whether a given theory has passed a given test or not (this problem demands the application of ‘norms related to action’). If such disagreements persist at a fundamental level within a scientific discipline, it will be difficult for its members to reach a ‘social contract’ on inference norms; nevertheless, in many disciplines it seems that such disputations are more or less easily resolved, which can be taken as an indication that norms of theory comparison and norms related to action work reasonably well in those areas.

Let us assume, then, that an ordering of epistemic levels has already been collectively defined. A priori, each researcher who is trying to solve a certain problem can produce a solution to it of any possible value (after an appropriate amount of testing), but not all the values will be equally likely. Let $F_i(x)$ be the probability that the solution proposed by the i -th researcher has an epistemic value of at most x . So, $1 - F_i(x)$ will be the probability that his theory *surpasses* the level x (see Figure 1). The basic property of this function is that it is increasing in x : the higher is the epistemic value, the more probable it is for a researcher to reach a theory of that value at most, and the less probable it is for him to reach a theory *better* than that level. It is neither essential that $F(0) > 0$, as in the figure (this would mean that there is a non-negative probability of devising a theory of null value), nor that $F(x)$ equals 1 at some point (this would

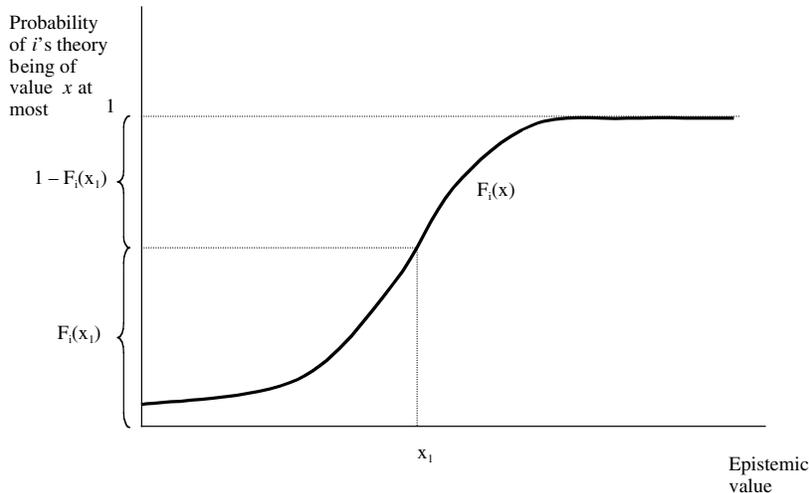


Figure 1.

mean that it is impossible for the scientists to devise a theory better than the quality level associated with that point). We might assume that F also depends on the size of the scientific community, since co-operation and the accumulation of knowledge can make the discovery of better theories more probable, but I will ignore this complication in the rest of the paper. In the next section I will make the additional simplification that F is the same for all community members, an assumption which will be removed in Section 6. Note also that high levels of epistemic value can only be reached after a long process of testing, since they entail that the theory will have surpassed *many* strong tests. So, the choice of a certain quality level will implicitly determine the amount of the time after which there is no need to keep on testing the hypothesis in order to determine its acceptability.

F (or better, $I - F$) is what I will call the ‘success function’ of a researcher. The basic question of this paper refers to the *collective choice* of a level of success such that theories surpassing that threshold may become (or must be) accepted by the members of a research community. If all the members of the scientific community had the same basic preferences, they would simply agree on choosing the threshold which maximises each researcher’s expected utility, and there would be no ‘competition’, since the optimal choice would be the same for all (real competition comes *after* the rule of inference has been chosen and each researcher tries to invent a theory which surpasses the chosen threshold, and also tries to show that this is the case). But if researchers’ preferences are not the same, or if they have different estimations of their probability of success, then it is possible that different scientists may prefer different inference thresholds; in this case, a real ‘negotiation’ to collectively choose a *unique* norm will be needed. I will study this possibility in Section 6.

5. Some Possible Rules of Inference. According to the fundamental assumption of the contractarian approach, the people who are going to be subjected to a norm could devise an indefinite number of alternative norms, and could negotiate which one to institute; it is even possible that the chosen norm does not coincide with anybody’s optimal rule, because the former can be a ‘compromise’ between people’s different preferences. Taking this into account, my goal will simply be to study the properties of some plausible rules of inference (‘plausible’ in the sense that they are intuitively similar to some patterns of theory choice actually used in science) in order to understand why they might have been chosen by recognition seeking researchers. The three alternative norms I am going to examine have a common feature: they establish a minimum level of epistemic value such that if a tentative solution to a scientific problem has surpassed

that level (or 'inference threshold'), then this would justify its acceptance as a *valid* solution. These three competing rules are the following:

(R-I) A level of epistemic value is established such that the theory with the highest value must be accepted by every member of the discipline, provided that it has surpassed that threshold. If no theory passes the threshold, the rule commands the suspension of judgement.

(R-II) A level of epistemic value is established such that the theory with highest epistemic value must be accepted by every member of the discipline, provided that it is the *only one* which has surpassed that threshold. If none or more than one theory has surpassed the threshold, the rule commands the suspension of judgement.

(R-III) A level of epistemic value is established such that the theories which have not surpassed that threshold can not be accepted. If only one theory passes the threshold, it must be accepted. If more than one pass it, then each researcher can either choose one of them, or suspend judgement.

The members of a scientific discipline have to decide in the first place which type of inference norm to establish, and secondly which inference threshold to choose (actually, this threshold may be defined by means of 'paradigmatic examples', instead of through a quantitative measure). The constitutional perspective suggests that these choices are prior to knowing which specific theory or solution is going to be invented by each member of the discipline, or, at least, prior to knowing whether these theories will pass enough tests. These choices can even be prior to knowing which specific problems will have to be solved in the future. Under these circumstances, the expected utility associated with choosing one of these norms plus a definite threshold will depend on three factors: the probability the scientist has of devising an acceptable solution, the probability of there being more than one acceptable solution (these two factors will depend on the success function), and the utility of getting one's theory accepted by all or some colleagues. We can add a fourth factor, which is that stronger thresholds will be preferred to weaker ones, *ceteris paribus*, indicating, at least, that the recognition for having won a strong competition is higher than for winning a weaker one. Nevertheless, I will also present the comparison of the three rules without assuming this fourth factor, because it is interesting to analyse the possibility of a constitutional agreement about a norm of inference even when scientists are motivated *exclusively* by the pursuit of fame and are given 'exogenously' the scale on which their theories are to be judged.

Assuming that every member of the community has the same basic preferences, it can be shown that R-I dominates R-II, since there can be

situations in which one has proposed a theory which is acceptable under R-I but not under R-II (i.e., if one's theory is the best one, but there is at least one more theory surpassing the inference threshold), but not vice versa. R-II is also dominated by R-III, for the same reason (if one has proposed a theory which is not the only one surpassing the threshold, one can get some utility under R-III, but not under R-II, while, if one's theory is the only acceptable one, the utility received is the same under both rules). On the other hand, no comparison is possible a priori between R-I and R-III: which rule provides a higher expected utility will depend on the particular forms of the utility function and the success function. All this entails that, facing a choice between R-I, R-II and R-III, a scientific community would discard the second, and would establish one of the other two. It is possible that those disciplines where a wider agreement on theories exists have opted for a rule close to R-I, while disciplines where disagreement is more frequent are regulated by R-III or some similar norm. Nevertheless, it is also possible that this disagreement about the right solutions to scientific problems is due, not to the character of the rules of inference, but to the inherent difficulty of applying any rule to specific cases.

Table 1 summarises the results on the *optimum inference threshold* for each rule, both assuming that scientists have only a preference for recognition (column *a*), and assuming that they also have a preference for knowledge (column *b*); the calculations are given in the appendix. The variable used to describe the optimum threshold t is $1 - F(t)$, i.e., the probability a researcher has, a priori, of finding an acceptable solution to the problem, if the level which determines acceptability is t . So, a high level of $1 - F(t)$ will correspond to a low threshold, and vice versa. The variable n represents the number of members of the research community who are trying to solve a given problem.

The interpretation of these results is the following: in the first place, for scientists merely worried about recognition, the optimum choice if R-I were established would be the absence of a threshold ($1 - F(t) = 1$): the best proposed theory should be accepted by every member independently of the epistemic value of *that* theory. It does not seem that scientists behave in such a way. For some problems, all the proposed solutions are taken to be too bad to be acceptable, and hence, either real scientists do have a preference for knowledge, or they have chosen some type of inference rule other than R-I. In the first case, it is shown in the table that when R-I is instituted, and scientists prefer a higher inference threshold to a lower one, *ceteris paribus*, the probability of devising an acceptable solution will be less than one, though the exact value of this probability can not be determined a priori; since, as we will see below, it is also impossible determine a priori the threshold associated with R-III assuming epistemic prefer-

TABLE 1. OPTIMUM INFERENCE THRESHOLDS.

RULE	KIND OF PREFERENCES	
	<i>a</i> Only with a preference for recognition	<i>b</i> With a preference for recognition <i>and</i> for knowledge
R-I (the best theory must be chosen, if it passes <i>t</i> ; in other case, suspend judgement)	$1 - F(t) = 1$	$0 < 1 - F(t) < 1$
R-II (the best theory must be chosen, if it is the only passing <i>t</i> ; in other case, suspend judgement)	$1 - F(t) = 1/n$	$0 < 1 - F(t) < 1/n$
R-III (the best theory must be chosen, if it is the only one passing <i>t</i> ; if more theories pass <i>t</i> , choose any of them or suspend judgement; in other case, suspend judgement)	$1/n < 1 - F(t) < 1$	$0 < 1 - F(t) \leq 1$

(Note: n = number of scientists who try to solve a certain problem; $1 - F(t)$ = probability of devising a theory better than threshold t).

ences, so the only possible way to establish empirically that R-III (or a similar rule) has been chosen instead of R-I would be to find some scientists who have decided to accept a hypothesis which they recognise to be inferior (for the time being) to another one, *and their colleagues find that decision to be legitimate* (which indicates that they do not think some rule has been violated).

In the second place, the value $1 - F(t) = 1/n$ corresponds to the case where the expected number of 'acceptable' solutions to each problem is exactly 1, if n researchers are trying to find one solution. The choice of such a threshold would institute a rule of acceptance which has some 'family resemblance' to *eliminative induction*,⁹ since both have as their main function the leaving of only one theory or hypothesis acceptable. However, three important differences exist between the two rules. Firstly, eliminative induction is assumed to 'prove' (at least under the assumption of 'background knowledge') that all non-acceptable theories are false, whereas the rule we are discussing institutes a conventional definition of what it is to be not-acceptable. Secondly, the application of eliminative

9. Kitcher (1993) argues, on the basis of historical cases, that eliminative induction is one of the most widely used methodological rules in scientific research.

induction (especially when it is not fully conclusive) requires us to take into account the probability *each theory* has of being true, whereas the other rule simply requires us to take into account the probability that *researchers* have of finding an ‘acceptable’ theory. Lastly, eliminative induction demands that we keep on testing until only one theory remains, while the other rule establishes a priori a level of epistemic value such that, on the average, there will be one acceptable solution per problem, but it is possible that, ex post facto, some problems will have more than one solution and others have none (R-I and R-II differ in allowing and not allowing, respectively, the acceptance of rival solutions if there are more than one). The hypothesis I would like to advance is that something similar to the rule derivable from R-II or R-III (with epistemic preferences) is what has been tacitly agreed upon in many scientific disciplines, and that philosophers of science have usually misinterpreted this agreement as an indication that another similar rule (eliminative induction) is actually used by scientists.

In the third place, Table 1 shows that, starting from R-II under the assumption that researchers have no epistemic motivations, introducing this type of preference will lead to an increase in the chosen threshold, whereas allowing the acceptance of any theory which has passed the threshold (i.e., accepting R-III instead of R-II) will lead to a reduction in the threshold. The consequence of this is that, if both changes are made, it will not be possible to tell, a priori, whether the chosen threshold will be higher or lower than that for which $1 - F(t) = 1/n$. Nevertheless, since the utility associated with being the only ‘winner’ is probably much higher than the utility associated with being one among several ‘winners’, the displacement of the optimum threshold upwards due to the presence of epistemic preferences will probably be more intense than its displacement downwards due to the possibility of accepting more than one solution (cf. the appendix). As a result, if scientists agree to establish a rule like R-III *and* they have epistemic preferences, then their optimum threshold will very likely be one for which $1 - F(t) \leq 1/n$; i.e., there will be at most one acceptable solution per problem.

6. The Choice of A Rule When Preferences Are Not Identical. In the previous section it was assumed that all the members of a research community had exactly the same basic preferences and the same success function, and hence, the same expected utility associated with each possible choice of norms. The aim of this section is to show that agreement on a rule is possible even if scientists do not agree about which norm would be the optimum one, provided that their preferences are not too different. The possibility of such an agreement should also serve to recall that the contractarian approach to scientific norms does not attempt to prove a priori

that a certain system of rules is optimum in an objective sense, since, according to this approach, scientists possess the freedom to establish the rules they want. No rule resulting from such an agreement should be taken as 'worse' than another in the epistemic sense, save when it is possible to prove that the choice goes against scientists' actual preferences.¹⁰

Suppose, for simplicity, that there are only two researchers working on a problem. One of them, whom I will call 'Strict', has either a stronger preference for knowledge, a higher estimation of his success function, or both, which makes him prefer a higher inference threshold. The second scientist, 'Lenient', prefers instead a lower threshold. Given the expected utility functions of both researchers, Strict would prefer to establish r^S as the minimum acceptable level of success, while the Lenient would prefer r^L . RU_L represents their reservation utility; the utility which would be received if no agreement were reached and each scientist joined another community or found a different job (it is not necessary to suppose that both scientists have the same reservation utility, but this assumption simplifies the figure). More specifically (see Figure 2), Strict would neither accept a threshold lower than a (since it would yield less than his reservation utility), nor one lower than r^L (since r^L is better for both than any

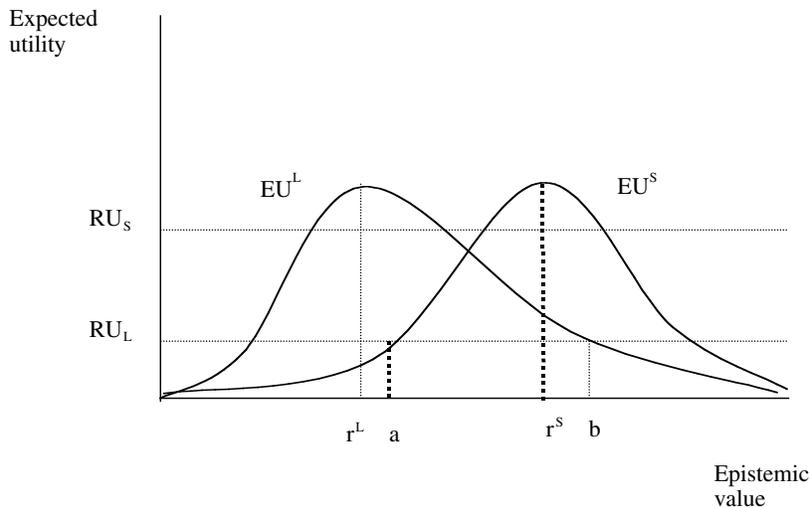


Figure 2.

10. That such inconsistent (or, more precisely, *inefficient*) choices are possible is shown, in a different context, in Zamora Bonilla (1999), where it is proved that it may be the case that all researchers believe that theory T is better than theory T' , but that T' is accepted by more scientists than T .

other point to the left), and Lenient would not accept any point to the right of r^S or b , for analogous reasons. So, any rule between a and r^S would be a Nash equilibrium, and hence, a possible 'point of contract' between Lenient and Strict. If, on the other hand, the reservation utility of these scientists were RU_2 , then no point of contract would exist, since any threshold giving an expected utility to Lenient higher than RU_2 would not be acceptable for Strict, and vice versa (this explains why the preferences of scientists must not be too different if they are to agree on an specific norm).

When the collective choice of a rule and an inference threshold is made by a group of more than two scientists agreement is more difficult, since it is more probable that there is not any level of epistemic value such that the expected utility associated with that level is higher than the reservation utility of every researcher. The problem for the scientific community becomes, therefore, that of choosing a threshold such that 'enough' members accept it, even if it does not coincide with their optima. It can be proved that, if the distribution of optimum thresholds is symmetric around an average point, the level of epistemic value associated with that point will be the threshold which maximises the number of researchers deciding to accept it; so, if the members of a discipline attempted to attract as many colleagues as possible to their research area, they will tend to choose their average optimum threshold. But other, more complicated possibilities are open, for example, it might be the case that some researchers decide to employ the choice of an inference threshold strategically, reasoning that, if they choose a level higher than their own optima, this will deter other researchers from entering the competition; this strategy will be rational as long as a decrease in the number of competitors may increase the probability of being the only one to find a solution. From the epistemic point of view, this 'collusive' behaviour would have the virtue of rising the average epistemic value of the accepted solutions. In any case, these negotiation mechanisms are mentioned here just as open lines of research, the most important point was simply to show that it is possible for the members of a scientific discipline to reach an agreement about a norm of inference, even when their preferences on such a norm are not identical.

7. Conclusion. In this paper I have tried to justify the view according to which the methodological rules of a scientific discipline can be understood as the result of an agreement between the members of that discipline. Seen from this perspective, methodological rules can be divided into those which are about theory *evaluation* and those which are about theory *choice*. It may be expected that, the thicker is the 'veil of ignorance' under which the 'methodological contract' is negotiated, the stronger will be the influence of epistemic preferences on the choice of theory evaluation rules

(these norms, nevertheless, have not been the topic of this paper). On the other hand, rules about theory choice, or inference rules, will mainly be influenced by the interest of researchers in having their own theories accepted by their colleagues/rivals, and hence, these norms will depend more on 'social factors', although the presence of cognitive interests will tend to make the preferred inference rules still more demanding. As a conclusion, we can assert that the 'pursuit of fame' determines *how good* scientific theories can be expected to be (and, with enough talented researchers engaged in competition, they will usually be *quite good!*), though it is the 'pursuit of knowledge' which, basically, determines *in what sense* 'good' scientific theories are *good*.

Another interesting conclusion is that an inference rule more or less similar to 'eliminative induction' may be a likely outcome of the negotiation of a methodological contract. Besides the work of devising and analysing other rules of inference, other lines of research derive naturally from the contractarian approach offered in this paper. To indicate just a few, the constitutional-choice model can be applied to the selection of the other kinds of rules mentioned in Section 3; it would also be interesting to explore whether the 'negotiations' studied by historians, sociologists and ethno-methodologists of science may be reinterpreted as the bargaining about the constitutional constraints which should be observed during subsequent research. This reinterpretation would be a reasonable one if it were regularly observed that methodological choices are seen by scientists as *compromises*, i.e., if a researcher's acceptance of a methodological rule when it helps to support her own theory is taken by her colleagues as a promise to follow the same rule in the future. Such a detailed survey of case studies falls, however, outside the scope of this more analytical paper.

I think that the main virtue of the contractarian approach presented here is that it offers a new point of view from which to consider the old problem of the *normative* character of methodological rules. According to this approach, what makes a scientific norm *a norm* is neither the fact that it is regularly followed by scientists (as many historians and sociologists might have assumed), nor the fact that it follows from logical or empirical arguments about wise strategies for achieving epistemic goals (as it is usually defended by rationalist or naturalist philosophers of science). Rather, what gives scientific norms their normative character is that they are taken as *compromises*, both by the scientists who submit their decisions to them, and by their colleagues, who expect that those compromises are going to be respected. The more freedom researchers have to decide what norms to abide by, and the more disconnected this decision is from short-term rewards, the more impartial this compromise will be, and hence, the stronger it will be from the normative point of view.

Appendix: The Calculation of Optimum Thresholds

For simplicity, I will assume that the utility of not getting one's own theory accepted is 0, that the utility of getting it accepted is $U > 0$ for scientists without a preference for knowledge, and that this is $U(t) > 0$ for scientists with a preference for knowledge if the chosen threshold is t . I will assume that $U(x)$ is positive and increasing, and $U'(x)$ decreasing. The optimum thresholds do not change for utility functions which are a linear transformation of these. In addition, n is the number of scientists trying to solve a problem.

R-I without a preference for knowledge: The expected utility of choosing threshold x is U times the probability of one's theory being the only one passing x ; i.e.:

$$\begin{aligned} EU(x) &= \int_x^\infty f(y)F(y)^{n-1}Udy = (U/n)[F(y)^n]_x^\infty \\ &= (U/n)(F(\infty)^n - F(x)^n) = (U/n)(1 - F(x)^n) \quad (1) \end{aligned}$$

From the last expression it follows that expected utility is maximised when $x = 0$.

R-I with a preference for knowledge: The expected utility is now:

$$\begin{aligned} EU(x) &= \int_x^\infty f(y)F(y)^{n-1}U(x)dy = (U(x)/n)(F(\infty)^n \\ &\quad - F(x)^n) = (U(x)/n)(1 - F(x)^n) \quad (2) \end{aligned}$$

The first order condition for maximisation is

$$U'(x)/U(x) = nF(x)^{n-1}(1 - F(x)^n).$$

For $x = 0$, the right hand side is 0, and increases as x does; our assumptions about $U(x)$ entail that $U'(0)/U(0)$ is positive and decreasing, and hence, the optimum threshold will be greater than 0.

R-II without a preference for knowledge: Now, the expected utility associated with threshold x is:

$$EU(x) = (1 - F(x))F(x)^{n-1}U = (F(x)^{n-1} - F(x)^n)U \quad (3)$$

The first order condition for maximisation is:

$$dEU(x)/dx = U[(n - 1)F(x)^{n-2}f(x) - nF(x)^{n-1}f(x)] = 0, \quad (4)$$

which entails (for $f(x) \neq 0$) that in the optimum threshold, $F(x) = (n - 1)/n$, and hence, that $1 - F(x) = 1/n$.

R-II with a preference for knowledge: Now equations (3) and (4) are transformed into:

$$EU(x) = (F(x)^{n-1} - F(x)^n)U(x) \quad (5)$$

and

$$[(n - 1)F(x)^{n-2}f(x) - nF(x)^{n-1}f(x)]U(x) + (F(x)^{n-1} - F(x)^n)(U'(x)) = 0. \quad (6)$$

Since the last three arguments ($U(x)$, $(F(x)^{n-1} - F(x)^n)$ and $U'(x)$) are positive, the first one $((n - 1)F(x)^{n-2}f(x) - nF(x)^{n-1}f(x))$ must be negative, and this entails that $F(x) > (n - 1)/n$, i.e., $1 - F(x) < 1/n$.

R-III without a preference for knowledge: Let V be the expected utility associated with having one's theory among a group of 'acceptable' theories. Obviously, $V < U$, since you get U when your theory is unanimously accepted. In this case, the expected utility associated to threshold x is:

$$EU(x) = (1 - F(x))F(x)^{n-1}U + (1 - F(x))(1 - F(x)^{n-1})V, \quad (7)$$

and the condition for optimisation is:

$$\begin{aligned} dEU(x)/dx &= Uf(x)[(n - 1)F(x)^{n-2} - nF(x)^{n-1}] \\ &+ Vf(x)[nF(x)^{n-1} - (n - 1)F(x)^{n-2} - 1] = \quad (8) \\ &(n - 1)F(x)^{n-2}f(x)(U - V) - nF(x)^{n-1}f(x)(U - V) \\ &- Vf(x) = 0 \end{aligned}$$

This entails (assuming again that $f(x) > 0$) that $(n - 1)F(x)^{n-2} - nF(x)^{n-1} = V/(U - V) > 0$, and hence, that $1 - F(x) < 0$.

R-III with a preference for knowledge: In this last case, the expected utility associated with threshold x is:

$$EU(x) = (1 - F(x))F(x)^{n-1}U(x) + (1 - F(x))(1 - F(x)^{n-1})V(x), \quad (9)$$

where $U(x)$ is as before, and $V(x)$ is the expected utility associated with having one's theory amongst a group of 'acceptable' theories, when the scientist also has a preference for knowledge. As was said above, it can not be known a priori whether the chosen threshold will be higher or lower than that of R-I without a preference for knowledge; nevertheless, the smaller $V(x)$ is as compared to $U(x)$, and the bigger $U'(x)$ is, the higher the chosen threshold will be, and vice versa.

REFERENCES

- Axelrod, Robert (1997), *The Complexity of Cooperation*. Princeton: Princeton University Press.
- Brennan, Geoffrey and James M. Buchanan (1985), *The Reason of Rules. Constitutional Political Economy*. Cambridge: Cambridge University Press.
- Brock, William A. and Stephen N. Durlauf (1999), "A Formal Model of Theory Choice in Science", *Economic Theory* 14: 113–30.
- Buchanan, James M. (1991), *The Economics and the Ethics of Constitutional Order*. Ann Arbor: University of Michigan Press.
- Elster, Jon (1989), *The Cement of Society: A Study of Social Order*. Cambridge: Cambridge University Press.
- Goldman, Alvin I. (1999), *Knowledge in a Social World*. Oxford: Oxford University Press.

- Goldman, Alvin I. and M. Shaked (1991), "An Economic Model of Scientific Activity and Truth Acquisition", *Philosophical Studies* 63: 31–55.
- Hands, D. Wade (1997), "Caveat Emptor: Economics and contemporary philosophy of science", *Philosophy of Science* 64 (Proceedings): S107–S118.
- Hayek, Friedrich A. (1973), *Law, Legislation and Liberty. Vol. I: Rules and Order*. London: Routledge and Kegan Paul.
- Hull, David (1988), *Science as a Process*. Chicago: University of Chicago Press.
- Kitcher, Philip (1993), *The Advancement of Science*. Oxford: Oxford University Press.
- Knorr-Cetina, Karen (1981), *The Manufacture of Knowledge. An Essay on the Constructivist and Contextual Nature of Science*. New York: Pergamon Press.
- Kuhn, Thomas S. (1962), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Lakatos, Imre (1977), "Falsification and the Methodology of Scientific Research Programmes", in John Worrall and Greg Currie (eds.), *The Methodology of Scientific Research Programmes*. Cambridge: Cambridge University Press, 8–101.
- Latour, Bruno (1987), *Science in Action. How to Follow Scientists and Engineers through Society*. Cambridge MA: Harvard University Press.
- Latour, Bruno and Steve Woolgar (1979), *Laboratory Life. The Social Construction of Scientific Facts*. London: Sage Publications.
- Laudan, Larry (1977), *Progress and Its Problems: Towards a Theory of Scientific Growth*. Berkeley: University of California Press.
- Popper, Karl R. (1959), *The Logic of Scientific Discovery*. London: Hutchinson.
- Rawls, John (1971), *A Theory of Justice*. Cambridge MA: Harvard University Press.
- Sent, Esther-Miriam (1999), "The Economics of Science: Survey and Suggestions", *Journal of Economic Methodology* 6: 95–124.
- Stephan, Paula E. (1996), "The Economics of Science", *Journal of Economic Literature* 34: 1199–1235.
- Sugden, Robert (1989), "Spontaneous Order", *Journal of Economic Perspectives* 3: 85–97.
- Vanberg, Viktor J. (1994), *Rules and Choice in Economics*. London: Routledge.
- Zamora Bonilla, Jesús P. (1999), "The Elementary Economics of Scientific Consensus", *Theoria* 14: 461–88.
- (2000), "Truthlikeness, Rationality and Scientific Method", *Synthese* 122: 321–35.
- forthcoming *a*, "Verisimilitude and the Dynamics of Scientific Research Programmes", *Zeitschrift für allgemeine Wissenschaftstheorie*.
- forthcoming *b*, "Economists: Truth-Seekers or Rent-Seekers?", in Uskali Mäki (ed.), *Fact and Fiction: Foundational Issues on Economics and the Economy*. Cambridge: Cambridge University Press.