Forthcoming in *The Logic Journl of the ILPG*. Special issue on Formal Social Epistemology.

SCIENCE: THE RULES OF THE GAME

Jesús Zamora-Bonilla[1]

ABSTRACT: Popperôs suggestion of taking methodological norms as conventions is examined from the point of view of game theory. The game of research is interpreted as a game of persuasion, in the sense that every scientists tries to advance claims, and that her winning the game consists in her colleagues accepting some of those claims as the conclusions of some arguments. Methodological norms are seen as elements in a contract established amongst researchers, that says what inferential moves are legitimate or compulsory in that game. Norms are classified in three groups: rules of internal inference (from claims to claims), entry norms (from events to claims), and exit norms (from claims to actions). It is argued that the value of a set of norms depends on how efficient they are in leading a scientific community to accept claims ranking high in a consensuated scale of epistemic value, and in giving each member of the community a reasonable expectation of winning some games.

1. VIENNA, 1934: THE ROAD NOT TAKEN.

ñMethodological rules are here regarded as *conventions*. They might be described as the rules of the game of empirical scienceò. This is one of the main statements in Popperôs *Logic of Scientific Discovery* (1959, p. 53). The theory of science developed there is basically a theory of the scientific method: it analyses what procedures in scientific research are *right*. Popperôs basic idea was that this analysis can not be deduced from logic alone -as perhaps positivists and Cartesians might have tried to do-, for we need also to take into account the *goals* of science: the órightô norms of the scientific method are those that are more *efficient* in approaching us to those goals. Hence, according to Sir Karl, science can be conceived as a game defined by certain rules; it *might* have had different rules (this is why they are óconventionsô), but this does

[1] Facultad de Filosofía. UNED. Madrid. E-mail: jpzb@fsof.uned.es. I thank participants in the Tilburgôs workshop on Formal Approaches to Social Epistemology, and in the London School of Economics ñChoice Groupò, for helpful comments.

not entail that the rules are capricious, for what are the ‘proper’ rules has to do with the actual goals of science.

Unfortunately, illuminating as it is, and in spite of Popper’s enormous influence in other respects, this metaphor of science as a game and scientific method as a set of conventions had only but an extremely marginal repercussion.[2] To remedy in part this situation, in a series of papers (Zamora-Bonilla 1999, 2002a, 2006a-b, 2007, Ferreira and Zamora-Bonilla 2006) I have been arguing for the applicability of game-theory reasoning to understand some essential aspects of the construction and evolution of scientific knowledge, aspects that have been strongly disputed within that slippery and misty field between philosophy of science and science studies.[3] The basic idea of those papers was to describe the scientific ‘situations’ from the point of view of scientists themselves, of their epistemic and non-epistemic interests, and to assume that the ‘final state’ to which their decisions lead must have the property of being a Nash equilibrium, i.e., a combination of decisions so that the choice made by each individual is optimal given the choices of the others (the rationale behind this idea is that, if a situation is *not* a Nash equilibrium, then at least one agent will realise she can get a better outcome by changing her choice, hence changing the collective state). This simple requisite proves to be extremely demanding on the set of ‘solutions’ that researchers can reach when engaged in a scientific disputation, and it also allows to the ‘external’ observer (e.g., a philosopher or sociologist) to consider the separate influences that the ‘epistemic’ and the ‘social’ forces behind scientists’ decisions have on the final collective agreement. More importantly, since the participants’ interests are an essential element in the description of the process leading to an equilibrium and of the equilibrium itself, this analysis (contrarily to other social approaches) also allows to show in a transparent way the *value* that the collective situation has for the participants themselves, showing, in some cases, how it might be improved through ‘social engineering’. Lastly, the approach is coherent with the naturalistic notion of epistemology as a branch of science itself, by employing the typical methodology of model building to understand our object of study, and by using those simplified models to make predictions that could, in principle, be empirically tested.

[2] Even philosophers trained in game theory didn’t take the road. The main exceptions are Bicchieri (1988) and Kitcher (1993). Radnitzky, a disciple of Popper, elaborated an ‘economic’ version of falisficationism, but with no connections to game theoretical applications (cf. Radnitzky, 1987).

[3] Any typical introduction to game theory (e.g., Binmore, 1991) is enough to understand the basic elements of a game-theoretic model.

The main difference with respect to the original Popperian metaphor of ôscience as a gameô is that Popper conceived the game in terms of a series of activities that, *in principle*, might be ôplayedô by a single individual (if she had enough mental and physical resources); it is true that mutual criticism plays an important role in Popperôs view of science and rationality, but what is important is that hypotheses *get* criticised, and it is secondary *who* criticises them. The proliferation of hypotheses is also good for science advancement, according to Popper, but, again, in principle one single scientists *might* invent many alternative theories and subject them to sever tests. If there is ôcompetitionô in the Popperian game of science, it is competition between *theories*, not between *scientists*. So, in order to apply game theory to Popperôs insight, it is necessary to consider the role of *human agents* in the game. This is important not only because we must take into account scientistsô goals (and not only the *philosopherôs* view of what the goals of science must be)[4], but also because a fundamental property of any set of rules that must be taken into account is its *implementability*: we might infer from ôurô vision of the goals of science that such and such rules were, *if followed*, mostly efficient in helping us to get those goals, but it can be the case that the real interests of the players make them act, when playing the game defined by those rules, in a way which is different from the favourite outcome, either because real players can decide to *break* the rules, or because they follow them in a Machiavellian way (as, e.g., when two fooball teams decide not to really compete and fix a draw), or, furthermore, because they prefer to establish a different set of rules, more fit to their interests. So, it is necessary to study also what economists call the *incentive compatibility* of the proposed systems of rules.[5]

In this paper I want to concentrate on the question of *what kind of game* (or games) scientific research consists in; in particular, I want to explain in more detail the *types of rules* that constitute those games, for this is a question I have devoted little space in the referred papers, in which simply some general assumptions about the rules of science were given (see, esp., Zamora Bonilla 2006a). Consequently, I shall devote little effort here to describe the scientistsô predicaments in mathematical terms (what I have done in the mentioned papers), but will ascend instead one step in terms of generality to discuss the fundamental game-theoretic structure of the rules of science. So, the first question is what are we talking about when we say ôrulesô or ônormsô

[4] Popper was obviously conscious of this difference, but apparently didnôt consider flesh and bone scientistsô goals as very relevant (cf. Popper, 1972, ch. 5).

[5] This is study by the branch of game theory known as ômechanism designô or ôimplementation theoryô. See, e.g., Mas-Collel, Whinston and Green (1995), ch. 23.

Typically, in game theory the órulesô of a game are simply a description of the *structure* of the game; this structure consists in:

1) a list of players, as well as a list of states of nature (events not depending on the playersô choices) with their probabilities;[6]

2) a list of the actions or decisions each agent can take in each position of the game;[7] the association of an action for each playerôs turn is called the agentôs *strategy*;

3) a *payoff* (or óutilityô) function, associating a distribution of gains or loses for each possible combination of strategies (one for each player) and states of nature.

This abstract definition, however, does not fit perfectly with what are usually referred to as óthe *rules* of a gameô, for the ócanô in the second statement is understood as physical possibility, or, more exactly, to the full set of *real* options the player has (and she knows she has), and not only to the options *allowed* by the órulesô in the ordinary sense. As it is obvious, however, players *can* indeed do what rules say they *mustnôt* do. So, we need to analyse the game in such a way that both behaviour according to the norms, and behaviour contrary to them, become describable within the analysis. A possible way of doing this is by decomposing games in (at least) two different steps or levels, that are really two different (but connected) games: in the first place, we can analyse the decision of the players about ówhat game to playô, i.e., what will be the órules of the gameô in the ordinary sense, including what sanctions will be applied to defeaters; this is what is usually called the óconstitutionalô level, in the sense of, e.g., the constitutional laws of a country. Secondly, we can analyse the expected behaviour of the players once the second game is in place. Though only the second step is what is ordinarily called óa gameô, it is important to notice that both are equally susceptible of being described and studied as ógame-theoretic gamesô. An important consequence of the constitutional level of the game of science being describable in these terms is that the norms of science (or of a particular group or institution within science) must respond to the interests of those players or coalitions ópowerful enoughô to determine what the norms are. This power, however, has usually very strong limits; for example, if less powerful scientists donôt like the rules imposed by the *élite*, the former can often ómigrateô to other scientific fields (or leave science); or, if the working of the imposed norms does not deliver products valuable enough from the point of view of

[6] In the simplest analysis, it can be assumed that all the uncertainty comes from the agentôs decisions, and so only one state of nature is taken into account (which can be obviated).

[7] For simplicity, ópositionsô are described in such a way that only one agent óplaysô each ótimeô, though there can exist simultaneous moves; this is not essential for our current analysis, however.

those providing funds to that branch of science, these may put pressure on the *élite* to change the rules. Taking this into account, the rules of a branch of science can be seen as a kind of ósocial contractô between, first, insider participants, who ónegotiateô *amongst themselves* what will be considered as óappropriateô behaviour in the game of research, and second, between insiders and relevant *outsiders*, those that have some resources scientists need, and can decide to whom will those resources be given depending on what they are getting in exchange.

So, the question I want to ask in this paper is the following: instead of considering scientific rules and norms from the point of view of a detached epistemologist who is trying to design an óidealô science, and instead of taking them just as a brute social fact, we can think of the rules from the point of view of the people that will have to óplayô according to them, and ask whether we would be interested in having exactly those norms or others instead, and, not less importantly, what *concessions* would we be willing to make as regards our óidealô norms in order to reach an *agreement* with other colleagues that would prefer different rules. Of course, in order to ask these questions, we need some information about the interests or preferences of scientists. There have been very strong disputations between órationalistô authors (mainly, philosophers) claiming that the basic goals of science are of cognitive or epistemic nature, i.e., that scientists pursue fundamentally *knowledge* about the world (though different philosophical schools deeply disagree about what knowledge is and how it must be looked for) on the one hand, and, on the other hand, other authors, mainly social scientists, asserting that real scientists pursue lots of other goals, mainly ósocialô goods, as power, prestige, income, or class interests. Though both visions of scientistsô goals are usually presented as deeply incompatible, I think there is no need of doing it so; the incompatibility is due more to the contradictory theses about the rationality and validity *of scientific knowledge* those authors attempt to derive, than to the fact that the two kinds of goals *are* different; for it is obvious that there is absolutely no *logical contradiction* in oneôs having *different* goals, values or interests: this is just what continually forces us to having to make *choices*, and choices are what economic theory is all about. Hence, what we have to do is to acknowledge that scientists have *both* epistemic and non-epistemic interests and values within their óutility functionsô (and that *both* epistemic and non-epistemic preferences can have a *variety* of conflicting elements), and to study how the circumstances in which scientists have to make a decision determine how much of every one of those goals must be honoured or

sacrificed, i.e., which is the *optimum* choice for them in each case. Stated differently: *the economic approach to these choices allows us to interpret the ócconflictô between epistemic and non-epistemic values not in terms of a contradiction, but in terms of a trade-off*, i.e., in terms of how much of some goals is one willing to sacrifice in order to get a little bit more of another goal. So, the right view of the relation between the ósocial interestsô of scientists and the ócepistemic valuesô that philosophers would want be realised in the production of scientific knowledge, is *not* that the former are incompatible with the latter, but, in the worst case, that the former (the ósocialô interests) are the *price* society has to ópayô for having a certain amount of good knowledge, and, in the best case, that the former are an essential part of the social mechanism (in the sense of the ómarket mechanismô[8]) that leads researchers to behave in an epistemically sound way. Hence, the relevant connection between scientistsô epistemic and non-epistemic interest can, hence, be formulated as the following question: by how much must we rise the level of satisfaction of scientistsô non-epistemic goals in order to persuade them to improve by a certain amount the epistemic quality of the knowledge they produce?

We must also take into account that, by their very essence, rules are chosen to be more or less *stable*, i.e., they will be valid during a period that will include a lot of different decisions, and it is difficult for a scientist to forecast exactly how well will those norms affect the acceptability of the results and ideas defended *by her* in the future; so, the defence of a norm must not be based on mere *short term* interests. In order to give more content to this question, we need to specify a little bit more what are the kinds of rules that are pertinent for our discussion. I am not referring simply to the óregularitiesô that can be observed in scientific practice, but only to those regularities that have, from the point of view of scientists, a *normative* content: they are principles (often implicit) that tell scientists what kind of óbehaviourô is *appropriate* and what is not, or that serve to determine the *valuableness* of a scientific output. In the next section I offer a classification of these types of norms, based on the comparison of the process of scientific research with a ólanguage gameô (cf. Zamora Bonilla 2006a), but before that, I shall briefly discuss what can be the content of the scientistsô preferences or interests.

[8] Zamora Bonilla (forthcoming), esp. section 2, offers a review of the proposals to understand science by analogy to the market.

Regarding the ósocialô component, a sensible assumption[9] seems to be that scientists have the same types of interests than the rest of us: they prefer more income, more power, and more comfortable jobs; but they also seem to fiercely pursue fame and honours, what is usually associated to have plenty of the other ógoodsô. The sociologist Pierre Bourdieu famously joined all these things under the concept of óscientific capitalô, one of whose main components would be ócreditô (something mixing the ócredibilityô one has as a researcher ï your results can be acceptedï , and the ótrustworthinessô people put on your capabilities when they assign some resources to you ï óthis will be a good investmentôï).[10] But these ósocialô interests do not exhaust the components of the óutility functionô of scientists: after all, when they *judge* a scientific result, they can see whether it is óbetterô or óworseô than other alternative solutions to the same scientific problem. Some sociologists, particularly Bruno Latour and Steve Woolgar, went further than Bourdieu and tried to dispense of these epistemic criteria, by reducing the scientific judgment to the application of a calculus of credit maximisation (a scientific claim is better if and only if accepting it leads to have more credit).[11] But this reductionist strategy is pointless: if *nothing of epistemic value* were produced by hard scientific research, then most scientists would simply look for less onerous means of getting wealth, fame and power, and more importantly, most people would plainly reject to give a penny from their taxes for scientific research.

The question is what these epistemic values exactly are, and what is their relation to the other goals. Regarding the first question, I think Popperôs answer in *The Logic of Scientific Discovery* (viz., that the rules of science must be designed in order to maximise ófalsifiabilityô) gives only a partial answer: if we observe what scientists say and do, it seems that they *are taking* some results as *real discoveries, not just as* ó*still-unfalsified-hypotheses*ô; so, in practice, they seem to value something like óconfirmationô. Since it is *them* (not Popper) who are going to óplay scienceô, my aim is to discuss what methodological norms would *they* (not Popper) want to establish. My suggestion is that a good strategy to find out what is the best representation of the scientistsô epistemic utility function is to look for those conjectures about this function which are more capable of explaining the most prevalent methodological criteria of

[9] The presence of this type of interests is grounded on the literature on sociology of science, particularly in empirical works (e.g., Latour and Woolgar (1979)). On the other hand, I know of hardly any research about what are the *actual epistemic* interests or values of scientists; in the next paragraph a strategy for research on this topic is suggested.
[10] Bourdieu (1975).
[11] Cf. Latour and Woolgar (1979).

ótheoryô choice;[12] I have defended in some other papers that a modified definition of another not-very-succesful Popperian concept (óverisimilitudeô) would make a nice job here.[13] Actually, Popper introduced the concept of verisimilitude, or ólikeness to the whole truthô, also as an attempt of explaining why the methodological criteria of ópreferring the less falsified theoryô was rational, even if all the available theories were known to be false. His definition of verisimilitude was that *A* is at least as verisimilar as B if and only if all false consequences of *A* are consequences of *B*, and all true consequences of *B* also follow from *A*.[14] If both *A* and *B* have *known* false consequences (i.e., if they have been falsified), but everything that falsifies *A* also falsifies *B*, and everithing that corroborates *B* also corroborates *A*, then the *hypothesis* that *A* is closer to the truth than *B* would have been *corroborated* by *A* and *B* being in exactly that relation with the empirical evidence. However, after David Millerôs and Pavel Tichyôs logical proof that Popper definition was not applicable to false theories, Sir Karl apparently lost interest in the question. My hypothesis is, instead, that the óactual distance to the truthô cannot be an epistemic *utility*, because utilities must be defined on statements that the agent is able of *noticing* whether they are true or false (they must be ópsychologicalô entities, so to say); so, I define instead the *empirical verisimilitude* of a theory as the similarity between the description of the world given by the theory and the description given by the known empirical facts, weighted by the informativeness of these facts.[15] As I shall comment in more detail in the next section, this definition has the virtue of explaining a much wider set of common methodological criteria than the other logical definitions of verisimilitude developed after the failure of Popperôs one.

Lastly, our strategy also allows to understand in a new way the connection between a *descriptive* and a *normative* view of scientific methods; first, there is here a *factual* assumption: scientists take norms as normative constraints on their decisions, decisions that, however, will be based on the pursuit of some personal goals under those constraints; investigating what these constraints are, and what effect they have on the outcomes of science, is a piece of ó*positive*ô research. However, it is also possible to investigate the actual norms of science from a *normative* point of view: are they

[12] Iôm using ótheoryô in the sense of any scientific claim that starts being hypothetical.
[13] See, for example, Zamora Bonilla (1996), (1999), (2000) and (2002b).
[14] Popper (1963), ch. 10. For a survey of research on this concept, see Niiniluoto (1998).
[15] Formally, $Sim(A,B) = p(A\&B)/p(AvB)$, $Inf(A)= 1/p(A)$, and hence, $Vs(H,E) = Sim(H,E)Inf(E) = p(H,E)/p(HvE)$. An alternative, more complex definition is that the verisimilitude of *H* given *E* is the maximum value of *Vs* for *H* amongst all the possible *subsets* of empirical data contained in *E*.

acceptable, or desirable, from the goals that ówe ô (as philosophers, practicing scientists, or citizens) actually have?

2. A TAXONOMY OF SCIENTIFIC NORMS.

In order to describe science as a game, we have to make some choice about what we think the game ós aboutô. As I have said, the pursuit of knowledge has been the main answer offered to this question from philosophy of science, but, if we want to honour the important competitive nature of research, we can say, instead, that the ópointô of the game is to get *the recognition of having made an important discovery*. You not only want to gain knowledge, but you also strive for the world (which may include, for that matter, just a fistful of colleagues) *acknowledging* that what you have discovered is real and valuable. Furthermore, all this recognition can be described from what scientists say, or more frequently, from what they write: basically, *what you want is that others write that what you wrote was right*. This does not mean at all that other things beside talking and writing are unimportant, as we shall see, but communication plays an essential role in the working of the game.

The problem for the órecognition seekingô scientists is that she does not exert a direct control on what *others* say. In principle, it is possible that you have written a marvellous paper, but no one of your colleagues reads it or quotes it approvingly. So, for the game of recognition to take place at all (i.e., for having a chance of being óincentive compatibleô), every scientist must have a reasonable expectation of

a) her colleagues deciding what results to approve according to some *predictable pattern*, and

b) this patter tending to approve results that have, on the average, a sufficient degree of *epistemic quality*.

If you are considering to enter the scientific career and do not expect these two conditions will be met, then you surely will opt for a different way of earning your living.

So, we can imagine scientists writing their papers (call a scientistsô list of papers her ôbookô), in which they present their own arguments, but also use, mention or criticise the claims made in other scientistsô books. Scientific norms can be interpreted, then, as norms telling what can or cannot be written (taking into account what has been

done and written before), and what is appropriate behaviour taking into account what has been written. We can classify scientific rules, hence, into three basic categories:[16]

1) norms of internal inference: these say what claims you must accept (or cannot accept, or are allowed to accept) given the other things you have written on your book before; these norms establish, then, what is a sound *argument*;

2) entry norms: these say what *claims* you must accept (or cannot accept, or are allowed to accept) given what other things have *happened* out of your book;

3) exit norms: these say what *actions* you must carry out (or cannot, or are allowed to carry out), given what is written on your book or on othersébooks.

The main role of the **norms of internal inference** is that of regulating what hypotheses, models, theories, and so on, will be accepted by the scientific community. An old philosophersédream was to reduce this type of norms to the bare rules of formal logic or mathematics, so that all scientific inference could be explained as algorithmic and apodictic inference. But, without denying that science makes abundant use of those types of inferential norms (in any piece of calculation or logical argumentation), it is clear that many of the conclusions that scientists derive are not so well grounded from a formal point of view, nor, when they are more or less uncertain (which is the fate of most of the cases), are they even expressed (or expressible) in the clean probabilistic terms that would allow the applicability of statistical calculus. But the absence of *algorithmic rules* to infer whether a scientific claim is true or false, or how probable it is, does not mean that scientists lack *real criteria* to determine whether the óweight of evidenceéis in favour ófor enoughéof a hypothesis, model, or theory. What happens is that these *real* criteria are tacit, and learned in a paradigmatic way (i.e, transmitted by means of *examples*); they constitute a *practice* to master, rather than a *canon* to be blindly applied (cf. Kitcher, 1993). This does not mean, as well, that these practices are analysable or can not be the object of conscious choice or explicit discussion (as a recent and interesting example, see Vul, 2008). So, I suggest to divide inference rules into two subtypes, that I shall call *microinferential* and *macroinferential* norms.

[16] Distinguishing the three types of rules is just for analytical convenience; the types refer rather to the *functions* that the norms of appropriateness have in the ópersuasion gameô; what we will empirically find out in real scientific communities are collections of normative criteria combining probably the three types of functions.

Microinferential norms are those that regulate those steps in an argument that are susceptible of being criticised as formally invalid; I prefer the term ómicroinferentialô to that of óformalô norms because the main point is not that they are *actually* formal, but that they are *used as if* they could be formalised (in fact, often their mathematical or logical reconstruction, or the formal proof of there validity, has come years or centuries after they started to be commonly used), and because they are applicable *step by step*. Macroinferential norms, instead, apply to those cases where there is *no formally valid argument* from the available premises to the conclusion, basically because the available evidence is heavily *varied*, and not all of it equally *supporting* (nor even equally *consistent* with) the conclusion; these norms, or argumentative practices, refer not to small inferential steps, but to the complex justification process of *theories* or *models*, whereas the microinferential rules are applied to warrant the single propositions that constitute the elements of bigger arguments. Macroinferential norms can be divided into two kinds. In the first place, we have the norms that say what properties count as *virtues* of a scientific claim, and under what conditions can a claim be consider as *better justified* than another. In the second place, we must have some rules telling when is a theory or model so well justified (probably as compared to others), that *its non acceptance becomes forbidden*.

So, the working of the inferential rules proceeds in three levels. In the lower level, some inferential rules allow to *construct chains of arguments*; they are basically the rules of *logic and mathematics*, as well as some general principles of *non-deductive* inference (generalisation, analogy, causality, etc.), not necessarily employed in formal terms by practicing scientists.

In the intermediate level, these chains of argumentation are employed to sustain or criticize the theories, hypotheses, or models advanced by each researcher, i.e., they build each element in the óevidenceô, or each ófactô, in favour or against those hypotheses. The macroinferential norms that govern what constitutes the óvirtuesô of a scientific claim direct this process by determining *what constitutes relevant, positive, or negative óevidenceô*, and, as a conclusion, *what is the epistemic value of a theory, hypothesis or model*. For example, the norm that favours simple theories over more complex ones, or the norm that says that successful predictions are a stronger argument in favour of a theory than the derivation of a previously known empirical result, are norms of this type; but note that what counts as ósimpleô, and the weight of predictions against explanations of known facts, varies a lot from field to field and time to time.

In the third level, once a certain amount of "evidence" has been assembled, other norms must tell whether it is *sufficient* to allow a choice amongst the proposed hypotheses, or to discard some ones, or to force the acceptance of only one of them; or, contrarily, if it is still necessary to collect more evidence before a decision is taken. Stated differently, the rules of the third level determine *when is a theory or model so good that its acceptance becomes compulsory within a scientific community*.

What can the game theoretic approach tell us about the rules that scientists would prefer for the second and third levels?[17] We must note that the definition of "epistemic quality" is something so central in the scientific game, and so relevant to the possibility of transporting results from one field to other field, that it is reasonable to assume that it will be very *stable*. Researchers learn what is what defines the epistemic value of a theory much earlier than they become capable of proposing theories that can be subjected to their colleagues' judgment. So, these norms are "constitutional", in the sense explained above: they must be chosen without taking into account to what cases they are going to be applied, or allowing as little interferences as possible from the desire of favouring specific theories. They must also be very stable in the sense that they are almost undisputed, for, more than other rules, they say what is the game scientist are playing, what is what they are "producing". Taking this into acount, the simplest assumption is that scientists will prefer to estabish those rules for defining the epistemic quality of the theories, that are coherent with the scientists' own "epistemic utility function". Hence, if this is utility can be represented by the concept of *empirical verisimilitude* we saw by the end of section 1, then there is a reason to expect that those norms for theory comparison that derive from the formal properties of that function are the norms that we will observe in actual scientists, which seems to be the case. Some of these norms are:

- between confirmed theories, those with more content are preferable;

- between theories explaining the same data, those with a higher probability are preferable;

- the more implausible a prediction of a theory is, the higher the increment in the value of the theory if the prediction is confirmed;

- being more empirically successul is not a sufficient condition for being preferable, if the theory has very low probability;

[17] I assume that, at the first level, they simply prefer ordinary logical and mathematical rules, plus simple rules of induction.

- if it is expected that new data are going to be found, only the existing data that *confirm* the hypotheses are taken into account to assess their epistemic value.[18]

The three first rules are consistent with Bayesianism; the first and second would be approved by Popper, but not the second one; the fourth is consistent with Kuhnôs description of the judgment of rival paradigms (even radical empirical successes or anomalies donôt force the defender of a paradigm to accept the rival one, if the principles of the other theory look incoherent), whereas the last rule has a Lakatosian flavour (at the beginning of a research programôs development, only confirmations, and not falsitications, are taken into account).

Regarding the norms of the third level (those commanding to accept a theory when it is ógood enoughô), the game theoretic approach leads us to consider the acceptance of a theory as the outcome of a competitive game: each scientists competes for being the ówinnerô, i.e., that having propose de accepted theory (or that having made the *discovery*). What the members of the scientific community have to decide in this case at the óconstitutionalô level is a certain degree of verisimilitude such that the theories that passes that level will be accepted (some especial rules can be established for those cases when more than one passes). So, what these rule define is *what is a discovery*. The relevant question is, hence, if you were a scientist, what definition would *you* prefer? The strategy to answer this question consists in determining the expected utility a researcher would get for every possible definition (i.e., for every possible degree of verisimilitude that were taken as the ódiscovery thresholdô); this demands to know the probability of finding a successful theory surpassing that threshold (i.e, how ódifficultô is to solve a problem in the field), but we can assume that this probability is intuitively known by practicing scientists. The optimal definition of ódiscoveryô would simply be the one that maximises this expected utility. Given certain formal assumptions, it can be proved that the preferred level of epistemic quality would correspond at least to that level such that the probability a researcher has of making the discovery is inversely proportional to the average number of competitors.

It must be stressed that the fact that the norms in the second and third steps are *conventional* and subjected to *choice* by the scientific community, does not necessarily support a *relativistic* interpretation of scientific knowledge. For relativism amounts to the claim that all *alternative* sets of norms would have been *equally valuable*, whereas

[18] See the papers quoted in note 13 for proofs and more examples.

what the "economic" approach advocated in this paper shows is, instead, that between alternative sets of norms scientists will have a *preference*, and then, we can expect that the *existing* norms will be those that *guarantee* in the best possible way the fulfilments of the *actual goals* of scientists (without forgetting all the complexity that the contractual nature of the decision can imply, as was argued in the past section).[19] Or, if they do not, then we can expect movements within the scientific community to *change* the norms. What relativist philosophers should do, instead of showing the *mere possibility* of alternative methodological or epistemic criteria, is simply to think about what criteria *they would like* science to work with, and show us why (or, at least, why they think that all possible criteria are *equally* effective or ineffective in satisficing the goals of scientists, whatever these goals might be).[20]

Before leaving the first group of norms (of internal inference), I want to stress the fact that the game theoretic approach allows us to approach the so called "problem of induction" in a very different way from how it is considered in other philosophical theories. In the typical exposition of the problem by Hume or Popper, it consisted in that no amount of confirmatory evidence can "proof" (in the sense of logical proof) a general hypothesis; another formulation would say that, even assuming that a finite empirical evidence can give a positive degree of confirmation to a general hypothesis, the choice of a particular threshold of epistemic value such that theories surpassing it can be "accepted" is completely arbitrary and has no rational foundation. According to the game theoretic approach, however, the first problem is "solved" by taking as an empirical datum the fact that scientists prefer to play a game in which there are rules that allow to accept a theory on a finite corpus of data, instead of playing a game in which there are no "discoveries" (but only "unfalsified hypotheses"); so, what counts is not that inductive inference can *logically* prove scientific hypotheses, but that there is a set of inferential norms that are accepted by scientists and lead to the compulsory acceptance of some hypotheses. To the second problem, what we can say is that the choice of this threshold is seen as *conventional, but not arbitrary,* for it is the outcome of a rational choice that takes into account the preferences of the scientists; so, the

[19] I understand relativism, so, as the thesis that no rational argument to prefer some rules in certain context can be given; the fact that in different contexts there are different rules does not support, per se, relativism, for there can be reasons why the rules that exist in each context are *more appropriate* to it.

[20] It can also be proved, regarding the norms for theory acceptance, that scientists with an interest *both* in fame and in the epistemic quality of the theories *per se*, will choose a quality threshold higher than scientists with only an interest in fame, and that these will choose a quality threshold *higher than scientists with only an interest in epistemic quality*. Cf. Ferreira and Zamora Bonilla (2006). So, competition tends to *increase* the epistemic standards of science.

ósolutionô to the problem of induction comes from considering the utility scientists derive from playing an inductivist game instead of other possible games. Of course, what one must do in order to epistemically assess the rules allowing to do just that, is to consider what is the average epistemic value of the hypotheses that turn out to be the ówinnersô in such a game.[21]

I pass now to discuss the other two groups of inferential norms. Rules of internal inference allow scientists to pass from some statements written in their óbooksô to others, but there must exist some regulation of the processes by which some propositions enter into the game óby the first timeô. This is the role of the **entry norms**. We can divide these into two kinds: norms telling that one must, or is allowed to, write a sentence in her book because the sentence is written in another scientistôs book; I shall call these *authority norms*. The second kind consists in those rules telling what ónon-linguisticô *events* license or force the introduction of a ódatumô into a scientific argumentation, and I shall call them *evidence gathering norms*. The first group of entry rules refer to óentriesô that are so just from the point of view of a single researcher, of course, but this does not mean that these rules are co-extensive to the ones discussed in the previous paragraphs; an obvious connection between authority norms and internal inference ones is that the former must be coherent with the latter: you can be obliged to accept the claim of another scientist *only* if this claim was also for *her* a commitment entailed by her arguments; but the reverse is not necessary at all: that someone is obliged by the rules of argumentation to conclude *her* argument with some proposition, does not force *other* scientists to accept also the same proposition, if they have not accepted the same premises (as a matter of fact, *most* of the claims made by real scientists in their papers are not accepted by their colleagues). Actually, it is possible to capture the difference between both types of rules by considering that internal inference norms are applicable in the context of the *discussion* between several researchers about what is the right solution to a problem, whereas authority rules apply when a solution has been determined, so that its acceptance becomes compulsory for all the *other* members of the community, i.e., even those not participating in the discussion; authority

[21] Another situation that can be analysed with the help of game theory is when the inferential rules allow each scientists to accept one amongst several incompatible statements (if these have enough quality). It is possible that how interesting it is to accept one of these proposition depends, amongst other things, on how many colleagues accept each one. In this case, we can assume that the community will be in a Nash equilibrium, though it is possible to show that more than one equilibrium can exist. Cf. Zamora Bonilla (2007), 659-666.

norms regulate, then, the *communication* channels going from the ‘original’ discovery to the average researcher, and, very importantly, to students (e.g., textbooks).

More important from an epistemic point of view are *evidence gathering norms*, for they are the ones that connect scientific claims with the real world. These norms are essential also from a game theoretic point of view, for, since the prize for a researcher is recognition, and this is often a competitive prize, in the sense that the recognition given to a scientists lowers the chances of other colleagues being recognized, then it might seem that the dominant strategy for every single researcher would be to *systematically deny* that a colleague has made a discovery; this might they do it not only by asking always for ‘more’ evidence (a strategy macroinferential norms attempt to stop), but by refusing to accept *any* confirmatory *evidence*, i.e., by not accepting the empirical premises necessary for the confirmatory arguments to function. So, scientist need, in their pursuit of recognition, that some clear cases exist where none of those taking part in a scientific discussion can discuss that *certain data are so an so*, or, at least, that the possible reasons to discuss this are clearly specified. This is again a deeply Popperian story: empirical data are not ‘data’ because some intrinsic epistemological reasons (or not *directly* because of that), but because of a *convention*, a convention about when some arguments ‘from world ï or observationï to language’ are *legitimate* moves in the game of persuasion. The first relevant question about these norms is, hence, which ones would you *prefer* if you were a scientist? How many observations of a phenomenon license an inductive generalisation (parallel to the question discussed above, on how many arguments in favour of a theory license its acceptance), what experimental protocols are appropriate (i.e., ‘warrant’ the reading of their results as an objective fact), or what statistical level of significance is to be chosen, all these kinds of questions are subjected to the agreement of each scientific community, and reasons to agree on certain answers instead of others will be based on the same combination of epistemic and professional interests we have mentioned above. For example, there has been a lot of discussion about the limits of replicability in scientific experimentation, and on if this entailed that all empirical data are a mere ‘construction’ (cf., e.g., Collins, 1985); the answer suggested from the point of view of this paper is that real scientists would not demand something as strong as perfect replicability as a necessary basis for accepting an empirical claim: even if it were feasible, it would be as costly as unexciting; they surely prefer to agree on accepting an empirical claim if there are some variety of *different* experimental protocols that independently confirm that claim, what allows to

give recognition to more researchers for more original work. This, however, leads to a greater probability of fraud, since a researcher, expecting that others will not replicate *exactly* her own experimental design, could simply opt for inventing her results. The game theoretic solution to this problem is to institute some rules that make the discovery of fraud more likely and the associated penalty discouraging enough, but the community can tolerate a certain frequency of misbehaviour, if the expected gain in epistemic and professional terms is high enough (cf. Zamora Bonilla (2006), 346-49). These rules can be different for different situations; for example, it is not the same case when one is trying to confirm an existing theory (and forges the data in order to make them agree with the theory), or when one is claiming a revolutionary discovery; incentives to reproduce the results, and hence to certify whether fraud has existed or not, will be different in each case, and the óprizesô given to those in charge of the reproduction, as well as the ópenaltiesô to the óconvictedô, will have to vary accordingly. I would suggest that the growing branch of game theory known as ómechanism designô could be fruitfully applied to this kind of problems, investigating the properties of these type of rules, and establishing something like an óeconomics of trustô in the empirical foundations of research.[22]

Discussion on óprizesô and ópenaltiesô leads us finally to **exit norms**, those that command or license *actions* on the basis of the contents of each researcherôs óbooksô. Since every action entails that certain amount of limited resources are devoted to some goals instead of others, these rules can be described as mechanisms for the *resource allocation*. Everything that is valuable for scientists and can be distributed amongst alternative aims counts as a resource: funds, positions, space for publication, time in meetings, grants, prizes, equipment, assistants, and so on. The role of the exit norms is to state what are the *appropriate criteria* for distributing these resources, including those that determine who can be taken as a member of the discipline.[23] The relevant question, again, is what rules of resource allocation would you prefer as a practicing scientist? For example, would you prefer those rules commanding to engage in self-critical research (*à la* Popper), or wouldnôt be better to let criticism to the órivalsô? Would you prefer ówinner-takes-allô norms, giving a disproportionate amount of

[22] Zamora Bonilla (2006b) applies this idea to the case of the choice of an *interpretation* for oneôs experimental or observational results: in this case, the situation is a game between the author of a scientific paper, who wants to make that interpretation that makes the discovery to be most important, and the readers, who want the conclusion to be as well empirically supported as possible, since this is what warrants its applicability.

[23] I thank to a referee of this paper the suggestion to include the latter type of norm.

resources to the ôbig starsô, or some kind of ôinsurance rulesô, that guarantee a decent chance of success for those not having the good luck of starting their careers in a top department? Would you prefer norms making it very difficult to publish, or more ôliberalô ones? Would you prefer peer review allocation mechanisms, or some other type?

My last observation is the following: along this section I have tended to adopt the perspective of the practicing scientist, on the assumption that the rules of each discipline are negotiated basically amongst its members, but it is also useful to consider the problem from a more general point of view, on the lines of the ôsocial contractô argument outlined in section 1. In the first place, we can also ask what of these norms would a ***ôcommon citizenô*** prefer, where she given the knowledge of the consequences each possible systems of norms will have: it is possible that they are not the same norms as those that are the likely outcome of scientistsô internal negotiation. This possibility suggests very interesting lines for future discussion: what are more ï morally, epistemically...ï ójustifiableô, the norms preferred for scientists, or the ones preferred by citizens? What social mechanisms could be implemented by the citizens in order to make scientists behave in a way more consistent with the formerôs preferred criteria? Or what mutual influence can be exerted by each relevant group, both within science and within society at large, in the construction of a particular normative agreement? Regarding all this questions (which by no means are novel ones), I think that the capacity of game theory to analyse these mechanisms and negotiation processes is an epistemic resource that science studies can simply not dispense with.

REFERENCES

Bicchieri, C., 1988, ñMethodological Rules as Conventionsò, *Philosophy of the Social Sciences*, 18:477-95.

Binmore, K., 1991, *Fun and Games: A Text on Game Theory*. D. C. Heath and Company.

Bourdieu, P., 1975, ñThe Specificity of the Scientific Field and the Social Conditions of the Progress of Reasonò, *Social Science Information*, 14 (6), 19-47.

Collins, H. M., 1985, *Changing Order: Replication and Induction in Scientific Practice*, University of Chicago Press, Chicago.

Ferreira, J. L., and J P. Zamora Bonilla, 2006, óAn Economic Theory of Scientific Rulesô, *Economics and Philosophy*, 22, 191-212.

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

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

Mas-Collel, A., M. Whinston and J. Green, 1995, *Microeconomic Theory*, Oxford, Oxford University Press.

Niiniluoto, I., 1998, ñVerisimilitude: the third periodò, *British Journal for the Philosophy of Science*, 49, 1-29.

Popper, K. R., 1959, *The Logic of Scientific Discovery*. London: Hutchinson.

Popper, K. R., 1963, *Conjectures and Refutations*, London: Routledge.

Popper, K., R. 1972, *Objective Knowledge*, Oxford: Oxford University Press.

Radnitzky, G., 1987, óCost-Benefit Thinking in the Methodology of Research: the óEconomic Approachô Applied to Key Problems of the Philosophy of Scienceô, in G. Radnitzky and P. Bernholz (eds.) *Economic imperialism: the economic approach applied outside the field of economics*, Paragon House, New York.

Vul, E., et al., forthcoming, ñVoodoo Correlations in Social Neuroscienceò, *Perspectives on Psychological Science*.

Zamora Bonilla, J. P., 1996, óVerisimilitude, Structuralism and Scientific Progressô, *Erkenntnis*, 44:25-47.

Zamora Bonilla, J. P., 1999, óThe Elementary Economics of Scientific Consensusô, *Theoria*, 14:461-88.

Zamora Bonilla, J. P., 2000, óTruthlikeness, Rationality and Scientific Methodô, *Synthese*, 122:321-35.

Zamora Bonilla, J. P., 2002a, óScientific Inference and the Pursuit of Fame: A Contractarian Approachô, *Philosophy of Science*, 69, 300-23.

Zamora Bonilla, J. P., 2002b, óVerisimilitude and the Dynamics of Scientific Research Programmesô, *Journal for General Philosophy of Science*, 33, 349-68.

Zamora Bonilla, J. P., 2006a, óScience as a Persuasion Gameô, *Episteme*, 2, 189-201.

Zamora Bonilla, J. P., 2006b, ôRhetoric, Induction, and the Free Speech Dilemmaô, *Philosophy of Science*. 73, 175-93.

Zamora Bonilla, J. P., 2007, ôScience Studies and the Theory of Gamesô, *Perspectives on Science*, 14, 639-71.

Zamora Bonilla, J. P., forthcoming, ôThe Economics of Scientific Knowledgeô, in U. Mäki (ed.), *The Philosophy of Economics*, Amsterdam, Elsevier.