Metadata of the Book that will be visualized in Online

Book Title	EPSA Epistemology and Methodology of Science	
Book SubTitle	Launch of the European Philosophy of Science Association	
Copyright Year	2009	
Copyright Holder	Springer Netherlands	
Corresponding Author	Family Name	**Zamora-Bonilla**
	Particle	
	Given Name	**Jesús**
	Suffix	
	Division	
	Organization	UNED
	Address	Madrid, Spain
	Email	jpzb@fsof.uned.es

Chapter 28
What Games Do Scientists Play? Rationality and Objectivity in a Game-Theoretic Approach to the Social Construction of Scientific Knowledge

Jesús Zamora-Bonilla

28.1 Introduction

In a series of papers (Zamora Bonilla 1999, 2002a, 2006a, b, c; Ferreira and Zamora Bonilla 2006). I have been defending an economic, game-theoretic approach to the understanding of the social construction of scientific knowledge; such an approach would complement the traditional efforts in using insights and techniques from other social sciences (esp. sociology and anthropology) to the idiosyncratic epistemic aspects of science, but would also have two fundamental virtues from a 'rationalist' point of view: in the first place, the game-theoretic, rational-choice approach allows to model in an *explicit* way the factors determining the scientists' decisions, as well as the interdependences between them, without dressing all this in a mystifying rhetoric which tends to obscure the analysis more than to illuminate it (many will say that the economic jargon can be no less mystifying and obscurantist than the Foucauldian one so often employed in post-modern studies of science, but the difficulty of rational-choice analysis is of the same kind as that of mathematics and logic: it serves, when used properly, the goal of making the inferential links of our reasoning explicit and subject to criticism); secondly, instead of launching a non-contestable accusation of lack of objectivity and rationality to the products and methods of scientific research, economic models allow to clearly see what are the *specific* shortcomings of certain ways research can be carried out (i.e., they permit us to identify specific *inefficiencies*), and point towards those changes in the modelled situations that would effectively improve the results that scientists are getting. Stated in other words, a game theoretic analysis of the social construction of scientific knowledge allows us not to renounce to the thesis that science is a pretty good method of finding out objective and significant truths about the world, nor to the claim that science is the product of typically human and social forces, nor to the goal of discovering the possible shortcomings of science and, more importantly, of discovering also some ways of overcoming them. I think the focus on the necessary

J. Zamora-Bonilla (✉)
UNED, Madrid
e-mail: jpzb@fsof.uned.es

M. Suárez et al. (eds.), *EPSA Epistemology and Methodology of Science: Launch of the European Philosophy of Science Association*, DOI 10.1007/978-90-481-3263-8_28,

formalisms I had to use in the papers quoted at the beginning may have precluded these virtues being well appreciated enough, so I would like to take this opportunity to state them in a clearer way.

28.2 The Elements of a Game-Theoretic Model of Science

Understanding the construction of scientific knowledge with the help of game theory demands to adopt a cognitive style which is not frequent in science studies (though it is employed more often in some areas of philosophy of science, in particular in the more formal ones, like, e.g., Bayesianism). I am referring to the design of *abstract models*. A formal model is an encapsulated argument, with which we just try to prove that certain interesting conclusions follow (or do not follow) from certain reasonable premises. This argumentative style strongly contrasts with the more usual practice of case studies, one that has flooded the literature in philosophy and sociology of science for the last three decades. I have of course nothing to oppose to that practice, and I recognise that a detailed view of how science is and has actually been done is a precondition for a proper understanding of scientific knowledge, but the case-study technology is essentially inductive, not too appropriate to illuminate the *regular mechanisms* that make science to be the way it is, and the claims we reach with it can hardly be generalised, as the diversity of conflicting views that have actually been defended by means of case studies shows. However, I do not propose to consider the game-theoretic approach to the social construction of scientific knowledge as an alternative to case studies, but as a complement of it; the idea is merely to take the case based literature as a corpus of *descriptive*, *empirical* facts about science, using rational choice considerations as a tool for discovering *theoretical* mechanisms (i.e., abstract models) that help us *explain* (partially, at least) why science is the way those empirical descriptions say it is.

The basic principles that such an approach offers for our attempt to discover this type of explanations are the following. In the first place, it points to scientists' *actions* as the basic elements of what has to be explained, i.e., why do scientists *do* what they do, and in the way they do it; this does not mean that other aspects of scientific constructs (e.g., the structure of theories, the connections between models and observations) fall out of the scope of a game-theoretic explanations, for we can ask, say, why do scientists *choose* theories with a certain structure, or *value* models according to some connections with empirical facts. Secondly, a rational choice approach forces to explicitly consider the *goals* scientists are pursuing through their actions, as well as the *information*, *capabilities* and *social mechanisms* that allow them to reach those goals to the extent they do it. Lastly, what a game theoretic approach more characteristically adds to this is the idea that, as long as the actions of *other* colleagues (or other relevant actors) influence the possible gains a scientist can expect from her own decisions, the social situations that we will expect to observe will be what is traditionally known as a *Nash equilibrium*, i.e., a situation in which the strategy each agent is choosing is a best response to the decisions the other agents are making.

Just to illustrate the way in which a game theoretic approach can be applied to traditional problems in philosophy of science, here are some examples of the kind of questions we can pose:

Scientific standards: many philosophers have longly discussed the virtues a scientific claim must have in order to become acceptable (e.g., 'degree of confirmation', 'corroboration', 'verisimilitude', 'simplicity', 'predictive capacity', and so on), but, since these properties come in degrees, we can reasonably ask how do scientists actually determine the *minimum* level of those virtues a claim must have so that its acceptance by the community becomes 'compulsory'; or, stated otherwise, when will a 'competing' scientists be 'forced' to recognise that the claim proposed by a rival is 'right'. A game theoretic analysis will show that the fact that every possible standard (or 'quality level') determines the chances a researcher has of 'winning' the race for a discovery, is a sufficient reason for scientists to have a preference for some particular standard over the other alternatives (cf. Zamora Bonilla 2002a).

'Theory' choice: Not all scientific claims within a field are 'compulsory' (in the sense that, if you do not accept them, you will not be given your degree in chemistry, say; there are also claims it is compulsory *not* to accept them, in this sense), but many are a matter of choice to some level. The famous 'underdetermination of theories' thesis is just a formal justification of that practice, though its name is misleading in that it does not only apply to real theories, but to any type of scientific claim, from experimental reports to megaparadigms. The problem is that saying that logic is not sufficient to determine the choice of a theory does not help us to know what do these choices *depend* actually on: are these factors 'social interests', 'cognitive biases', or just a matter of 'mob psychology'? The game theoretic approach allows to say that, as long as the profitability for a researcher of accepting a claim depends on who else in her community is also accepting it, then the only stable situations will be those that constitute an equilibrium, in the sense that everybody is making her best choice given the ones made by the rest. In general, there will only be a few equilibria in each case, and sudden changes from a situation to other are possible ('scientific revolutions'?), but it is also possible to proof that, as long as the profitability of accepting a claim depends even very partially on how good the claim is according to the epistemic standards, then it can be expected that 'better' claims will have a tendency to become more 'popular' as their epistemic quality improves (cf. Zamora Bonilla 1999, 2006a, sec. 3).

The 'construction' of an empirical fact: It is almost a platitude from science studies that the way in which empirical 'discoveries' are presented is the result of a 'negotiation'. This entails, at least, that an empirical finding can be presented in more than *one* way. But this does not entail in any sense that all those ways are equally good for every agent engaged in the 'negotiation'; rather on the contrary, if they were equally good, there will be no negotiation at all (which is usually costly), for the mere flip of a coin will be enough to select an interpretation or other. Game theoretical modelling of the situation allows to see how the interests and preferences of the different 'negotiators' induce a distinction between those interpretations that *cannot* be the outcome of the negotiation, and those that can; it also shows some ways in which the outcome can be judged as efficient or not, and, furthermore, it allows to devise institutions that may improve those outcomes (cf. Zamora Bonilla 2006c).

28.3 The Epistemic Quality of Scientific Products

What makes a scientific theory, or model, or hypothesis, a *good* one from the scientific point of view? The economic approach to the social construction of scientific knowledge has just two insights to offer as a way to the answer of this question; they are not very deep insights (as a matter of fact, they are nearly trivial), but they are also controversial when examined from several philosophical or sociological approaches to the matter, and furthermore, there is even a certain amount of tension between both ideas. The first insight consists in the claim that, since in the game theoretic analysis we are assuming that scientists do *rationally* pursue their goals (whatever these happen to be), we are then forced to assume that they will also have a non negligible capacity for understanding what 'sound reasoning' is. In other words: if they are wise enough to know how to navigate the ocean of their social relationships in pursuit of resources, publications, and honours, they must not be necessarily inept when trying to discover the laws governing a physical phenomenon. The second, more important insight, is that the *definition* of the epistemic quality of scientific products is basically *not* a question for the philosopher of science (nor, for that case, for the sociologist or the economist of science), but for *scientists* themselves, or, dare we to say, for citizens in general; i.e. *de epistemicibus gustibus non est disputandum*. Real scientists will have some 'epistemic utility function', probably not everyone the same function, and the first role of the analyst of science is, rather than that of proposing what this utility function *should* be, that of *discovering* which one it is. The point is not to deny that many interesting things can be analysed from a philosophical point of view about the epistemic virtues of scientific items: what our approach precludes is just adopting a *patronising* attitude towards these questions. After all, scientists are the society's *best experts* in the production and use of knowledge, and so, if some people know what knowledge consists in and how to discriminate 'good' pieces of knowledge from not so good ones, these are scientists. Perhaps they are not particularly good in making this practical knowledge explicit (actually, when scientists say what good science consists in, they tend to do it worse than philosophers). We should concentrate, hence, on the way that scientists' *behaviour* 'reveals' what their actual criteria of 'good scientific practice' are. But, how to do that? As a matter of fact, all the theories that have been proposed by philosophers in order to explain the nature of scientific knowledge, its virtues and its progress, and so that some methodological rules of 'good science' can be derived from those theories, have been 'refuted' by showing that, in real scientific

BookID 188902_ChapID 028_Proof# 1 - 11/09/09

practice, researchers often do not behave as if they were pursuing those epistemic goals and the methodologies derivable from them; for example, scientists are usually not 'falsificationist', nor 'confirmationist', but they also are not 'anarchist', nor strict followers of the Lakatosian methodology of research programmes. A big part of the debates of the last 50 years between the different schools in philosophy of science has consisted in showing through historical or contemporary examples that scientists *don't do* what other rival philosophical theories assume they should be doing. My suggestion is that, in these debates, almost all parts were *right* when they were criticising the theses of the rivals (more or less like in politics), but partially wrong when they were proposing a philosophical explanation of scientific practice. So, the outcome of the debate must be seen as a rich corpus of *empirical evidence* about how science is practiced, in which we should try to find out some *regularities* about what scientists actually consider as *good practices*.

The fact that scientists behave in many different, often conflicting ways, is not an argument against this goal: in every society there can be conflicting practices and conflicting norms, both because people have different interests, values, and preferences, and because they face different situations, incentives, and constraints. In the case of science, it is not necessary to discover many norms that *all* scientists in *all* times and places have considered appropriate; it would also be interesting to show that, when certain *specific* conditions are given, then *such and such* norms of 'good practice' tend to be accepted. Our main goal as students of science should be, once the historical evidence is organised in such a way, that of trying to answer the following question: *what hypothetical utility function explains in the best possible way the acceptance of precisely these ideas of 'good practice' by part of real scientits?* Every economic model starts by making some reasonable assumptions about the agents' preferences, and game-theoretic models of science are not unlike the rest. My hypothesis (obviously a very simplified one, for it is applied to very simple models, and also not too original) is that a typical scientist's utility function has two main components: a 'social' one, and an 'epistemic' one. The social component can contain many different variables (income, control over resources, class interests, political or human values, and so on), but the most important one is 'recognition': scientists strive for being recognised by their colleagues as good, or even 'excellent' practitioners of their disciplines; this creates an incentive to agree, within a scientific community, on how a 'good practice' is defined, i.e., to agree on the 'rules of the game', for, if such an agreement is lacking, 'recognition' becomes simply impossible. The question (for scientists, nor for philosophers) is: what criteria to use in order to determine those rules? I think that the more general, more basic rules, i.e., those that allow to say that it is really a type of 'science' the game a community of researchers are playing, and not literature, or music, or football, or car mechanics, or politics, must be rather similar in all scientific communities (though they may have very strong differences in the details), and must not suffer very significant changes with the passing of time; so, by accepting these rules, scientists can not take into account their own *social* goals (or at least, their *private* social goals; things like 'social justice' will be different, and perhaps also 'class interests', but I doubt it), for it is impossible to know how the adoption of some methodological rules instead of others will affect their individual *chances* of getting recognition and things like that. Stated differently, the more basic a scientific norm is, the more it will be chosen by scientists 'behind a veil of ignorance', to use this well known Rawlsian metaphor. Here it is, then, where the *epistemic* elements of the scientists' utility function enter, in determining the choice of what an appropriate criterion of 'good science' is.

A hypothetical epistemic utility function, that tries to explain the prevalence of certain very basic and general criteria of preference over 'theories' (criteria that, as we have seen above, are in apparent mutual conflict), is the one I offered in my own research on verisimilitude, where this notion was explained not as '*objective* closeness to the full truth', but as '*perceived* closeness to what we *empirically know* about the truth, weighted by the *perceived* amount of information this empirical knowledge contains'. In formulae: $Vs(H,E) = [p(H\&E)/p(H \vee E)][1/p(E)] = p(H,E)/p(H \vee E)$, where H is a model or theory, E is the relevant empirical evidence, and p is a Bayesian, subjective probability function, which allows for different scientists attaching different degrees of verisimilitude to the same theories; this does not condemn scientists to relativism, for the shared *form* of their epistemic utility function guarantees that they will agree on certain conditions under which a theory is necessarily better than another (e.g., if H entails H', and both entail E, then H' will be judged to be better than H), and these conditions will give us the 'criteria' of epistemic preference we were looking for. A more sophisticated measure assumes that E is structured as a set of known empirical regularities, and verisimilitude is defined then as the maximum perceived closeness to *one subset* of those regularities. (Cf. Zamora Bonilla 1996, 2000 and 2002b).

28.4 Epistemic Efficiency and Scientific Institutions

One thing is to *define* what the quality of a scientific item consists in, and a very different thing is to *determine* how good that item is according to that definition of quality. The point made in the previous section amounts to saying that scientists will know better than the rest of us the answers to those questions (though they will not *tell* us those answers: we shall have to get them by studying scientists' behaviour), but, even if they agree on what a good theory, model, hypothesis, experiment, etc., is, this does not guarantee by itself that the actual *outputs* of society's investment in scientific research are of a 'high epistemic quality'. This depends on a number of variables, the effort and talent of individual scientists not being the least, but an essential factor is also the *efficiency of the scientific institutions*. I will not refer here to their *economic* efficiency, though I admit this is a topic of fundamental importance, but will restrict myself to discuss the *epistemic* efficiency: do scientific institutions work in such a way that 'high quality' outputs tend to be produced? In order to illustrate how the game-theoretic approach can offer some answers to this question, I will briefly examine in turn the three examples given at the end of Section 28.2.

Scientific standards: The process of collectively choosing a model, hypothesis, etc., as the *right* solution of a scientific problem demands, as we have seen, that the relevant community has agreed on certain standards specifying the minimum level of epistemic quality the solution must have. Will scientists choose a 'low' standard, or a 'high' one? Of course, this question only makes sense in comparison with some independent criterion of 'lowness' or 'highness', and the most helpful one is just that of *individual* scientists' epistemic preferences: imagine that an 'individual' researcher (in the sense that she is not competing with other colleagues for the solution of her problem, but wants just to find out the solution for its own cognitive sake) has to decide when will she be satisfied with the number and variety of tests a solution has passed in order to be accepted by her; so, her choice in this 'isolated' situation will give us a certain epistemic benchmark (I would suggest to science students, and particularly to those more sympathetic to relativism, to ask themselves what standard would *they* choose in a similar situation . . . perhaps it is not very different from the scientists' standard, but, even if it is, it would be nice to have some arguments about *why* both standards are not the same). The interesting thing about the models presented in Zamora Bonilla (2002a) and Ferreira and Zamora Bonilla (2006) is that they show that the *collective* choice in the case of a *competitive* search for the solution generates a standard of quality that is *higher* than the one that most individual scientists would have chosen if they only cared about epistemic considerations, not about recognition! That is, the search of solutions in a competitive environment ends in having solutions *epistemically better* (according to scientists' own criteria of epistemic goodness) that those that would have been chosen in the absence of a competitive pressure. By the way, perhaps they are 'too' good, in the sense that *we, citizens*, would be content with a little bit worse solutions, if this allowed to have solutions to *more* problems. But the point of my argument is just showing that, from an *epistemic* point of view, there is no point in saying that the pursuit of recognition leads to scientific claims that are not 'good enough' on the average.

Theory choice: As we saw in Section 28.3, if the acceptance of a scientific claim by an individual scientists depends on which ones of her colleagues also accept the same claim, it can be the case that more than one 'social situation' (i.e., a description of who accepts and who rejects the claim) is possible. For example, it can be the case that the hypothesis H being accepted by a 20% of the community is an equilibrium (i.e., everybody is happy with her choice, given the choices of the colleagues) and that a 70% is also an equilibrium; which one of both equilibria is the actual one will depend on historical causes. This seems to constitute, by itself, a reason in favour of certain degree of relativism: the scientific consensus is what it is, but *with the same amount of information and the same social relations* it could have been a different one (it is, however, a *limited* relativism, for there are *more* states that are *not* an equilibrium, than states that are possible equilibria). But the situation is still worse: for imagine that H' is a hypothesis that *all* the members of the community agree that is better than H. It can be proof that, in a case like this, for every equilibrium of the worse hypothesis there will be an equilibrium of the better one which is *more inclusive* than the former; so, it can be the case that the equilibria for H' are 30%

and 80% respectively. So, H might be accepted by a 70% of the community, and H' by a 30%, in spite of everybody agreeing that H' is better. Contrarily to the analysis of the past example, this is a case where the game theoretic analysis can show that the interaction between researchers can lead to an epistemic *inefficiency*. The good news are that it can also be proved that, the better becomes a hypothesis (e.g., for having been confirmed by new data), it does not only happen that its equilibria go upwards, but smaller of them are *disappearing*, till in the end only one equilibrium, close to unanimity, remains. So, the growth of empirical knowledge can solve by itself those epistemic inefficiencies.

The 'construction' of an empirical fact: Suppose you have performed an experiment, and are planning to report its result in a paper. It is now a platitude within science studies that you will have at least some possible choices, for 'facts do not speak for themselves', but need to be 'interpreted'. For example, you can report the result as a 'very important', unexpected finding, which forces your community to look for a novel explanation, or you can present it as something uncontroversial. The problem you face is that, the more 'radically' you interpret your result, the less 'credible' it will be, i.e., the less 'well confirmed' by the experimental data you present will be from the point of view of your colleagues. As we saw above, the mere existence of many alternatives does not show that 'all of them are equal': it can certainly be the case that some of them are 'better' than others in specific senses; for example, some can be better from the *epistemic* point of view, whereas others (or, if scientists are fortunate enough, the same ones) are better from the point of view of the social elements of the researchers' preferences; it can also be the case that some options are clearly better for some researchers, whereas other claims are better from the point of view of other colleagues. This *plurality* of valuations has not to be confused with some kind of 'fundamental indifference': game-theoretic models analyse precisely those situations in which people have different interests, and determine what choice they will make in those cases; and, by comparing the outcome determined by these choices with the value that other outcomes would have had for the agents themselves, the models allow to evaluate the *efficiency* of the interaction. It is also possible *for the analyst* to select the values or preferences *she* would like to be enhanced in that interaction (think of her as a science policy maker, for example, or just as a mere epistemologist), and then to think about ways of changing the way the agents interact, so that the chance of getting a 'better' result is bigger. In the case I'm discussing now, what can be proved is that, as long as the 'readers' of the paper the experimentalist is writing value her claim according to a positive function both of the result's 'novelty' and of its 'credibility', but the author of the paper simply wants that the most possible novel claim is accepted, and as long as there is a negative correlation between novelty and credibility, then authors will have an incentive to interpret the results of their experiments in the least possible credible way that is *compatible* with the results' acceptance. This means that, if authors were given a full liberty to describe their experiment in the way they preferred (do not forget we are assuming that they are choosing only between descriptions which are *legitimate* according to the methodological rules of the community), all the 'gains' to the scientific community from the interaction between authors and readers would go to the authors, and these gains will be in the form of 'social', not 'epistemic' values. It is reasonable to expect that communities will have designed some institutional ways of framing that interaction in such a way that epistemic gains are higher, e.g., by standardising the forms of interpreting experimental results, or by making peer review processes more severe.

28.5 Conclusion

Is science 'rational'? Is scientific knowledge 'objective'? The game theoretic approach suggests that these questions can be better rephrased in the following way: Are scientific methods and scientific institutions efficiently 'designed' (or have they appropriately evolved) to provide us the best possible knowledge of the world, according to scientists' epistemic values? And, do *you* find the epistemic goals of scientists appropriate? If the answer to the last question is 'yes', then a positive answer to the first question must be enough for satisfying (to a reasonable extent) our doubts about the rationality and objectivity of science; if the question to the first answer where negative, then we might employ the game theoretic approach to find where is that scientific institutions are failing, and how they can be improved. On the other hand, if your answer to the second question is not, i.e., if you think that scientist should pursue a *different* set of epistemic goals, then what you *owe* us is a specification of what these other goals should be, and game theoretical models of science can help you to discover how science should be organised in order to promote the epistemic values that you prefer. My own judgment is that, if I am right in my hypothesis that something similar to the verisimilitude function I summarised in Section 28.3 rightly describes the epistemic values of real scientists, then most branches of science are leading us to have progressively more and more theories and models that describe, predict and understand better and better an increasing number of empirical facts about the world, and this is the maximum I personally can ask from the epistemic point of view.

References

Ferreira JL, Zamora Bonilla JP (2006) An economic theory of scientific rules. Econ Philos 22: 191–212

Zamora Bonilla JP (1996) Verisimilitude, structuralism and scientific progress. Erkenntnis 44: 25–47

Zamora Bonilla JP (1999) The elementary economics of scientific consensus. Theoria 14:461–488

Zamora Bonilla JP (2000) Truthlikeness, rationality and scientific method. Synthese 122:321–335

Zamora Bonilla JP (2002a) Scientific inference and the pursuit of fame: a contractarian approach. Philos Sci 69:300–323

Zamora Bonilla JP (2002b) Verisimilitude and the dynamics of scientific research programmes. J Gen Philos Sci 33:349–368

Zamora Bonilla JP (2006a) Science studies and the theory of games. Perspect Sci 14:639–671

Zamora Bonilla JP (2006b) Science as a persuasion game. Episteme 2:189–201

Zamora Bonilla JP (2006c) Rhetoric, induction, and the free speech dilemma. Philos Sci 73: 175–193