The referee’s dilemma. The ethics of scientific communities and game theory

Bracanović, Tomislav

Source / Izvornik: **Prolegomena : Časopis za filozofiju, 2002, 1, 55 - 74**

Journal article, Published version
Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:261:985931

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-04-25**

Repository / Repozitorij:

Repository of the Institute of Philosophy

DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

The referee's dilemma. The ethics of scientific communities and game theory

TOMISLAV BRACANOVIĆ

University of Zagreb – Studia Croatica,
Ulica grada Vukovara 68, HR-10000 Zagreb
tomislav.bracanovic@hrstud.hr

Professional article
Received: 14–02–02 Accepted: 06–05–02

ABSTRACT: This article argues that various deviations from the basic principles of the scientific ethos – primarily the appearance of pseudoscience in scientific communities – can be formulated and explained using specific models of game theory, such as the prisoner's dilemma and the iterated prisoner's dilemma. The article indirectly tackles the deontology of scientific work as well, in which it is assumed that there is no room for moral skepticism, let alone moral anti-realism, in the ethics of scientific communities. Namely, on the basis of the generally accepted dictum of scientific endeavor as the pursuit of knowledge exclusively for knowledge's sake, scientifically »right« behavior is seen to be clearly defined and distinguishable from scientifically »wrong« behavior. After elucidating the basic principles of game theory, the article illustrates – by using imaginary and real cases, as well as some views from the philosophy of biology (the units of selection debate) – how this sort of reasoning could be applied in an analysis of the functioning of science.

KEY WORDS: Game theory, prisoner's dilemma, iterated prisoner's dilemma, scientific communities, pseudoscience, biological and scientific selection, units of selection.

Game theory: a few principles

According to Simon Blackburn's *Dictionary of Philosophy*, game theory is »the mathematical theory of situations in which two or more players have a choice of decisions (strategies); where the outcome depends on all the strategies; and where each player has a set of preferences defined over the outcomes« (Blackburn, 1996, p. 153). The word »game« featuring in »game

theory« symbolizes various interactions between individuals, where each individual tries to act in an economically rational manner, i.e. to maximize his/her personal benefit and minimize his/her personal cost. The key to the game and its curiosity rests in the fact that maximizing one's benefit does not depend solely on one's own actions, but also on the actions of all other individuals participating in the game. The possible outcomes of such »games«, therefore, depend on anticipating another individual's intentions and, similarly, on the anticipations of others with respect to one's own intentions. In a way, one of game theory's central concerns is the *origin* and *evolution* of cooperation in an environment in which personal and communal successes are closely connected and mutually dependent.

Though game theory *qua* theory is of modern origin (founded by mathematician John von Neumann), its principal concerns are also evident in the past. Don Ross (2001), for example, suggests that Plato was one of the first thinkers to come across problems and paradoxes typical of game theory. In *The Republic*, Socrates raises the following paradox for consideration:

> Consider a soldier at the front, waiting with his comrades to repulse an enemy attack. It may occur to him that if the defense is likely to be successful, then it isn't very probable that his own personal contribution will be essential. But if he stays, he runs the risk of being killed or wounded – apparently for no point. On the other hand, if the enemy is going to win the battle, then his chances of death or injury are higher still, and now quite clearly to no point, since the line will be overwhelmed anyway. Based on this reasoning, it would appear that the soldier is better off running away regardless of who is going to win the battle. Of course, if all of the soldiers reason this way – as they all apparently *should*, since they're all in identical situations – then this will certainly *bring about* the outcome in which the battle is lost. Of course, this point, since it has occurred to us as analysts, can occur to the soldiers too. Does this give them a reason for staying at their posts? Just the contrary: the greater the soldiers' fear that the battle will be lost, the greater their incentive to get themselves out of harm's way. And the greater the soldiers' belief that the battle will be won, without the need of any particular individual's contributions, the less reason they have to stay and fight (quoted in Ross, 2001).

The prisoner's dilemma

Though the situation described by Plato focuses on the moral or psychological characteristics of soldiers, it nevertheless lays open the problem of »rational choice«. The most frequently used game-theoretical illustration of this sort is the prisoner's dilemma, originally construed by mathematician Albert W. Tucker.

One possible scenario of the prisoner's dilemma is as follows: Police inspector Herlocka Sholmes has arrested two criminals suspected of committing a bank-robbery. Since inspector Sholmes lacks any substantial evidence for their crime, she can only hope to divulge a confession from them. Inspector Sholmes applies the following procedure: She places the criminals in two

separate cells, so that they cannot communicate with one another, and offers each of them the following options:

- If you confess, and your partner does not, you will go free and your partner will get 10 years;
- If you don't confess, and your partner confesses, you will get 10 years and your partner will go free;
- If you both confess, you will get 5 years each;
- If you both refuse to confess, you will get 2 years each.

(The assumption is that the two criminals will not meet again, so they don't have to be afraid of mutual revenge.)

The following matrix represents the outcomes of the possible combinations of their actions:

	Confesses **A**	Refuses to confess **A**
Confesses **B**	**5** **5**	**10** **0**
Refuses to confess **B**	**0** **10**	**2** **2**

As for the question as to what the prisoners will do in this situation, the answer is that they will both confess, i.e. »defecting« (mutual confession) will be the *dominant strategy* of this particular game. Naturally, a further question arises: Why is »defecting« going to be the dominant strategy, and why will they both confess, and therefore get 5 years each, instead of refusing to confess (cooperate), and therefore get only 2 years each? It should be remembered that both prisoners will endeavor to do what is »economically rational«, putting all moral (altruism) or emotional (fear) concerns to one side (there is no »shadow of the future« or fear of potential revenge). The logic of self-interest that engenders mutual defecting is rather simple: If the other prisoner confesses, it is better to confess, since that way one reduces one's own prison sentence from 10 to 5 years. If the other prisoner does not confess, again it is better to confess, since that way one reduces one's own prison sentence from 2 to 0 years. This is bad for the prisoners, good for inspector Sholmes, and interesting for game-theorists.

The iterated prisoner's dilemma

In addition to the prisoner's dilemma, game theory also pays attention to the »iterated prisoner's dilemma«. The iterated prisoner's dilemma relates to

the *sequences* of the aforementioned situation (series of prisoner's dilemmas) in which two or more prisoners find themselves in from time to time. Now, when the »shadow of the future« (the fear of revenge) is manifest in the game, let's assume that inspector Sholmes deals with several criminals, who occasionally change partners. When faced with the prisoner's dilemma, it is very likely that the temptation to cheat will continue to be great for each of them – viz. by defecting (confessing) one can still either avoid or considerably reduce his/her prison sentence. But owing to the well-known »maxim« that »stool pigeons« end up badly – usually with »concrete boots« at the bottom of some bay or lake – the end-result is that confessing ceases to be the dominant strategy, and the possibilities of other strategies are open, which the prisoners will consider more advantageous in the long run. Game theory demonstrates that, in the iterated prisoner's dilemma with multiple players, the dominant and (in the long run) most profitable strategy will rarely be fully fledged cheating. The dominant strategy will depend, namely, on several factors: the amount of reward and punishment, the number of players involved and the nature of their strategies, the frequency of mutual interactions, the possibility of opponent-recognition in iterated interaction, remembering opponents' earlier moves, the finite or infinite character of the game, etc.

Tit-for-Tat

The prominent game-theoretician Robert Axelrod (1984) sought to determine whether it is possible to establish a strategy that would yield the best results in playing the iterated prisoner's dilemma. He organized a prisoner's dilemma tournament in which various scholars participated, including mathematicians, psychologists, economists, political and social scientists. In order to participate in the tournament every player sent his/her own computer program (strategy) of the game, which played the prisoner's dilemma with every other program 200 times over. It was played by making simultaneous moves and remembering the opponent's earlier moves. The points system satisfied the following inequalities:

$$(A > B) \wedge (B > C) \wedge (C > D)$$

[A = points of successful defection, B = points of mutual cooperation, C = points of mutual defection, D = points of being defected].

Fourteen different programs/strategies appeared at Axelrod's tournament. The winning program was Tit-for-Tat, submitted by economist Anatol Rapoport. Rapoport's program was quite simple: it cooperated in the first move, and later it continued to play the same thing its opponent previously played. All other programs consisted of various proportions of cooperation and defection. It should be noted that Tit-for-Tat never defeated anyone in

a one-on-one game; it always lost by a small margin or tied. But at the end of the tournament, after summing up the points, Tit-for-Tat was the overall winner. One of the basic reasons for its success was that the »cooperative programs« lost too many points from »defecting programs«, while »defecting programs« lost too many points in their mutual interactions. Tit-for-Tat won the tournament thanks to three »features«: (1) it was oriented towards cooperation, (2) it retaliated (defected) only when necessary, and (3) it had a simple and comprehensible strategy (namely, other cooperative but complicated or incomprehensible strategies were subjected to unintentional exploitation in the game).

Having presented a brief introduction to game theory, I will try to demonstrate how it could be applied in an analysis of the structures and functioning of scientific communities. But before doing that, I will outline one general model of the functioning of science.

Science as scientific selection

It is an obvious fact that the appropriate functioning of modern sciences heavily depends on the direct or indirect collective work of many scientists in the specific field. Moreover, scientific cooperation is generally considered the key mechanism for filtering good from bad science or non-science. A suggestion or criticism, agreement or disagreement from impartial colleagues plays a vital role in achieving the highest possible quality of the final scientific product – be that a theory, an article, a book, etc. The essential point is that the evaluation of individual scientific labor proceeds *impersonally* (independently of an individual's gender, race, religion, etc.), i.e. it proceeds exclusively in accordance with objective methodological standards.

We can illustrate this with an quasi-analogy taken from biology. In the theory of evolution and philosophy of biology there is an ongoing debate about what is a *basic unit* upon which natural selection works: groups of organisms, an individual organism, or individual genes?[1] If we assume, for the sake of this quasi-analogy, that there is such a thing as *scientific selection*, then the question turns out to be: Does scientific selection work on groups of theoreticians, on individual theoreticians, or on individual theories? The logical answer is, of course, that scientific selection, if it really is scientific, works exclusively on *theories*, i.e. given the selective pressure of the scientific community, »good« theories survive, while »bad« theories become extinct.[2]

[1] For a discussion on »units of selection«, see Sober (1993). For the recent »resurrection« of previously abandoned group selection in the philosophy of biology, see Sober and Wilson (2000).

[2] A typical evolutionary model of the growth of (scientific) knowledge can be found in Popper (1972), as well as in Campbell's (1974) »blind variation and selective retention« model of »epistemic growths«.

How does this »evolutionary« model of the scientific selection of theories work in the real world? Consider the example of writing peer reviews (blind-reading) for determining whether or not articles should be published in scientific journals. The referee evaluates an author's manuscript exclusively on the basis of the well-foundedness and justifiability of its content. The referee cannot, and indeed should not, know who the *person* under review is (the name of the author[s] and all self-identifying references are usually removed from the article). By the same token, the person under review cannot, and indeed should not, know who his/her referee is, for academic »good manners« forbids the journal's editors to make this public. What counts in the referee's evaluation of the article are the adequacy or accuracy of the data presented, the logical validity of inferences, general consistency, etc.

The aforementioned »rules of the game« are supposed to guarantee the highest possible level of objectivity in the evaluation of an author's scientific work. Due to anonymity, excluded from the refereeing procedure are all potential »deals« between the two parties, as well as scientifically inappropriate – altruistic or egoistic – gestures the referee may display towards the refereed. In other words, refereeing rules prevent any kind of Tit-for-Tat strategy to evolve – simply because neither referee nor refereed knows with whom they are dealing. The only relation that really exists is *referee-text*, not *referee-refereed*, which, due to the »ontological asymmetry« of its counterparts, is not sufficient for developing any decipherable strategy that would be of interest for game theory. The only »strategy« the referee can employ is to be as rigid as possible in his/her evaluation of the article, and the only »strategy« the refereed can employ is to write the best possible article he/she can. A science wherein these strategies dominate would definitely flourish. However, as we shall see, different situations are possible, where things drastically diverge from the above model.

Before considering such situations, let's first see which features of scientific communities could justify applying game-theoretical models (like iterated prisoner's dilemma) in its analysis:

(1) *Time*. The iterated prisoner's dilemma is an unlimited game or a game of very long duration. The same can be said of scientific communities, which usually »last longer« than individual scientists. For example, just as game theory is used for analyzing behavioral traits in animal populations over long periods of time, in which many generations are born and die,[3] we can assume, by way of analogy, that its methods could also be efficiently utilized for analyzing behavioral traits in scientific populations. In any case, »game theory« and »scientific community« are minimally connected by the fact that the former *investigates* the complex systems of long-term interactions among large numbers of individuals; the latter *is* one such system.

[3] On the application of game theory in the theory of evolution, see Maynard-Smith (1982).

(2) *Action*. In the iterated prisoner's dilemma the same possibilities of action are open to all players: cooperation, non-cooperation, or a certain proportion of the two. The members of scientific communities, by participating in one such »long-term game«, can also freely choose how they will react to each other; they can even choose whether they want to be »active« or »passive« members of a scientific community.

(3) *Rational selfish interest*. In both the iterated prisoner's dilemma and scientific communities, players/scientists behave in an economically rational manner, i.e. they minimize personal cost and maximize personal benefit. Being a member of a given scientific community involves certain »costs«, which are closely linked to the amount of »benefits« the member receives – either from particular colleagues or from the entire scientific community. A common indicator of this relationship are *scientific societies*, whose statutes include, almost without exception, the following proposition in one form or another: »The society operates in order to (a) support and protect the interests of science as such, and (b) to support and protect the interests of its individual members.«

The emergence of pseudoscience

Though scientific communities contain all the aforementioned elements in order to advance the quality of scientific work, the very same elements occasionally produce, by means of a somewhat mysterious inversion, an entirely opposite outcome, i.e. some scientific communities begin to loose their scientific reputation and assume pseudoscientific nuances. It is not particularly surprising, then, that there are highly respected and less respected scientific communities. For example, the scientific community of biologists in Stalin's Soviet Union (whose leading protagonist was the »renowned« Trofim Lysenko, whose »scientific reputation« was based mainly on adjusting biological facts and theories to current political affairs) was undoubtedly much less respectable than, say, the British community of biologists of the same period. But even though the Soviet biological community was pseudoscientific, and the British wasn't, they *both* had the demeanor of serious scientific communities (both »sciences« were taught in universities, had their institutes, organizations, journals, etc.). In both instances, scientists cooperated amongst themselves, but in one of them this produced good and in another bad science.

Notwithstanding, the claim that scientific cooperation can result in pseudoscience should not be taken to mean that parasitic scientists, who intentionally undermine the foundations of their profession, inhabit certain scientific communities. In one sense, the evolution of pseudoscience proceeds gradually, much like biological evolution, through small and seemingly neutral modifications, which in the long run, unfortunately, result in something en-

tirely new, daunting and damaging. The question, therefore, that requires answering is: *What mechanisms*, *and under what circumstances*, can cultivate the occurrence and evolution of pseudoscience in scientific communities?

To begin with, one thing is certain: pseudoscience does not emerge *ex nihilo*; it grows like a weed on the soil of real science. If it stems from real science, then this indicates that some scientific mechanisms are not functioning well. Pseudoscience must bear resemblance to real science, viz. it must at least »formally« satisfy the customary scientific standards. This, of course, is best attainable *within* science, in an environment in which this »mimicry« is possible and profitable. Consequently, pseudoscience will also depend on the cooperation of »scientists« who aren't guided by maximal, but minimal methodological standards. One should thus expect pseudoscience to be neither anarchistic nor chaotic, although it originates from ineffectiveness, anarchy, and major or minor anomalies in real science. In other words, the Smithian *invisible hand* can create pseudoscience only when the Hobbesian *visible hand* is not around to prevent it from doing its work. In order to illustrate this, we can imagine the possible atmosphere at the »macro« and »micro« levels of a typical pseudoscientific community.

Pseudoscientific individual selection

First, we should remind ourselves of the previously mentioned quasi-analogy between the units of selection in biological evolution and the units of selection in »scientific evolution«. Whilst in biology controversies still persist regarding the basic units of natural selection (gene, individual, or groups of individuals), the established view in the theory of science is that the unit of scientific selection is neither a group of theoreticians nor a single theoretician, but the *theory itself*. A scientific community (»environment«) exerts epistemic pressure (»selection«) on a theory, which is then acknowledged, if it resists (»survives«), as part of real science. The selection of a theoretician *qua* theoretician is not legitimate and is, in fact, epistemologically trivial and counter-productive.

The state of affairs in pseudoscience seems different, however. Pseudoscience *qua* pseudoscience cannot have a theory as its basic unit of selection. Namely, this could come »hazardously« close to postulating a set of *invariant criteria*, on the basis of which some »theories« would be accepted and others rejected, and that might eventually – like a boomerang – bring into question the entire pseudoscientific paradigm. Subsequently, the unit of pseudoscientific selection will in most cases be the individual theoretician, wherein *selfish interest*, accompanied with a large dose of »rational behavior«, assumes the central role. To illustrate, let's imagine the following situation, similar to the prisoner's dilemma, but one in which two scientists play the role of prisoners. We will call it the »referee's dilemma«.

The referee's dilemma

The editors of a scientific journal sent two received articles for refereeing to two scientists. As it turns out, scientist A received the article of scientist B, and scientist B received the article of scientist A. Let's also suppose that the »conditions« under which this situation occurred are as follows:

(1) The journal is published in the national language of a small country;

(2) The circle of the journal's contributors consists for the most part of scientists from the country in which the journal is published;

(3) The editors sent articles for refereeing to scientists A and B because they are both experts in the same scientific field.

Owing to conditions (1) and (2), it isn't particularly difficult for scientist A and scientist B to guess whose article they received (the probability of guessing is even greater if the articles were previously read at some local conference, which happens quite often). From condition (3), which states that the two scientists referee each other, it follows that the two scientists are – in one sense – *competitors*. For example, if they both work in the same scientific field (let's call it »posthermeneutic physics«), then they are competitors to some degree, e.g. competing for the position of director of the Institute of Posthermeneutic Research, or the position of chair of posthermeneutic physics, or the position of editor of *The Posthermeneutic Review*, etc. Competitiveness between the two is also intensified by condition (1), which implies that there are very few unoccupied »scientific niches« in the country. Finally, we introduce the additional fact – by no means imaginary! – that published articles bear important scientific points for »obtaining« or »maintaining« some of the said positions.

In these circumstances, the two scientists only know who they are refereeing and that someone is refereeing them. Academic rules and courtesy do not, of course, allow them to reveal publicly who they are refereeing, and the same holds for the journal's editors. For that reason, they don't have to fear revenge or the »shadow of the future«, even if they write negative reviews.

How will our two scientists act and what will be the dominant strategy employed in their game? Though this situation is not completely identical to the prisoner's dilemma (especially because of additional »environmental conditions«), the scientists will most likely reason in a manner not altogether unlike the prisoners. In other words, if the other scientist writes a negative review, it is better to write a negative one. This way the mutual score is maintained at the same level as before the game – don't let the other take the lead! If the other scientist writes a positive review, it is again better to write a negative one. In this way, one considerably increases one's points and decreases the opponent's points, i.e. one increases one's chances of obtaining or maintaining a particular job. Needless to say, both scientists think

in the same fashion – and they both write negative reviews. Bad for scientists, very bad for the journal and posthermeneutic physics, interesting for game theory and the theory of science.

The outcomes of the possible combinations of the two scientists' actions are shown in the following matrix (now the numbers do not represent years spent in prison, but scientific points):

	Negative review **A**	Positive review **A**
Negative review **B**	**2** **2**	**0** **10**
Positive review **B**	**10** **0**	**5** **5**

In keeping with the prisoner's dilemma, the points system in the »referee's dilemma« satisfies the following inequalities:

$$(A > B) \wedge (B > C) \wedge (C > D)$$

[A = points of given negative and gained positive review, B = points of given and gained positive review, C = points of given and gained negative review, D = points of given positive and gained negative review.]

Though the »referee's dilemma« assumes that the »shadow of the future« poses no threat, it is significant that specific proto-conditions appear for the potential growth of the Tit-for-Tat strategy in the *symmetrical sense*. This is due to the »ontological equalization« of counterparts in interaction, i.e. the relation referee-text disappears and the relation referee-refereed comes into sight. What would happen, then, if the refereed, by some means or another, discover (or can guess with high probability) who their referees are?

Once the prospect of the »shadow of the future« returns into the game, certain types of strategic cooperation, in which Tit-for-Tat will be the most appealing and promising, will probably emerge among a number of scientists. Tit-for-Tat will originally develop amongst *individual scientists* having an inferior scientific »pedigree«. On the other hand, however, this will influence good scientists as well – their scientifically justified negative review will, from time to time, be »penalized« with a reciprocal negative review, but this time without scientific justification. Hence the »referee's dilemma« suggests that pseudoscientific selection will work on *each individual member of a scientific community*, regardless of the quality of one's scientific work.

Postmodern »theory« and Sokal's »hoax«. A case of pseudoscientific group selection?

As we previously saw, genuine science does not operate at the level of »group selection«. But what about pseudoscience? It seems that in pseudoscience there can be selection of *groups of individuals* without epistemic justification, just as there can be *individual selection* without epistemic justification. Namely, now and then we can locate in pseudoscience, and especially in the social sciences and humanities, »groups« of »scientists« engaged in a ruthless »theoretical struggle«, but with unclear epistemic standards. An example of this practice might be found in »eminent« postmodern criticism of the modern natural sciences.

In short, postmodern aggression towards science represents a certain blend of »epistemological relativism« and »social constructivism«, i.e. it claims that there is no »'real' reality, no objective truths external to mental activity, only prevailing versions disseminated by ruling social groups« (Wilson, 1998, p. 22). However, one should be aware of the fact that postmodernism isn't just the latest type of traditional epistemological skepticism. Postmodernists, namely, »reject the very notion of 'truth' itself. They argue that there is no 'objective knowledge' and no 'facts', only personal interpretation, and that 'reason' and 'science' are no better than any other 'myth', 'narrative', or 'magical explanation'« (Cherry, 1998, p. 20). (Of course, »postmodernist authors would be undeserving of discussion if they were not so famous« [Bricmont, 1998, p. 25].) E.O. Wilson offers an apt description of postmodernism:

> Usually leftist in orientation, the more familiar modes of general postmodernist thought include Afrocentrism, constructivist social anthropology, 'critical' (i.e. socialist) science, deep ecology, ecofeminism, Lacanian psychoanalysis, Latourian sociology of science, and neo-Marxism. To which add all the bewildering varieties of deconstruction techniques and New Age holism swirling round about and through them (Wilson, 1998, p. 22).

Today it is easy to notice that postmodernism has became one of the leading »intellectual fashions«. It is frequently encountered not only in serious academic circles, but also in the mass media. But just as the Trojans were fatally deceived by Ulysses' famous »Trojan horse«, so too was an epicenter of postmodern thought – the North American journal of cultural studies *Social Text* – hoodwinked by a modern type of »Trojan horse«, namely the »Trojan article«.

Alan Sokal, a physicist from the University of New York, played the role of Ulysses in this battle of »two cultures«. In 1996 he submitted to *Social Text* an article bearing the exotic title »Transgressing the Boundaries: Towards a Transformative Hermeneutics of Quantum Gravity«. The article was a cunningly devised fraud, filled with a bunch of scientific and non-scientific ele-

ments connected in a logically suspicious manner. But the »fundamental silliness« of the article, claims Sokal, was its central thesis »that quantum theory – the still-speculative theory of space and time on scales of a millionth of a billionth of a billionth of a billionth of a centimeter – has profound *political* implications…« (Sokal, 1996b). By performing this »experiment«, Sokal asked himself the question: »… would a leading North American journal for cultural studies – whose editorial collective includes such luminaries as Frederic Jameson and Andrew Ross – publish an article liberally salted with nonsense if (a) it sounded good and (b) it flattered the editors' ideological preconceptions?« (Sokal, 1996b).

The answer, alas, is affirmative. Sokal was surprised at how readily the editors »accepted [his] implication that the search for truth in science must be subordinated to a political agenda, and how oblivious they were to the article's overall illogic«, and emphasizes that »[t]he editors of *Social Text* liked [his] article because they liked its *conclusion*« (Sokal, 1996b). Naturally, the problem we are interested in is not postmodern ideology, but the epistemically dubious theoretical work – and celebrity! – of numerous postmodern authors. As Jean Bricmont says, »we're dealing with people who obviously want to make a theoretical work«, but he also believes that the entire case surrounding Sokal's hoax »uncovered an extreme form of intellectual abuse – namely, academics trying to impress a nonscientific audience with abstruse scientific jargon that the academics themselves do not understand very well« (Bricmont, 1998, p. 23–25). In keeping with this, our principal concern is whether the case of »Sokal's hoax« can be integrated into our discussion on game theory and units of selection.

At least two facts support the hypothesis that the case reveals pseudoscientific group selection at work (to be sure, on the »postmodern side«):

1. That it was an instance of *pseudoscientific selection* is supported by the fact that the article (in spite of numerous and very specific physical and mathematical sections) wasn't evaluated according to appropriate scientific standards – moreover, the editors later declared that the referees didn't even evaluate the article at all (Polšek, 1998, p. 231).
2. That it was an instance of pseudoscientific *group selection* is supported by the fact that the article was accepted *exclusively* on the basis of its *general tone* and *conclusion*, which »flattered the ideological preconceptions« of a specific *group of theoreticians* (who, according to Sokal, represent »a self-perpetuating academic subculture«).

Michael Ruse's following portrait of this clique of self-proclaimed »critics« could also substantiate the hypothesis that the editors of *Social Text* accepted Sokal's article with the intention of »signing« another »theoretical alliance«, thereby strengthening their »group«:

> Searching out allies and molding opinion to their ends, these critics have no limits to their intentions and their arrogance. Little wonder, then, that the editors of *Social Text* seized happily on Sokal's submission – a piece rubbishing the pretensions of modern science and from a scientist himself. Exposing a piece to referees could only lead to criticism, and that is precisely what the editors did not want (Ruse, 2001, p. 4).

If we are to believe Sokal and Ruse, the same »group« or »subculture« would most likely reject the article if it had a different conclusion. For the present discussion, it is important to bear in mind the fact that the decisive factor of acceptance wasn't scientific, and that »selection« – *in ultima linea* – was performed by a homogenous group of like-minded theoreticians that was (surprisingly?) enthusiastically interested in criticizing the methods of the modern natural sciences. On could object, no doubt, that this was not a case of a group *being selected* (as required by the group selection hypothesis), but of *group selecting*. This, however, only appears to be a semantic distinction, not a real one – at least in the case of »Sokal's hoax«. In other words, the postmodernists grouped around *Social Text* accepted Sokal's article because it flattered their *collective* theoretical biases, and the selfsame postmodernists felt *collectively* attacked when Sokal revealed his »hoax« and their ignorance. It seems, therefore, that the motto of this postmodern »super organism« is »select or be selected« – and it is always better to select and be selected as a group, especially in postmodernism. Furthermore, the same conclusion about postmodern pseudoscientific group selection, paradoxically or not, can also be drawn from the postmodernists' canon. If »objective truth« does not exist, and the »acceptance« of a certain theory depends solely on its being »socially constructed« or »enforced«, then this process, by definition, must proceed via group selection in the guise of forming alliances or groups around – to paraphrase Richard Dawkins – various »meme pools«, such as journals, institutes, university departments, etc. Expressed rhetorically, isn't it all just a »social construction« of »social construction«?

»Sokal's hoax«, together with numerous subsequent attempts by the editors of *Social Text* to defend themselves, confirms that a number of »scientists«, without all (usual) methodological criteria, identified their individual theoretical beliefs with the firm theoretical canon of a specific group. In such a group, naturally enough, there can be no place for »modern« scientists like Sokal or Ruse, because their views on science diverge and conflict with the »methodology« or »metatheory« of the group. The criterion that was in one sense primary, and which is actually implicit in postmodern »theory«, is blind agreement with the »revealed truths« of the group. As already noted, this is not surprising, since postmodernism »questions accepted standards and emphasizes how social context affects beliefs and theories« (Cherry, 1998, p. 20). This definitely caused certain shifts in the »ontology« of pseudoscientific selection, whereby the central unit of selection, besides individuals, also became *groups of individuals* in their social context.

Tit-for-Tat: A pseudoscientifically stable strategy?

Now we can ask one of the central questions of game theory: Can Tit-for-Tat become *evolutionary stable strategy* in a certain pseudoscientifically infected scientific population? Evolutionary stable strategy (ESS) is defined as »a set of rules of behavior that, once adopted by members of a group, is resistant to replacement by an alternative strategy« (Cartwright, 2000, p. 347). In other words, ESS is a strategy that, once it spreads, cannot be substituted by any other strategy, in which case all other strategies are condemned to extinction. Tit-for-Tat will most likely become ESS if the following conditions prevail: first, selection does not work on theories, but on theoreticians (i.e. the evaluation of »scientific contribution« is determined *ad hominem*, instead of *ad rem*); secondly, certain »protection mechanisms« (such as the »veil of anonymity« in the refereeing procedure) in the scientific community do not function; and, finally, a sufficient number of »scientists« (even less than 50%) start by using this strategy.

As an illustration, we can conceive of a population of 10 scientists, out of which 6 are genuine scientists and 4 are pseudoscientists. They play the referee's dilemma amongst themselves a certain number of times. But information, we assume, is leaking through the »veil of anonymity«. Let's presuppose that just one peer review is required – negative implies not publishing the article, positive implies publishing the article. Members of both groups want to maximize their own fitness, i.e. they all want to publish their articles. Whilst genuine scientists employ *exclusively methodological criteria* in the refereeing procedure, pseudoscientists, on the other hand, play Tit-for-Tat. In the first interaction, each pseudoscientist writes a positive review for everyone, thus anticipating a reciprocal outcome in the next interaction. The real scientists – because of their serious methodological standards – review positively only the real scientists' articles, and review negatively the pseudoscientists' articles. They act in the same manner in every subsequent interaction. But the pseudoscientists in the second and every subsequent interaction review negatively all the articles of real scientists (because of their non-reciprocal return), and continue to review each other's articles positively.

The following table shows the development of interpersonal relations and individual »scientific careers«. Letters A, B, C and D represent pseudoscientists; letters E, F, G, H, I and J represent real scientists. 1 in subscript format represents a positive review; 0 in subscript format represents a negative review. The sum total of positive reviews is equivalent to the number of published articles.

First interaction		
Scientist under review	**Referees and their reviews**	**Published articles**
A	$(B_1+C_1+D_1)+(E_0+F_0+G_0+H_0+I_0+J_0)$	**3**
B	$(A_1+C_1+D_1)+(E_0+F_0+G_0+H_0+I_0+J_0)$	**3**
C	$(A_1+B_1+D_1)+(E_0+F_0+G_0+H_0+I_0+J_0)$	**3**
D	$(A_1+B_1+C_1)+(E_0+F_0+G_0+H_0+I_0+J_0)$	**3**
E	$(A_1+B_1+C_1+D_1)+(F_1+G_1+H_1+I_1+J_1)$	**9**
F	$(A_1+B_1+C_1+D_1)+(E_1+G_1+H_1+I_1+J_1)$	**9**
G	$(A_1+B_1+C_1+D_1)+(E_1+F_1+H_1+I_1+J_1)$	**9**
H	$(A_1+B_1+C_1+D_1)+(E_1+F_1+G_1+I_1+J_1)$	**9**
I	$(A_1+B_1+C_1+D_1)+(E_1+F_1+G_1+H_1+J_1)$	**9**
J	$(A_1+B_1+C_1+D_1)+(E_1+F_1+G_1+H_1+I_1)$	**9**

Second and each subsequent interaction		
Scientist under review	**Referees and their reviews**	**Published articles**
A	$(B_1+C_1+D_1)+(E_0+F_0+G_0+H_0+I_0+J_0)$	**3**
B	$(A_1+C_1+D_1)+(E_0+F_0+G_0+H_0+I_0+J_0)$	**3**
C	$(A_1+B_1+D_1)+(E_0+F_0+G_0+H_0+I_0+J_0)$	**3**
D	$(A_1+B_1+C_1)+(E_0+F_0+G_0+H_0+I_0+J_0)$	**3**
E	$(A_0+B_0+C_0+D_0)+(F_1+G_1+H_1+I_1+J_1)$	**5**
F	$(A_0+B_0+C_0+D_0)+(E_1+G_1+H_1+I_1+J_1)$	**5**
G	$(A_0+B_0+C_0+D_0)+(E_1+F_1+H_1+I_1+J_1)$	**5**
H	$(A_0+B_0+C_0+D_0)+(E_1+F_1+G_1+I_1+J_1)$	**5**
I	$(A_0+B_0+C_0+D_0)+(E_1+F_1+G_1+H_1+J_1)$	**5**
J	$(A_0+B_0+C_0+D_0)+(E_1+F_1+G_1+H_1+I_1)$	**5**

What will be the ratio of scientific and pseudoscientific articles in the population, i.e. which group will publish more articles, after, say, four interactions?

The group of genuine scientists will have 144 articles. However, this number should be reduced by 24 because of the pseudoscientists' reviews from the first interaction – remember: those articles were not positively reviewed according to methodological criteria, but according to the Tit-for-Tat strategy. These 24 articles, therefore, are »pseudoscientific« articles published »accidentally« by genuine scientists. Nevertheless, the situation does not yet look all that bad for genuine science. The score still remains 120 scientific articles as opposed to 60 pseudoscientific articles (the pseudoscientists published 48

articles thanks to their reciprocal positive reviews, plus 12 »accidental« pseudoscientific articles written by genuine scientists). If we observe the situation at the individual level, every genuine scientist will have 24 published articles (including 4 »accidental« pseudoscientific articles), and every pseudoscientist will have 12. Consequently, genuine scientists will again occupy all the positions vital for controlling scientific policy in this particular scientific community.

But this conclusion is invalid because one of its premises is false. Namely, it is erroneous to think that genuine scientists will *always* write a positive review for each other. This is evident, for example, from the research conducted by Harriet Zuckerman and Robert Merton (Zuckerman and Merton, 1971; reference in Lelas, 1990, pp. 215–229). In analyzing the archive of a prominent scientific journal, *The Physical Review*, for the period 1948–1956, the authors noticed that the journal published more than 80% of the manuscripts received. But in the analysis of other journals, the statistics were different: three journals for history rejected on average 90% of manuscripts; five journals for philosophy rejected 85% of manuscripts; fourteen journals for sociology rejected 78% of manuscripts; five journals for mathematics rejected 50% of manuscripts; and five journals for chemistry rejected 31% of manuscripts, etc. (Zuckerman and Merton, 1971; Lelas, 1990, p. 222). Zuckerman and Merton concluded that »the more humanistically oriented the journal, the higher the rate of rejecting manuscripts for publication« (quoted in Lelas, 1990, p. 222).

If we assume, for the sake of argument, that our 10 (pseudo)scientists belong to a certain »humanistically oriented« scientific community (recall *Social Text*!), then the number of genuine scientific articles – individually and totally – must be reduced by 2/3. Now, the ratio of scientific and pseudoscientific articles in the community will be 60:40 for the latter. Each pseudoscientist will have 12 published articles, and every genuine scientist will have 8. In spite of there being a majority of genuine scientists, the outcome will be that »pseudoscientific heresy«, in the form of the Tit-for-Tat strategy, will spread throughout the community and begin to dominate. At one point it will become the evolutionary stable strategy because it cannot be substituted by any other alternative strategy – be it scientifically justified or not. One of the crucial conditions for the development of this state of affairs is that the basic factor of »success« in the scientific community ceased to be *methodological* (as is the case in genuine science), and became *strategic* (just like in politics, where occasional coalitions are the only way of surviving).[4]

[4] If Tit-for-Tat can become ESS at the level of individual selection, one could ask whether it is possible to become ESS at the level of group selection. This question is difficult to answer, especially because pseudoscientific groups can be tricked by hoaxes like the one performed by Alan Sokal. However, it seems that Tit-for-Tat cannot become ESS at the level of group selection, because this would imply convergence towards one »supergroup«. This converging is possible, but only up to a certain point at which – due to the rational selfish interests of individuals and limited scientific »resources« – the disintegration into competing sub-groups begins.

Pseudoscientific mass extinctions

If Tit-for-Tat becomes the evolutionary stable strategy in a particular scientific community, it will be almost impossible to dismantle it from *within*. Nonetheless, even if dismantling this »pseudoscientific reciprocal altruism« may not be possible from *within*, it is possible from *without* – with help of certain *non-scientific* influences. Here's another biological analogy as an illustration. A well-known fact from the theory of evolution is that species unadapted to their environment do not reproduce and, consequently, die out. But there are also instances when a highly adapted species still goes extinct owing to rapid changes of its environment.

The most famous example of mass extinction during the course of evolution is that of the dinosaurs. They quickly disappeared, it would seem, because a large comet crashed into the Earth and drastically changed its ecosystem. Dinosaurs simply didn't have the adequate »survival program« that could help them overcome the new situation, and thus were wiped off from the face of the Earth. As Bertrand Russell says, »[t]he dinosaurs were, in their day, the lords of creation, and if there had been philosophers among them not one would have foreseen that the whole race might perish« (Russell, 1962, p. 7). The same could be said for large pseudoscientific populations (recall Lysenko's Stalinist biology!). Some of them are very quick and successful in adapting to new environments, and their members are often genuinely convinced that what they do is genuine science that can't »go extinct«. This, however, is not always the case.

That *non-scientific* facts (»a comet striking«) sometimes influence the collapse (»extinction«) of pseudoscientific populations (»dinosaurs«) and their evolutionary stable strategies can be confirmed, for example, by the consequences the fall of communism has had on Marxism. It is well known that within the »Marxist paradigm« (especially in Eastern European countries) there existed many »scientifically based« Marxist approaches and influences in philosophy, social science, economics, art theory, and even in the natural sciences – not just in biology, but in physics as well. MA's and PhD's have been written about numerous Marxist topics. Moreover, Marxism was frequently considered to be a scientific field *sui generis*, and was taught »as such« at secondary and tertiary educational institutions. Various groups of »Marxist scientists« likewise created their own »interpretations« of the »original Marx«, proclaiming them to be »final« and »objective«.

Unfortunately, this alleged »objective« character of certain Marxist interpretations was more often than not just a synonym for »official«. Beneath the surface, as Neven Sesardić claims, the situation was of such a kind that

> philosophical arguments were not used for establishing one philosophical position and its victory over alternative approaches, but rather non-philosophical devices [were used], such as propaganda, institutional protection and the en-

> forcement of particular beliefs, various sorts of administrative and unofficial pressure on different opinions, the creation of a general atmosphere of intolerance and dogmatism (Sesardić, 1991, p. 217).

We don't want to get side-tracked here by a discussion on the scientific dubiousness of Marxism. However, we can point out that the massive production of books and journals, which are today seen mainly in library garbage dumps, testifies quite convincingly to the »stretchy standards« of this »science«. In short, in the wake of the social and historical falsification of Marxism through the collapse of communist regimes, there is very little valuable scientific material left from the former gigantic Marxist paradigm. And the earlier passionate advocates of Marxism that didn't manage to adapt to the new conditions simply »scientifically died«.

»Concluding (pseudo)scientific *postscript*«

Pseudoscience (a) *emerges* in circumstances where the mechanisms for filtering science from non-science do not function; (b) *evolves* in scientific communities where such circumstances exist over a long period of time; and (c) *spreads and stabilizes* when sufficient numbers of individuals (less than a majority) »realize« which type of behavior in such an environment is the most economical. Process (c) is particularly rapid when (pseudo)science is co-determined by influential non-scientific interests, including financial motivation and political, ideological or religious belief (e.g. modern creationism).[5] But even in the absence of non-scientific co-determination, the individual interests of scientists – perhaps the desire for quickly advancing one's career or gaining an important position – can »psychologically« suppress their sense of professional responsibility and tempt them to establish specific »scientific alliances«. These selfsame »scientific alliances« then devise »internal« rules (individual selection, using the Tit-for-Tat strategy) and »external« rules of behavior (group selection, using any strategy that gives optimal results).

Almost all the situations depicted in this article are, of course, imaginary, but the emergence and progression of pseudoscientific trends is very difficult to trace or predict. In actual scientific communities, things are definitely more complex, and mainly because of details and processes not mentioned here. Some of them include changes and »revolutions« in the world of science, the expansion of interdisciplinary research programs, international scientific collaboration, the influence of politics and public opinion on financing science, the application of scientific discoveries in technology and industry (economic justification), psychological and financial motives for doing science, etc. Notwithstanding, the network of various interactions in scien-

[5] On modern creationism and its dubious scientific ambitions, see Ehrenreich and McIntosh (1998), Dawkins (1998), Ruse (1998).

tific communities unquestionably involves a certain *ordered meta-structure* that can be modeled by game theory. The optimal corroboration of this hypothesis would probably be a detailed examination of scientific processes in which individuals and their »preferences« play a decisive role: in the mutual writing of peer reviews, publishing reviews of other scientists' books in journals and magazines, citing other scientists' articles, recommending one's colleagues for scientific awards or honorary degrees, reciprocal invitations to conferences, forming »alliances« in the elective assemblies of professional societies, etc.

All the »processes« mentioned consist of the mutual interactions of many individual scientists. Moreover, they are – at least to some extent – accompanied by records and files that provide insight into their development and history, i.e. they open up the possibility of a subsequent scientific evaluation of the outcomes of those interactions. Game theory, therefore, together with the history, sociology and philosophy of science, might illuminate various individual and group (pseudo)scientific strategies, as well as trace the causes of their positive and negative consequences. Likewise, game-theoretical models of long-term scientific interactions could also assist in keeping the invisible hand of pseudoscience under the control of the visible hand of genuine science: not just *a posteriori*, as the reconstruction of the past, but *a priori* as well, as the construction of the possible future.

BIBLIOGRAPHY

Axelrod, R. 1984. *The Evolution of Cooperation* (New York: Basic Books).

Blackburn, S. 1996. *The Oxford Dictionary of Philosophy* (New York: Oxford University Press).

Bricmont, J. 1998. Exposing the emperor's new clothes. *Free Inquiry* 4: 23–26.

Campbell, D. T. 1974. Evolutionary epistemology. In Schilpp, P.A. (ed.) *The Philosophy of Karl Popper* (LaSalle, Ill.: Open Court Publishing Co., *The Library of Living Philosophers*,Vol. 14, I & II), 413–463.

Cartwright, J. 2000. *Evolution and Human Behaviour* (London: Macmillan Press).

Cherry, M. 1998. Truth and consequences. *Free Inquiry* 4: 20.

Dawkins, R. 1998. When religion steps on science's turf. *Free Inquiry* 2: 18–19.

Ehrenreich, B. and McIntosh, J. 1998. Sizing up »secular creationism«. *Free Inquiry* 2: 23–25.

Lelas, S. 1990. *Promišljanje znanosti* [*Contemplating Science*] (Zagreb: Hrvatsko filozofsko društvo).

Maynard-Smith, J. 1982. *Evolution and the Theory of Games* (Cambridge: Cambridge University Press).

Polšek, D. 1998. Sokalova »psina«: Nova metoda znanstvene prijevare i njezina relevancija za sociologiju znanosti i kulture [Sokal's »trick«: A new method of scientific

fraud and its relevance for the sociology of science and culture]. *Društvena istraživanja* 33–34: 223–239.

Popper, K. R. 1972. *Objective Knowledge: An Evolutionary Approach* (Oxford: Clarendon Press).

Ross, D. 2001. Game Theory. In Zalta, E.N. (ed.) *The Stanford Encyclopaedia of Philosophy* (http://plato.stanford.edu/entries/game-theory).

Ruse, M. 1998. Answering the creationists. *Free Inquiry* 2: 28–32.

Ruse, M. 1999. *Mystery of Mysteries: Is Evolution a Social Construction?* (London: Harvard University Press).

Russell, B. 1962. Science and human life. In Newman, J.R. (ed.) *What is Science?* (New York: Washington Square Press), 3–21.

Sesardić, N. 1991. *Iz analitičke perspektive* [*From an Analytical Point of View*] (Zagreb: Sociološko društvo Hrvatske).

Sober, E. 1993. *Philosophy of Biology* (Boulder and San Francisco: Westview Press).

Sober, E. and D. S. Wilson 2000. *Unto Others: The Evolution and Psychology of Unselfish Behaviour* (London: Harvard University Press).

Sokal, A. 1996a. Transgressing the boundaries: Towards a transformative hermeneutics of quantum gravity. *Social Text* 46–47: 217–252.

Sokal, A. 1996b. A physicist experiments with cultural studies. *Lingua Franca*, May/June: 62–64.

Wilson, E. O. 1998. Back to the enlightenment. *Free Inquiry* 4: 21–22.

Zuckerman, H. and Merton, R. 1971. Patterns of evaluation in science. *Minerva* 9: 66–100.

Recenzentova dilema. Etika znanstvenih zajednica i teorija igara

SAžETAK: U članku se tvrdi da se različite devijacije od osnovnih načela etosa znanosti – prije svega pojava pseudoznanosti u znanstvenim zajednicama – mogu formulirati i objasniti pomoću specifičnih modela teorije igara kao što su zatvorenikova dilema i ponovljena zatvorenikova dilema. Članak se, prema tome, neizravno bavi i deontologijom znanstvenoga rada, pri čemu je ključna pretpostavka da u etici znanstvenih zajednica nema mjesta moralnom skepticizmu, a kamoli moralnom antirealizmu: naime, znanstveno »ispravno« ponašanje smatra se jasno definiranim i razlučivim od znanstveno »pogrešnog« ponašanja na osnovi općeprihvaćene maksime znanstvenoga rada kao potrage za znanjem isključivo radi znanja. Nakon izlaganja osnovnih načela teorije igara, pokazuje se – koristeći se imaginarnim i zbiljskim slučajevima, kao i nekim stavovima iz filozofije biologije (rasprava o jedinicama selekcije) – kako bi se ovu vrstu razmišljanja moglo primijeniti u analizi funkcioniranja znanosti.

KLJUČNE RIJEČI: Teorija igara, zatvorenikova dilema, ponovljena zatvorenikova dilema, znanstvene zajednice, pseudoznanost, biološka i znanstvena selekcija, jedinice selekcije.